A Machine Learning-based Framework for Preventive Maintenance of Sewer Pipe Systems

by

Xianfei Yin

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Civil (Cross-disciplinary)

Department of Civil and Environmental Engineering
University of Alberta

ABSTRACT

The municipal drainage system is a key component of every modern city's infrastructure. However, as the drainage system ages, its pipes gradually deteriorate at rates that vary based on the conditions of utilization (i.e., intrinsic conditions) and other extrinsic factors such as the presence of trees with deep roots or the traffic load above the sewer lines, which collectively can impact the structural integrity of the pipes. As a result, regular preventive maintenance of the drainage system is extremely important since replacement is not only costly, but, more importantly, can disturb the daily routines of citizens. As for preventive maintenance, closed-circuit television (CCTV) inspection has been widely accepted as an effective inspection technology for buried infrastructure. In practice, the CCTV inspection of sewer pipes is scheduled by the municipal department according to a set of criteria that prioritize the pipe sections and the frequency of the visits. After the onsite CCTV inspection process, the CCTV video footage is sent to the offsite office, where technologists (trained professionals) must watch through the entirety of the video footage in order to assess the condition of the corresponding pipes. When the technologists watch the video, pipe defects (such as cracks, fractures, roots, deposits, broken, and holes) that appear in the video need to be classified according to certain standardized nomenclature (such as PACP, WRc). After recording the existing defects, a pipe score is calculated, which will serve as the foundation for scheduling the pipe inspection and repair plan in the future. The above-mentioned preventive maintenance process is a time-consuming and costly operation. In the meantime, machine-learning technologies and computing power have increased rapidly in recent years and both are used in various engineering areas to improve productivity and the level of automation. In this context, this research proposes a

machine learning-based framework to facilitate the preventive maintenance of sewer pipe systems. The ultimate goal of the research is to improve the productivity, consistency, and automation of the CCTV inspection-based sewer pipe preventive maintenance. To accomplish this aim, the following five objectives targeting on the optimization of each step of the maintenance process guide the activities of the research: 1) Develop a data-driven framework for modeling the productivity of the CCTV recording process for sewer pipes. This involves modelling the video recording process to predict the video duration. 2) Develop a deep learning-based framework for an automated defect detection system for sewer pipes. The targeted defects and construction features will be detected and labeled in the CCTV video. 3) Develop a video interpretation algorithm and corresponding software program. The text information, defect information, and other information included in the CCTV video will be exported to a tabulated format (e.g., EXCEL, database), which will serve as the source of information for pipe rating purposes. 4) Develop and analyze a data-driven bi-level sewer pipe deterioration model, which will provide city managers with a basis for scheduling preventive maintenance at the neighborhood- and individual-level. 5) Develop an input model of the CCTV inspection data of sewer pipes by examining the inherent characteristics of the historical dataset employing Markov chain-based random generation and validation. The generated dataset will be the input for scheduling preventive maintenance in a situation where there is insufficient data. The outcome of this research is expected to make significant contributions by proposing a machine learning-based framework for sewer pipe preventive maintenance by reducing human work, increasing productivity, and increasing assessment consistency.

PREFACE

This thesis is the original work of Xianfei Yin. Five journal papers related to this thesis have been submitted or published are listed below. This thesis is organized in paper format following the guidelines for paper-based theses.

1. Yin, X., Chen, Y., Bouferguene, A., Zaman, H., Al-Hussein, M., and Russell, R. (2020). "Data-driven framework for modeling productivity of closed-circuit television recording process for sewer pipes." *Journal of Construction Engineering and Management*, 146(8), 04020093. (Chapter 2).

2. Yin, X., Chen, Y., Bouferguene, A., Zaman, H., Al-Hussein, M., and Kurach, L. (2020). "A deep learning-based framework for an automated defect detection system for sewer pipes." *Automation in Construction*, 109, 102967. (Chapter 3).

3. Yin, X., Ma, T., Bouferguene, A., and Al-Hussein, M. "Automation for sewer pipe assessment: CCTV video interpretation algorithm and sewer pipe video assessment (SPVA) system development." Submitted to *Automation in Construction*. (Under review). (Chapter 4).

4. Yin, X., Chen, Y., Bouferguene, A., and Al-Hussein, M. (2020). "Data-driven bi-level sewer pipe deterioration model: Design and analysis." *Automation in Construction*, 116, 103181. (Chapter 5)

5. Yin, X., Bouferguene, A., Al-Hussein, M. (2020) "Markov Chain-based CCTV Inspection Data of Sewer Pipe Random Generation and Validation." *Journal of Construction Engineering and Management*. 146(12), 04020131. (Chapter 6).

ACKNOWLEDGEMENTS

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AHP	Analytic Hierarchy Process
AP	Average Precision
BMA	Bayesian Model Averaging
CC	Crack Circumferential
CCTV	Closed-circuit Television
CH	Crack Hinge
CL	Crack Longitudinal
CM	Crack Multiple
CNN	Convolutional Neural Networks
CS	Crack Spiral
DES	Discrete event simulation
DFC	Defect Frame Cluster
FN	False Negative
FP	False Positive
FPS	Frames per Second
GIS	Geographic Information System
HPF	High-Pressure Flushing
IoU	Intersection Over Union
IQR	Interquartile Range
IR	Infra-red
LOF	Likelihood of Failure
LOS	Level of Service
LPR	Linear Polarization Resistance
mAP	Mean Average Precision
MCMC	Markov Chain Monte Carlo
MHD	Modified Hausdorff Distance
MLP	Multilayer Perceptron
MSCC	Manual of Sewer Condition Classification
NLP	Natural Language Processing

NN	Neural Network
NSM	Non-maximum Suppression
PACP	Pipeline Assessment and Certification Program
PNN	Probabilistic Neural Network
RANSAC	Random Sample Consensus
R-CNN	Region-based Convolutional Neural Networks
RoI	Regions of Interest
RPN	Region Proposal Network
SA	Simulated Annealing
SDD	Single Shot Detector
SPVA	Sewer Pipe Video Assessment
SVM	Support Vector Machine
TN	True Negative
TP	True Positive
UI	User Interface
VIASP	Video Interpretation Algorithm for Sewer Pipes
YOLO	You Only Look Once

Chapter 1: INTRODUCTION

1.1 Background and motivation

The deterioration of urban infrastructure is commonly regarded as a critical problem that all countries worldwide are currently facing, a challenge that is compounded by the reality of limited budgets and finite resources. For instance, it is estimated that Canada's municipal infrastructure deficit is $123 billion and growing by $2 billion annually, where water and wastewater systems account for a quarter ($31 billion) of the total deficit (Mirza, 2007). Sewer pipes represent one of the most important components of the municipal infrastructure, essential to public health and quality of life in urban environments (Meeker, 1971). To curb the potentially severe consequences of an aging sewer pipe (e.g., urban waterlogging, ground subsidence), conducting regular preventive maintenance (which aims at detecting as early as possible the structural and operational deterioration of sewer pipes in order to take the most appropriate corrective actions, e.g., repair, rehabilitation, or if the damage is extreme, replacement of the faulty sections) has proven to be effective (Fenner, 2000). The purpose of maintenance work is to maintain the health of any infrastructure system at a state where wide disruptions in operation are minimized, thus allowing a near-optimal level of service, which is expected to function without interruption (McDonald & Zhao, 2001).

"We want to see our communities become cleaner and more digitally connected. We need to find ways to use data more wisely … adopt new technology that can identify leaks in our sewer systems…" (Infrastructure Canada, 2018). The trend of using advanced techniques, including machine learning, artificial intelligence, data mining, etc., is increasing rapidly in many engineering areas, including the civil infrastructure area, in recent years. In this context, this

research takes advantage of advanced computing power, high-efficiency machine learning algorithms, and the huge amount of related engineering data available to aim at realizing the automation of preventative maintenance of sewer pipe system, increasing the productivity and consistency, and decreasing the amount of required labour for maintenance of the sewer pipe system.

1.1.1 Sewer pipe system

A sewer pipe system (also known as a drainage system or wastewater collection system) serves as a foundational part of municipal infrastructure to keep cities or communities clean and healthy by transporting waste/stormwater to treatment plants or streams in the vicinity (Zaman, 2016; Zaman et al., 2015). The pipes could be made of several different materials, including clay, concrete, metal, polyvinyl chloride, etc. Sewer pipes can also be categorized by the sewer type that is transported by the pipes. For example, there are sanitary pipes, stormwater pipes, and combined pipes (the pipes that transport both sanitary and stormwater) in the city of Edmonton, Canada. The inventory of sewer pipes is extremely large, which results in a large replacement value: for example, there is more than 6,500 km of sewer pipes in city of Edmonton, which has a replacement value estimated at $18 billion (Kurach et al., 2019) and the replacement value of the sewer pipe system infrastructure in Canada is $16,380 per Canadian household according to CIRC (2016).

However, one of the problems for the sewer pipe system is that the health condition of sewer pipes deteriorates as they age, which negatively affects the functionality of the system. The factors that could influence the deterioration rates are complicated, and the problem is investigated thoroughly in Chapter 5. To get an overview of the condition of the infrastructure,

Canada's Core Public Infrastructure Survey (CCPIS) uses the following ratings to indicate the health of the civil infrastructure including the wastewater collection system (CIRC, 2019):

- Very poor: The asset is unfunctional or unstable.

- Poor: Significant deterioration has happened on the asset, and the asset is approaching the end of its service life.

- Fair: The asset starts requiring attention for maintenance. The signs of deterioration start to show and some minor defects appear.

- Good: The asset is adequate, which is usually in the mid-age of its estimated service life.

- Very good: The asset is either in the condition of newly built or recently rehabilitated.

- Unknown: Not enough data has been collected.

According to these ratings in 2019, around 28% of wastewater pipes and 30% of the stormwater pipes are in the condition of poor to fair in Canada, which means a significant amount of the pipes need maintenance operation (e.g., rehabilitation or replacement) in the coming years (for pipes in very poor to poor condition, they need rehabilitation or replacement in next 5-10 years; for pipes in fair condition, they could deteriorate to poor or very poor over the next decades) (CIRC, 2019). Therefore, a proper maintenance strategy should be adopted to keep sewer pipes operating at an adequate level of service, to extend their service life, and to decrease the annual investment required for replacement.

1.1.2 Preventive maintenance: CCTV inspection for sewer pipes

The sewer pipe system plays an important role in modern cities not only because of its functionality but also for the large amounts of resources and budget required to carry out the maintenance operation for the sewer pipe system (Zaman et al., 2015). In practice, two maintenance strategies, preventive maintenance (or proactive maintenance) and reactive

maintenance (or corrective maintenance), are widely employed in sewer pipe maintenance. Of the two, preventive maintenance has proven to be more effective (Fenner, 2000). The primary goal of preventive maintenance is to extend the service life or to extend the amount of time to the next failure for the infrastructure asset considering that replacement is more expensive than maintenance (Fontecha et al., 2016).

In order to support evidence-based decision making for the maintenance and investment plan for the sewer pipe system, the health condition of the pipes are rated according to certain specifications, such as Pipeline Assessment and Certification Program (PACP) (NASSCO, 2015), and Manual of Sewer Condition Classification (MSCC) (WRc, 2013). To collect the data regarding sewer pipe condition that is used in the pipe rating process, closed-circuit television (CCTV) is widely employed in North America for several reasons such as the processes are safe to operate (Navab-Kashani et al., 2015), and the outputs (i.e., videos) are easy to understand (Duran et al., 2002). CCTV inspection is one of the most important tools in the sewer pipe preventive maintenance process since it provides visual information pertaining to the sewer pipe condition, based on which the pipes can be rated or scored, which will serve as the basis for prioritizing pipes in critical condition to receive immediate attention, and also for making future maintenance plans (Cheng & Wang, 2018; Kumar et al., 2018). Note that, there are other tools that are used for water pipe inspection, such as laser scan, pulsed eddy current testing, ground-penetrating radar, SmartBall, etc. (Liu & Kleiner, 2013). The summary of the water pipe inspection technologies is presented in Chapter 3 in detail.

The process of CCTV inspection includes two major processes, namely, on-site video collection, and off-site video assessment (see Figure 1.1).
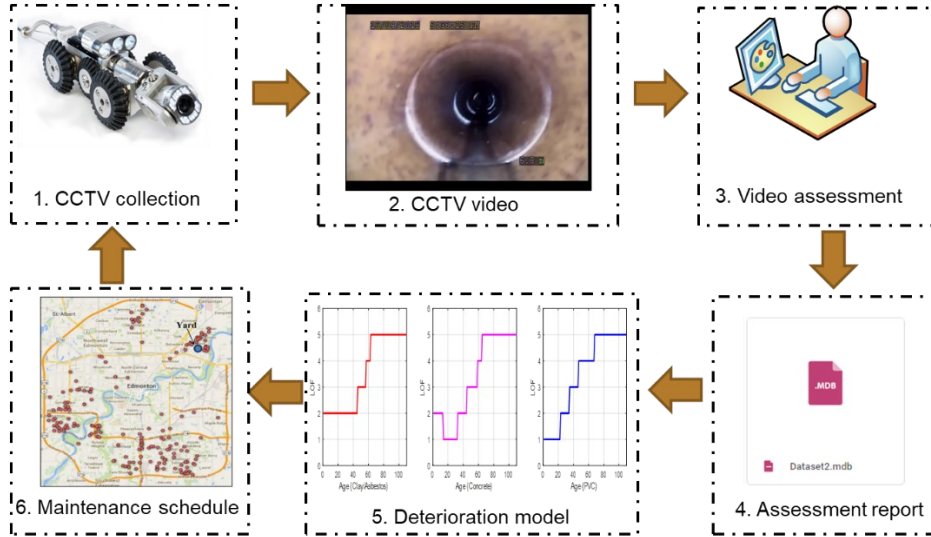
Figure 1.1: Workflow of CCTV inspection for sewer pipes.

The truck crews get work orders from the maintenance office to conduct the CCTV on-site collection as part of their daily operations. After the video collection, CCTV videos are sent to the off-site assessment office where pipe technologists (i.e., trained individuals) watch through all of the newly collected video footage and record the sewer pipe condition in a detailed manner in the sewer pipe assessment report. The report records the detailed information (e.g., type, distance to the starting manhole, clock face location) for the defects (e.g., crack, fracture, broken, hole, deposits, and root) and the construction features (e.g., taps that connect service branches to the mainline) that appear in the video. With the pipe assessment report, the municipal department could develop a deterioration model to predict the future health condition of sewer pipes, which could serve as the foundation for decision-making in planning the future maintenance schedule (including future CCTV inspection plans, rehabilitation plans, and replacement plans, etc.). For instance, sewer pipes that are in bad condition will be assigned a high priority for future CCTV inspection, otherwise, the sewer pipes are assigned a low priority. With the developed maintenance schedule, the next round of CCTV collection can proceed.

5

Note that the above-mentioned video collection and assessment processes are regulated by specifications such as PACP (NASSCO, 2015) and MSCC (WRc, 2013). For example, for the CCTV video recording, the recommended maximum traveling speed is 9 m/min, and the camera crawler should travel smoothly along the pipe unless it encounters taps or severe defects, in that case, the camera crawler should stop beside the targeted taps or defects and turn to the taps or defects to conduct a detailed inspection. The recording process is presented in Chapter 2 in detail. The dataset used in this research was collected following the standards specified in PACP.

Preventive maintenance operations are costly and also require a large amount of resources and labour. However, long-term CCTV inspection processes have accumulated large amounts of valuable data that has not been fully utilized and exploited. Therefore, this research proposes a data-driven framework to fully investigate the potential value of the available data using advanced machine learning techniques to improve the productivity of preventive maintenance for sewer pipes and to increase the level of automation in the CCTV inspection process at the same time.

1.1.3 Machine learning techniques

In recent years, computing power has grown exponentially, which offers enough power to process much more complicated real-life problems that could not have been processed before. In the meantime, more and more data are available in many domains of science and engineering due to advanced data sensing techniques, especially in the context of the Internet of Things (IoT) (Gao et al., 2015). Machine learning techniques serve as a powerful tool to deal with problems found in many aspects of modern society, for instance, to identify objects in images, to translate speech into text, and to select relevant searching results (Lecun et al., 2015). The machine learning techniques have been adopted in construction engineering for many types of problems,

such as object detection techniques for safety management in the construction field (Kanan et al., 2018; Roberts & Golparvar-Fard, 2019), and natural language processing (NLP) techniques for construction contract management (Lee et al., 2019). Machine learning techniques are also used to solve maintenance problems for the sewer pipe system, especially in the area of using object detection techniques for the detection of sewer pipes defects. The relevant literature regarding this issue has been summarized in detail in Chapters 3 and 4. In this research, several cutting-edge machine learning algorithms are used for the purpose of data cleaning (Chapter 2), simulation (Chapters 2 and 6), object detection (Chapter 3), clustering (Chapter 4), optimization (Chapter 4), prediction (Chapter 5), and classification (Chapter 5). The ultimate goal of using these algorithms is to increase the productivity, consistency, and the level of automation of the preventive maintenance process of the sewer pipe system.

1.2 Research objective

The research presented herein is built upon the following hypothesis:

"*The application of machine learning techniques in the analysis of the preventive maintenance process of sewer pipes will improve the productivity of the preventive maintenance process, improve the efficiency of maintenance planning, improve the consistency of the sewer pipe assessment output, and increase the level of automation of the sewer pipe assessment process.*"

The research identifies, in the context of preventive maintenance of a sewer pipe system, one of the limitations of current practices is that the vast amount of data in various formats (i.e., video, images, and numerical database) are not effectively investigated and utilized to facilitate the maintenance operation. Four research questions are proposed aiming at optimizing the CCTV inspection-based preventive maintenance process for sewer pipes:

1. How can the total duration of the CCTV recording process be predicted?

2. How can the level of automation for the CCTV assessment process be improved?

3. How can an effective deterioration model of sewer pipes be developed?

4. How can the size of the dataset for machine learning-based/data-driven research be increased?

To answer these questions, the following objectives (see Figure 1.2) are pursued in this research:

1. Develop a data-driven framework for modeling the productivity of the CCTV recording process for sewer pipes with the intent to model the process of the CCTV recording and to predict the duration of the collected videos.

2. Develop a deep learning-based framework for an automated defect detection system for sewer pipes. The targeted defects and construction features will be detected and labeled in the CCTV video automatically with this proposed system.

3. Develop a video interpretation system and corresponding software. Text information, defects information that is included in the CCTV video will be exported to a tabulated format (e.g., EXCEL, database.), which can serve as the basis for further pipe rating purposes.

4. Design and analyze a data-driven bi-level (i.e., neighborhood-level, and individual pipe level) sewer pipe deterioration model. A deterioration model that focuses on deterioration at the neighborhood-level and the individual-level of sewer pipes will be developed to provide city managers with a basis for scheduling preventive maintenance at the neighborhood and individual pipe level.

5. Generate and validate random sewer pipe inspection data using Markov chain-based random generation. The inherent characteristics of the historical CCTV inspection data for sewer pipes will be explored, and based on that, the artificial dataset will be generated

and validated. The generated dataset is useful for preventive maintenance in situations where there is insufficient data.
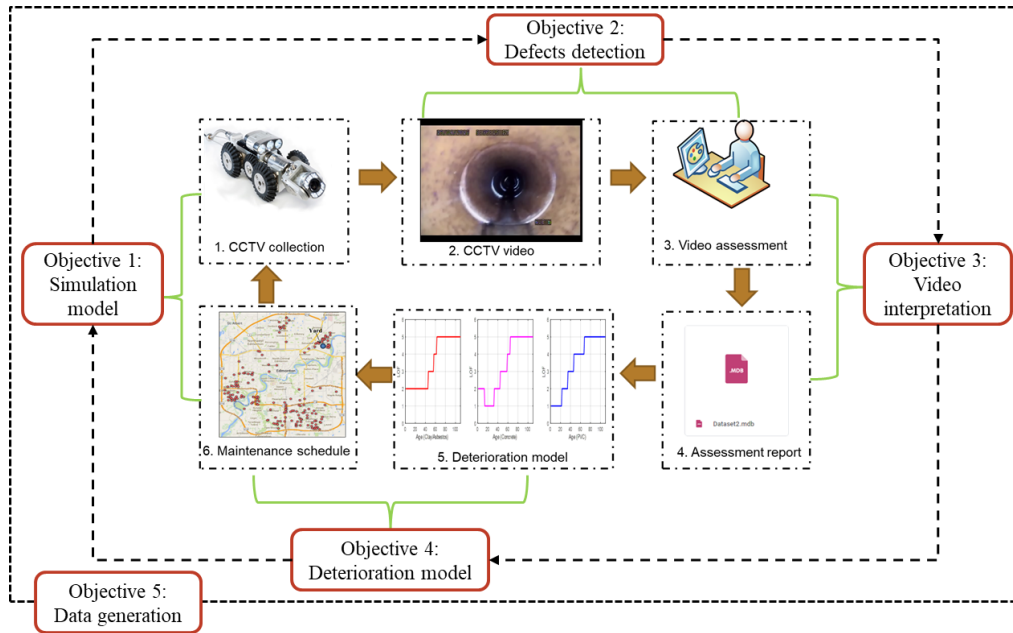
Figure 1.2: Research Objectives.

These five objectives focus on the workflow of the CCTV inspection process for sewer pipes. The sequence of the Objectives 1 through 4 follows the workflow of the CCTV inspection process. The total duration of the CCTV video recording process predicted by Objective 1 can serve as a benchmark for allocating the resources for the video assessment, which is the target of Objectives 2 and 3. The historical assessment reports (which record the health conditions of sewer pipes) generated from Objectives 2 and 3 could serve as the inputs for developing the deterioration model in Objective 4. Objective 5 is a complementary study focusing on the data deficiency problem since data is the foundation for machine learning-based techniques, which are the core techniques used in this research. In addition, the dataset generated in Objective 5 can serve as the input for improving the models proposed in Objectives 1 and 4. Objective 1 aims at solving research question 1; Objectives 2 and 3 focus on research question 2; Objectives 4 and 5 target research questions 3 and 4, respectively.

1.3 Thesis organization

This thesis consists of seven chapters and one appendix. Chapter 1 presents the background and motivation of this research, and includes brief introductions to sewer pipe systems, current best practices in preventive maintenance of sewer pipes, and the application of machine learning techniques. The hypothesis and objectives of this research are also outlined in Chapter 1.

Chapters 2 to 6 present the five objectives of this research mentioned above. In Chapter 2, the CCTV recording process, which is the primary technique used for sewer pipe inspection, is described in detail. To extract the benchmark for the CCTV recording process, a dataset that records the CCTV recording process is processed by the random sample consensus (RANSAC) algorithm. Finally, with the extracted benchmark and the results of the regression analysis, a simulation model to simulate the CCTV recording process is developed and further validated by comparing the results from the simulation model with the historical data. Chapter 3 develops a deep learning-based system to detect, in real-time, the defects and construction features of a sewer pipe shown in a CCTV video. The detailed process followed to build the defect detection system, together with the data and information flow of the system, are highlighted in this chapter. Appendix A presents supplementary data for Chapter 3. In Chapter 4, a video interpretation algorithm for sewer pipes (VIASP) is developed that aims at translating the information from video format to text format. The VIASP and the defect detector described in Chapter 3 are integrated to constitute the video interpretation system, which realizes the automation of the CCTV assessment process. Chapter 5 aims at developing a bi-level deterioration model based on historical sewer pipe maintenance data. A thorough literature review is conducted that focuses on previously developed deterioration models for sewer pipes. Subsequently, two complementary deterioration models, focusing on neighborhood-level deterioration and individual-level

deterioration of sewer pipes, are developed and validated separately. The proposed deterioration model provides the basis for decision making in the sewer pipe maintenance process. Chapter 6 proposes an input model of CCTV inspection data of sewer pipes. The characteristics of the original CCTV inspection dataset are investigated thoroughly. Then, an input model is developed that integrates the Markov chain model with distribution fitting techniques. Finally, the developed input model goes through a rigorous validation process. Chapter 6 provides a framework of how to develop an input model and how to validate the results for other similar studies. Finally, conclusions and the research contributions are summarized in Chapter 7, in addition to limitations and future research directions.

Chapter 2: A DATA-DRIVEN FRAMEWORK FOR MODELING PRODUCTIVITY OF CCTV RECORDING PROCESS FOR SEWER PIPES[1]

Preface

This chapter investigates the details of CCTV recording operation, which is one of the major processes during the CCTV collection. The CCTV recording operation provides the raw data (which is the CCTV video) that serves the foundation for decision-making in the preventive maintenance of sewer pipe systems. The main objective of this chapter is to develop a model to predict the total duration of the CCTV recording process. The productivity of CCTV recording could be influenced by many factors, which makes the prediction difficult. In order to have a deep understanding of the CCTV recording process, the RANSAC algorithm is used to extract the benchmark from the original dataset and separate the dataset into three sections. Each section of the data is analyzed separately, and then all the findings are integrated into a simulation model to mimic the CCTV recording process and to predict the total duration of this operation. Chapter 1 investigates the first step (i.e., CCTV recording) of preventative maintenance of sewer pipes (see Figure 1.2). A simulation model is developed in this chapter to predict the total duration of the CCTV recording, which can be beneficial to the preventive maintenance in two-folds: 1) facilitate to make detailed scheduling for the future CCTV recording operation based on the available budget; 2) the predicted total duration of CCTV videos could help the municipal

---

[1] A version of this chapter has been accepted for publication in *Journal of Construction Engineering & Management*, as follows: Yin, X., Chen, Y., Bouferguene, A., Zaman, H., Al-Hussein, M., and Russell, R. (2020). "Data-driven framework for modeling productivity of closed-circuit television recording process for sewer pipes." *Journal of Construction Engineering and Management*, 146(8), 0402009. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001885. It also has been reprinted with permission from ASCE.

department allocate their resources on the video assessment process, which is the main focus in Chapters 3 and 4.

2.1 Introduction and background

Regular preventive sewer pipe maintenance is an important activity for municipalities since the consequences of undetected deterioration problems of pipes, which generally are due to factors such as aging, external force, deposits, etc., (Chughtai & Zayed, 2008; Wirahadikusumah, et al., 1998) may be dire since they can not only be costly but more importantly can lead to tragic accidents, e.g. fatalities because of collapsing roads due to sinkholes. Generally, several activities are involved in the maintenance operation for sewer pipes, such as visual inspection, low-pressure flushing, high-pressure flushing, catch basin cleaning, closed-circuit television (CCTV) inspection, and hydromechanical cleaning (Zaman et al., 2017). These activities are performed according to a pre-defined schedule except in the case of emergency circumstances (e.g., pipe blockage). However, the maintenance of the sewer pipe system is time-consuming and costly due to the fact that the system is buried underground and consists of complex components (Zaman et al., 2015). Municipalities typically spend a substantial portion of their budgets on sewer pipe maintenance due to a large amount of inventory underground. For example, there is around 6,500 km of sewer pipes in the city of Edmonton, Canada, and 120-180 km of sewer pipe are scheduled to be monitored by CCTV each year, which typically costs more than $2 million Canadian dollars per year for the CCTV inspection alone (Navab-Kashani et al., 2015).

In this era of digitization, CCTV inspection technology has become the cornerstone of sewer preventive maintenance programs for thousands of municipalities since not only does this technology prevent the exposure of humans to the issues associated with confined spaces (Navab-Kashani et al., 2015), but more importantly it allows drainage professionals to build, at a

low cost, a digital history for each component of the infrastructure (Duran et al., 2002; Kumar et al., 2018). Clearly, the availability of this digital history makes it possible, at any moment, to conduct data analytics investigations, conduct longitudinal studies, and integrate these data with other information technologies, e.g., automatic defect detection and classification, geographical information systems, visualization, etc. Another important reason for the widespread employment of CCTV inspection is that the inspection process is standardized as per the protocols defined by professional associations, such as Water Research Centre (WRc) (2013) and National Association of Sewer Service Companies (NASSCO) (2015).

In the past few years, several studies have focused on analyzing and improving the productivity of the sewer maintenance process from different perspectives. For instance, Navab-Kashani et al. (2015) investigated how to increase the productivity of sewer CCTV inspection. Specifically, they conducted a time study for the CCTV inspection activities and split the whole process into several sub-processes (e.g., flushing, recording, traveling between manholes, locating manholes, camera installation, etc.). Among all the sub-processes, locating manholes, traveling between manholes, recording, and flushing accounts for the majority of the time in the CCTV inspection process (23%, 11%, 29%, 25% of the total time, respectively) (see Figure 2.1). Therefore, locating manholes and traveling between manholes was selected as the optimization subject in their study, and consequently, the research focuses on the optimization of traveling routes by solving a traveling salesman problem. Another research study by the same authors focused on measuring the CCTV inspection productivity for sewer pipes with respect to weather conditions and temperature to determine their influence on the overall productivity (Navab-Kashani et al. 2019). Zaman et al. (2015) developed a framework to model the process of high pressure flushing activity (which is also responsible for a large portion of the total time required for sewer

14

pipe maintenance activities (see Figure 2.1)) by means of linear regression and ordered probit analysis. As a result, they proposed an efficient algorithm to optimize the high pressure flushing schedule, which led to an increase in productivity of 20% (Zaman et al., 2017). Agbulos et al. (2006) applied the lean concepts and simulation analysis to improve the low pressure flushing and catch basin cleaning process: 10% and 4% efficiency improvements were realized through their research. With regards to CCTV inspection, it turns out that although video recording takes up 25% of the total time of CCTV inspection (Navab-Kashani et al., 2015), it has not been thoroughly investigated yet (see Figure 2.1). Therefore, there is a need to model the recording process in order to predict the duration of video recording process time required for the CCTV inspection of existing sewer pipes.
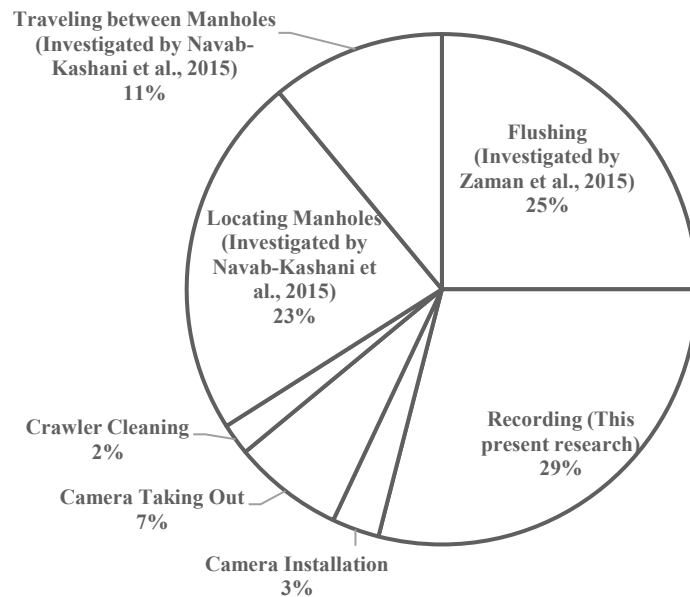
Figure 2.1: The time distribution for the CCTV inspection, and summary of previous studies. (Adopted and modified from Navab-Kashani et al., (2015)).

The model can then be used for scheduling the CCTV inspection as well as the CCTV video assessment process. In fact, since CCTV videos recorded onsite are usually sent to an offsite office where offsite-technologists (well-trained individuals) classify and annotate every defect, the duration of the assessment operation in the best-case scenario, i.e. no defect present in the

footage, will be at least equal to the length of the videos. In other words, under ideal circumstances (assuming there are no defects and no interruption during the video assessment process), a 20-minute video will take at least 20 minutes to be annotated. Therefore, an accurate model to predict the time required to record the video (and, by extension, the lengths of the recorded videos) could not only help for scheduling the data collection but also can give managers a lower bound of the resources that will be needed during the assessment phase. The CCTV recording process will be discussed in the next section. The objective of this research is to model the CCTV recording process for sewer pipes. The results can be used for CCTV collection process scheduling and CCTV video assessment process scheduling. In addition, this research can also extract the standard CCTV video recording process duration (for a video with a regular recording speed, and with no occurrences of waiting time), which can serve as a benchmark for assessing the CCTV recording process. For example, CCTV collection time within a newly developed neighborhood should be aligned with the standard CCTV collection time; while it could be slower in an older neighborhood, theoretically, considering the deterioration of the pipe, since more defects may cause extra time spend on inspecting these defects.

The logic of the methodology for the research follows the procedure of trials and errors. With the ultimate objective of developing a simulation model for the CCTV recording process, the research starts with the identification of the influencing factors through conventional regression analysis. If the preliminary regression model cannot pass the performance test, further operations need to be performed to improve the performance of the regression analysis (data segmentation procedure by machine learning algorithm is used in this research). After that, the simulation model will be developed based on data segmentation and regression analysis. Finally, the model will go through a rigorous validation process to prove the effectiveness of the simulation model.

2.2 CCTV recording process

The CCTV inspection process contains several sub-processes, namely, locating manholes and transportation, upstream manhole setup, downstream manhole setup, flushing, camera installation, recording, additional flushing, camera removal, and cleaning crawler. As indicated in the literature, the CCTV recording process is one of the more time-consuming parts of the entire CCTV inspection process, which represents approximately 25% of the total time required to complete the CCTV inspection process (Navab-Kashani et al., 2015). The CCTV operator usually travels to an assigned location to perform the CCTV inspection as per the schedule set up in advance, unless there is an emergency task that could interfere with the pre-defined schedule. During transportation, the operation crew locates the manhole in order to gain access to the targeted sewer pipe. After locating the targeted manhole, the flushing operation is usually required before the CCTV recording can start in order to provide an acceptable data collection environment by cleaning the pipe and eliminating deposits and obstacles that could hinder the movement of the CCTV crawler (Zaman et al., 2015). Following flushing activities, operators begin set-up work for CCTV data collection on the ground. The setup includes the adjustment of a remotely controlled crawler equipped with a specialized television camera. CCTV recording is the following process, which requires the operators on the ground to follow the standard procedure (e.g. pipeline assessment certification program (PACP) code (NASSCO, 2015)) in order to obtain high-quality information that can be assessed in the later assessment stage. After watching a large quantity of CCTV footage showing the sewer pipes of Edmonton and also talking with the professionals from our research partner at EPCOR Drainage Services, the recording patterns are summarized as follows: Once the recording starts, the CCTV crawler will travel smoothly (at a constant speed (less than 9 m/min according to PACP)) along the pipe.

When there is a tap or severe defect (e.g. broken and hole), the onsite technologist who controls the crawler will stop the crawler and perform a careful inspection of it. For example, the tap recording process is presented in Figure 2.2. As we can see from Part a of Figure 2.2, the crawler is traveling along the pipe at a constant speed, and there are two taps (Tap A and Tap B) that appear on the monitor screen in front of the onsite technologist. Tap A is closer to the crawler, therefore, as the crawler travels to the location of Tap A, the onsite technologist will slow down the crawler to travel closer to the tap, and then stop the crawler when the crawler reaches the location beside the tap (as we can see from part b of Figure 2.2, Tap A is at 11 o'clock with respect to the crawler). Then, the onsite technologist turns the camera to Tap A and performs a detailed inspection of it to check if there are any defects (as we can see from Part c of Figure 2.2, the condition of Tap A is recorded). After the inspection of Tap A, the crawler will accelerate until it reaches a certain speed that is controlled by the onsite technologist, and then continues to travel along the pipe for the inspection of Tap B. For severe defects such as broken and hole, the crawler will operate according to a similar pattern as detailed for the aforementioned tap inspection. For minor defects, such as crack, the crawler will pass them directly without stopping. It should be noted that, for sewer pipe CCTV inspection, the crawler may stop at any defect even for minor defects (e.g. crack) in other cities that follow a different recording standards (NASSCO, 2014). However, the video footage we have that was recorded in the city of Edmonton follows the pattern described above.

| a. Crawler is traveling along the pipe | b. Crawler stops besides the Tap A | a. Camera turns to Tap A and performs an inspection |
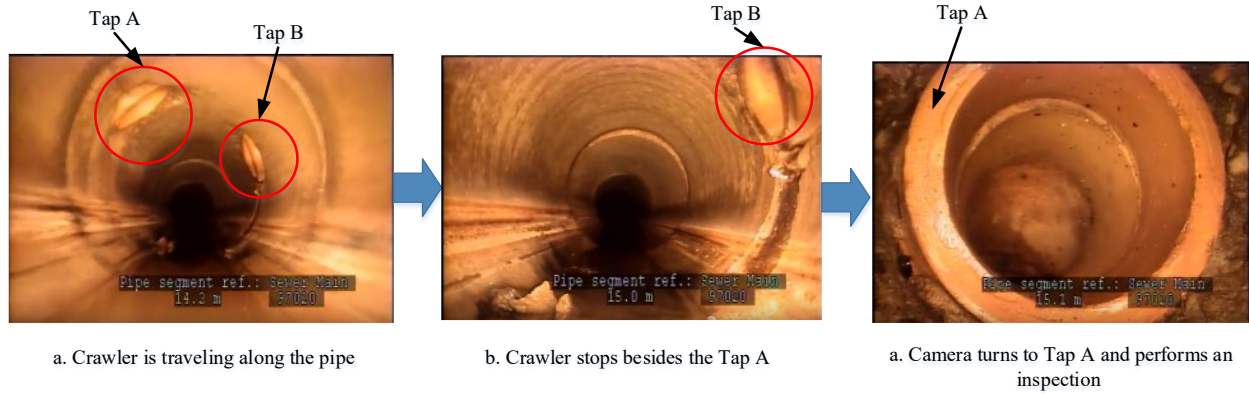
Figure 2.2: Tap recording process.

In general, the total CCTV recording duration contains three parts: time for the crawler to travel along the pipe, waiting time for tap inspection, and waiting time for inspection of severe defects or unpredictable situations (e.g. machinery breakdown). The crawler needs to travel through the pipe, thus, the length of the pipe can influence the total recording time. Every tap in the sewer pipe needs to be inspected, which will lead to some time spent on each tap. Also, severe defects would also contribute to the collection time since the onsite technologist needs to record the detailed condition of them: this inspection time is referred to here as waiting time. A linear model can be developed to represent the total recording time for a sewer pipe, which can be seen in Equation (2-1).

$$Time_n = a \bullet L_n + b \bullet T_n + W_n + C + \varepsilon_n \tag{2-1}$$

where $Time_n$ represents the total CCTV recording time for pipe segment, n; L_n is the length of the pipe segment (structural information known in advance); T_n is the number of taps in this pipe segment (structural information known in advance); W_n is the total waiting time, which may include the time for inspection of severe defects or unpredictable situations (random number, unknown beforehand). C is the fixed time for the CCTV collection process, which may include time for equipment adjustment and other routine processes (e.g. sometimes after the controller clicks the recording button, he still needs to adjust the angle of the camera or other settings

19

before the crawler starts to move, or sometimes the camera starts recording before the equipment is sent into the pipe); and ε_n is an error representing the uncertainties regarding this process, which has a mathematical expectation of 0. The parameters a and b have practical meanings in Equation (2-1), specifically, a is the average time spent to travel along each meter in sewer pipes, by which the average travel speed can be calculated, and b is the average time spent on each tap.

From a theoretical viewpoint, if there are no severe defects or other incidents that hinder the CCTV collection process observed in one pipe, no waiting time will occur in the recording process. If during a recording process, the onsite technologist controls the crawler at a constant speed that aligns with the inspection code, this process will be defined as a standard CCTV recording process. The standard CCTV recording time can be represented by Equation (2-2),

$$Time_n = a \bullet L_n + b \bullet T_n + C + \varepsilon_n \qquad (2\text{-}2）$$

The waiting time is eliminated in this equation. For all the CCTV recording processes, if one process can be expressed by Equation (2-2), then it should be a standard recording process, which means there is no severe defects or unpredictable situations causing waiting time. The traveling speed extracted from the standard process can be compared to the recommend traveling speed (9 m/min) to determine that if the crawler's traveling speed is too fast or too slow. The standard recording process can be extracted by a machine learning algorithm called the Random Sample Consensus (RANSAC) algorithm, which will be described in Section 2.3.2.

2.3 Methodology

The framework for the CCTV recording process productivity analysis and the development of the simulation can be described in four basic steps, namely data collection, linear regression analysis, benchmark extraction, and simulation model development, which correspond to Parts A, B, C, and D, respectively, shown below in Figure 2.3, where techniques utilized in each step are

listed in the oval frame. The objective of linear regression analysis is to extract the factors that delay or accelerate the collection speed and quantify this as well. Then, a machine learning algorithm called the RANSAC algorithm is employed to extract the standard CCTV collection time, which was defined in Section 2.2. The original dataset is split into three parts after Part C: standard data; slower data, which is the process having a recording time greater than the recording time in the standard situation; and faster data, which is the process having a recording time less than the recording time in the standard situation. Successively, two linear regression models are built regarding these two parts of the data (slower data and faster data). Finally, with the impact factors and the benchmark identified in the former steps, a discrete event simulation model is developed based on Simphony.Net (AbouRizk & Mohamed, 2000). The following sub-sections will describe the methodology in more detail.
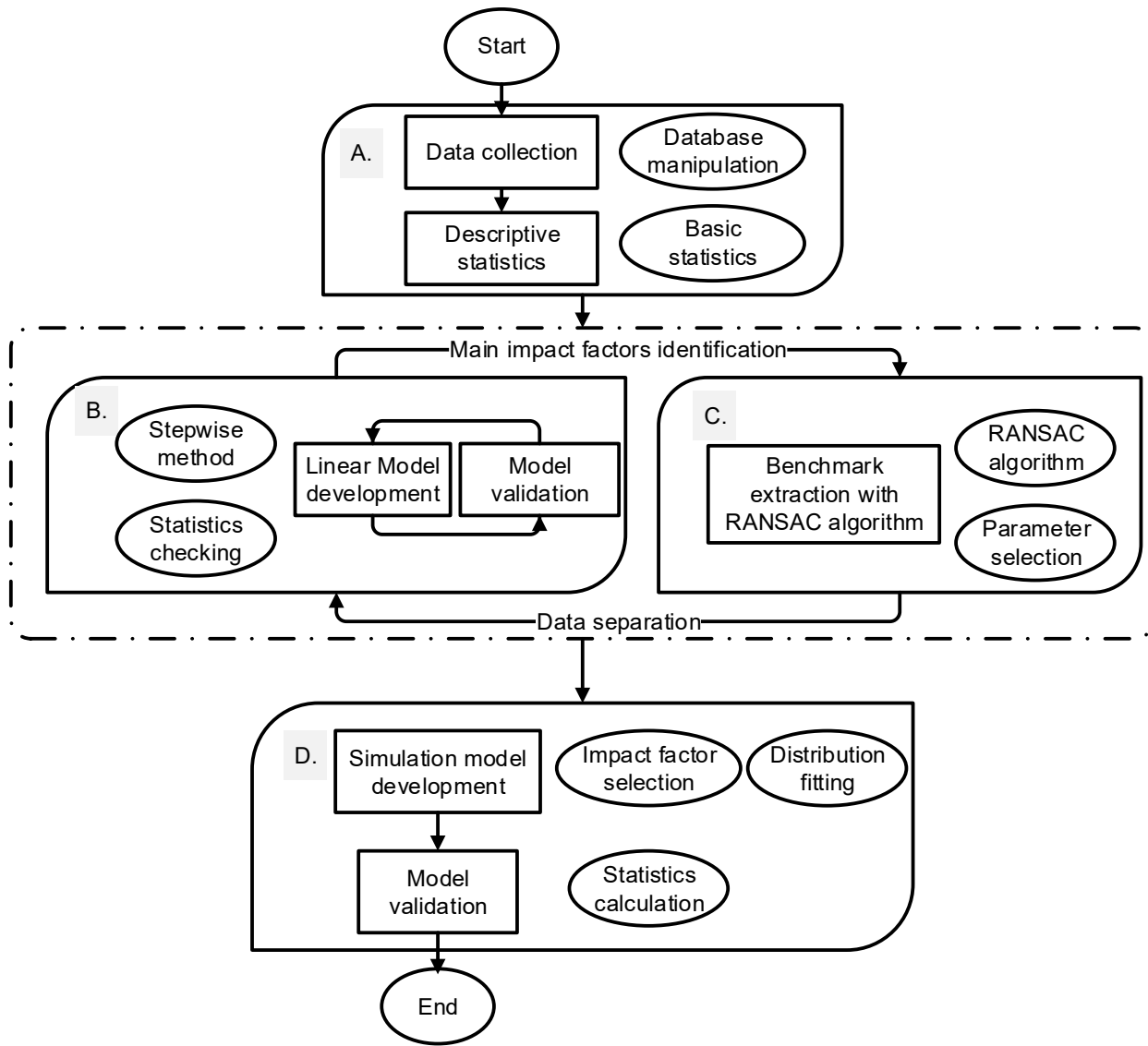
Figure 2.3: Overview of the Framework.

2.3.1 Productivity analysis

The CCTV recording process has been chosen as the focus of this research. The CCTV video recording process starts from the time video acquisition starts and ends with the turning off of the video recording equipment. The time period of the total video recording process duration serves as the CCTV recording duration in this research. The productivity of this process can be represented by Equation (2-3).

$$P = \frac{Total\_pipe\_length}{Total\_video\_duration} \qquad (2\text{-}3)$$

Where P is the overall productivity of the CCTV recording process. For scheduled inspections, the total pipe length in Equation (2-3) is known beforehand and is retrieved from a database where the structural properties of the pipes belonging to the municipal drainage system are stored. The video recording duration, however, is an unknown variable that varies based on the structural properties of the pipes as well as the severity of their deterioration. Predicting this duration constitutes the subject of this research. The total duration of the video recording process can be calculated as shown in Equation (2-4).

$$Total\_video\_duration = \sum_{i=1}^{n} T_i \qquad (2\text{-}4)$$

where T_i is the duration of video i. In order to identify the impact factor of T_i, a multiple regression model is developed to figure out the relationship between T_i and several independent variables. The multiple regression model development follows a common procedure that is utilized in other research articles (e.g. Zaman et al., 2015). The procedure contains five parts, which are variable selection, linear model construct, parameter estimation, performance checking, and model interpretation. For a CCTV recording process, the basic linear model to predict the video recording process duration can be constructed as per Equation (2-5).

$$T_i = \beta_0 + \sum_{j=1}^{m} \beta_{ij} X_{ij} + \varepsilon_i \qquad (2\text{-}5)$$

where β_0 is the constant term, β_{ij} is the parameters for independent predictors X_{ij}, m is the number of predicted variables, and ε_i is the error term.

The stepwise regression method, which is an automatic procedure to choose the appropriate independent variables for fitting the multiple regression model (Hocking, 1976), is used to

23

predict the duration of the CCTV recording process in this research. Specifically, the bidirectional elimination method, which is a method to measure whether a variable needs to be included or excluded recursively at each step (Liao et al., 2008; Wang et al., 2013), is used in the stepwise regression analysis. It can benefit the regression analysis by managing a large number of variables efficiently and eliminating the multicollinearity of the independent variables (Wang et al., 2013). The detailed procedure of multiple linear regression model development will be described in Section 2.4.2.

2.3.2 Benchmark extraction

2.3.2.1 The RANSAC algorithm

In order to extract the benchmark of CCTV recording time (which is the standard CCTV recording time as defined earlier), this research applies the RANSAC algorithm to automatically extract targeted data. RANSAC, proposed by Fischler & Bolles in 1981, is a non-deterministic machine-learning approach that uses a repeated procedure to fit an objective function that has the best performance on the targeted subset of the original dataset (Fischler & Bolles, 1981). Unlike linear regression using least squares estimation, which seeks to minimize the distance from all the data points to the fitted function, the RANSAC model seeks to extract the largest possible subset from the initial dataset that best fits a pre-defined parametric function thus segmenting the data into a set of inliers, i.e. points fitting the model, and a set of outliers, which are not likely to be described by the pre-defined function (see Figure 2.4). In Figure 2.4, the blue points with a red marker are the inliers, while the others are outliers in the RANSAC algorithm model. The two dashed lines that separate the inliers and outliers are the threshold (upper threshold and lower threshold) for the algorithm. The red line is the linear model constructed by the RANSAC algorithm and the blue line is the linear model developed by least squares linear regression

analysis. Note that the RANSAC model cuts the 2D space into three parts, which are benchmark space, upper space, and lower space, as indicated in Figure 2.4. It can be seen that the red line constructed by the RANSAC algorithm performs better in describing the trend for a high density of points than the blue line developed by the least-square estimation. Outliers, which collectively constitute noisy data, e.g. contaminated or potentially incorrect, are identified and excluded from the fitting step by the RANSAC procedure, thus leading to a regression model that better captures the main trend in the data.
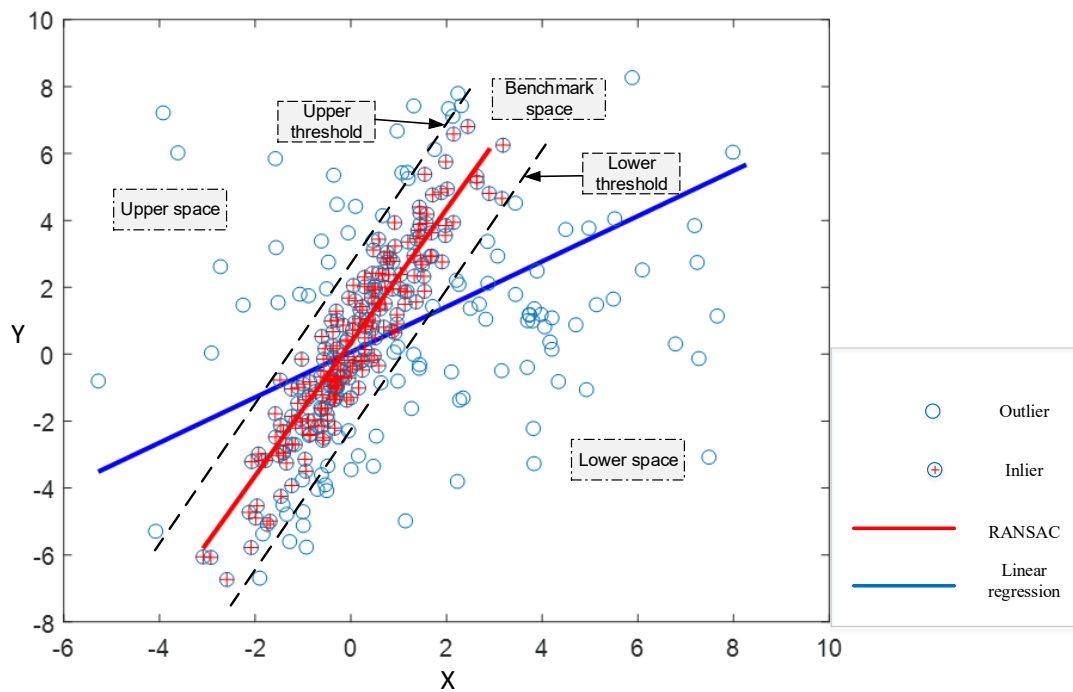
Figure 2.4: Comparison of least squares linear regression and RANSAC algorithm.

The mechanism of the RANSAC algorithm is presented in Figure 2.5 (Yin et al., 2019). The algorithm starts with the selection of X points (where X= number of independent variables + 1). For instance, two points are needed to fit a line (2-dimensional problem) while three points are needed to fit a plane (3-dimensional problem). From Equation (2-2), we know that the extraction of the standard CCTV video recording process time is a problem with two independent variables,

namely, the length of the pipe segment associated with the video (L), and the number of taps within the pipe segment (T). Therefore, three points need to be selected for each iteration. An objective function (Y_i) can be formed based on the selected points. Then, the Euclidean distance of each point to the objective function is calculated. A threshold (t) should be selected in order to decide whether the point is an inlier or an outlier. If the distance is within the threshold, it is an inlier; otherwise, it is an outlier. The number of inliers (N_i) should be counted and compared with the N_{i-1} (i.e., the greatest number of inliers among all the historical iterations). If the N_{i-1} is bigger, we save the N_{i-1} as N_i, and Y_{i-1} as Y_i. Otherwise, we update N_i and Y_i accordingly. The process is repeated until the predetermined number of iterations (M) is reached. As for the number of iterations (M), it can be calculated by means of Equation (2-6) (Altaf et al. 2018).
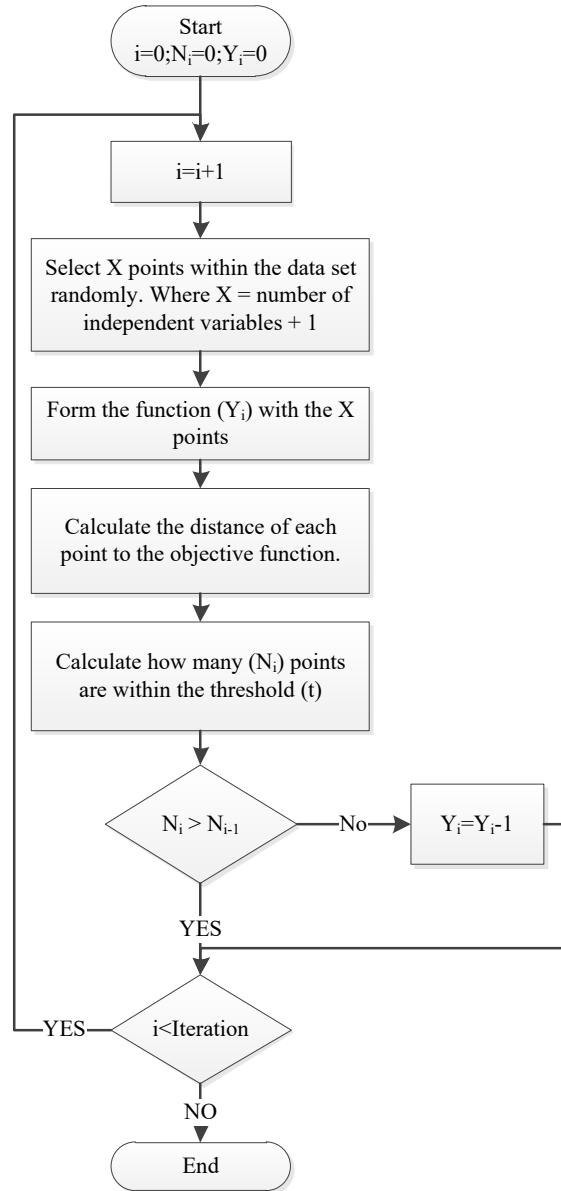
Figure 2.5: The mechanism of the RANSAC algorithm.

$$1-p = (1-\lambda^X)^M \Rightarrow M \geq \frac{\log(1-P)}{\log(1-\lambda^X)} \qquad (2\text{-}6）$$

where X is the number of points needed to construct the objective function; M is the number of

iterations; P is the probability that at least one of the objective functions built in all M iterations

is constructed by X inliers; and λ is the inlier ratio, which can be calculated using Equation (2-7).

Although λ is a priori unknown ratio, it can be updated during the algorithm progress (Raguram et al. 2008).

$$\lambda = \frac{Inliers}{Inliers + Outliers} \qquad (2\text{-}7)$$

Considering the probabilistic nature of the RANSAC algorithm, the parameters (e.g., *a* and *b* in Equation (2-2)) of the objective function will vary from one run of the algorithm to the next. Therefore, the probability distribution of each parameter can be calculated from a large number of run-times of the algorithm. The distributions of the parameters will be used to form the objective function for the convenience of constructing the simulation model.

2.3.2.2 Data separation

The RANSAC algorithm could classify the data into inliers and outliers. The inliers are represented for the standard CCTV recording process, which can be described by the RANSAC algorithm with a linear model, while the outliers are the noise data. Therefore, a model should be developed to describe the outliers in order to simulate the overall behavior of the data. In this research, we separate the outliers into two parts: data of slower recording process (which has a recording time longer than the standard process), and data of faster recording process (which has a recording time shorter than the standard process). As a result, the whole dataset is separated into three parts: standard process, slower process, and faster process. For the standard process, RANSAC could provide a linear model to predict the recording duration. For the faster and slower process, two multiple linear regression models are developed to describe the behavior of these two parts of the data. Finally, these three linear models will be fed into a simulation model to simulate the CCTV recording process.

2.3.3 Simulation model development and validation

Discrete event simulation (DES) is employed in this research to simulate the CCTV recording process and serves as a validation tool for the productivity analysis model. DES has been widely used in process modeling research, for example, Abourizk (2010) employed the DES for modeling the utility tunneling operation using a tunnel boring machine, Liu et al. (2015) used the DES to model the on-site construction process of a panelized building project, Taghaddos et al. (2014) developed an integrated DES for scheduling the construction of modules in an assembly yard. Specifically, the drainage maintenance process is simulated by Agbulos et al. (2006) integrated with lean concepts. The DES in this research is carried out on a simulation system called *Simphony.Net* developed by researchers at the University of Alberta (AbouRizk & Mohamed, 2000).

2.3.3.1 Input modeling

The input of the DES model is the product of the simulation process, which is the sewer pipe CCTV recording process in this research, represented by the concept of the entity in the simulation model. The variables used in the simulation model are determined in productivity analysis (part B in Figure 2.3) and benchmark extraction (part C in Figure 2.3). Two types of variables are included in the simulation model, which are scale variables (e.g., length of pipe, number of taps) and nominal variables (e.g., functions of the pipe (sanitary, storm, or combined), season condition (winter or other seasons)). For scale variables, distribution fitting techniques are utilized to simulate the behavior of parameters. For example, the length of each pipe is a significant parameter in the simulation model, thus, the exact length of each pipe is assigned randomly according to a realistic mathematical distribution retrieved from the historical data. Other scale variables, such as the number of taps, the height of the pipe, etc. are modeled

according to a similar method. For nominal variables, probabilities of each category are calculated and fed into the simulation model. For instance, based on our historical data, the probabilities of each type of sewer pipe in terms of materials are 46%, 22%, and 32% for clay, PVC, and concrete, respectively. The final simulation model integrates all the variables in a logical manner, which for this research are presented in Section 2.4.3.1.

2.3.3.2 Output analysis

The CCTV recording process duration is the focus of this research and also its main output. Each CCTV recording process has inherently different variables and will have a different recording duration. For a large number of recording instances, the CCTV recording duration should follow a similar pattern as that found in a real-life situation (retrieved from historical data). The final duration pattern will be compared in pairs (linear regression model output versus real-life situation, RANSAC model output versus real-life situation, and simulation output versus real-life situation). The output analysis will provide visual and statistical validation to prove the validity of the proposed framework.

2.4 Case study

The proposed framework is implemented and validated with the data retrieved from a drainage system in Edmonton, Canada. The framework development will follow the process described in Figure 2.3. Data collection serves as the foundation of the model development, and the regression analysis and RANSAC algorithm will be conducted in succession. Finally, a simulation model will be formed based on the regression model. The output of the simulation model will be compared with the real-life situation and other regression techniques to prove its performance. Data collection, regression model and simulation model development are described in the following sections.

30

2.4.1 Data collection

The data was collected from 1,557 pipe segments spread over 100 km in total length. Predicted variables are retrieved from the sewer pipe inspection database, which is maintained by the EPCOR Drainage Services located in Edmonton, Canada. The variables can be categorized into two classes, namely, scale variable and nominal variable. Note that nominal variables are converted to dummy variables for the purpose of feeding into the regression analysis. For example, for the function variable, the nominal value of sanitary (one value of function variable) has a numerical value of 0 and 1 (1 means that the pipe function is sanitary, otherwise the function is not sanitary). Other nominal variables will be converted according to this pattern.

The predicted variable is the video recording process duration, which has a wide range from 2 seconds to 1747 seconds. Any video durations of extremely small values (e.g. 2 seconds) is considered to be a failed recording due to reasons such as equipment failure. These data are included in the model development process since the failure also needs to be simulated in order to get a practical model to simulate the CCTV recording process. There are other variables besides that have been summarized in Table 2.1. However, a selection process is performed to filter some unrelated variables (such as the shape of the pipe) manually by checking the scatter plot and the trend line. The scatter plots of the scale variables in Table 2.1 are partially shown, with respect to the recording duration, in Figure 2.6. It is shown that the recording duration has a linear correlation with the length of the pipe, the number of the taps, and the number of minor defects. However, the relationship between the age of the pipe and the recording duration seems to be weak. Despite that, the variable of age is still used for the regression analysis and whether this variable should be used in the final model will be further checked by the stepwise regression method. Some variables are excluded, such as the drainage community (the community that the

pipe service for), due to the complexity of the variable itself, thus, a more comprehensive model

may be needed in the future to include the geographical variable. The multiple regression model

and simulation model proposed in this research are only targeting the variables listed in Table 2.1.

Table 2.1: Summary of variables in the modeling dataset.

Variable name	Description	Range	Variable type
Season	The season when the inspection performed	Winter or Non-winter	Nominal
Time of the day	The time when the inspection performed	Day, Night	Nominal
Function	The function of the sewer pipe	Sanitary, Stormwater, Combined	Nominal
Material	The type of materials of sewer pipes	Clay, Concrete, Reinforced concrete, PVC, Other	Nominal
Purpose of inspection	The purpose of the inspection	Maintenance, Routine assessment, Other	Nominal
Pre-cleaning method	The pre-cleaning method before the inspection	Jetting, Heavy-cleaning, No pre-cleaning	Nominal
Weather	The weather condition during the inspection	Dry, Rain, Snow, Other	Nominal
Road condition	The road condition of the sewer pipe	Alley, Light Highway, Other	Nominal
Pipe joint length	The joint length of the inspected sewer pipe segment	0.6-7.5 m	Scale
Total length	The total length of the inspected sewer pipe segment	1.3-1013 m	Scale
Age	The age of the inspected sewer pipe segment	2-111 years	Scale
Height	The height of the sewer pipe	150-3000 mm	Scale
Number of taps	The number of taps of the inspected sewer pipe segment	0-33	Scale
Minor defects	The number of minor defects of the inspected sewer pipe segment	0-94	Scale
Major defects	The number of major defects of the inspected sewer pipe segment	0-6	Scale
Recording duration	The video duration of the CCTV video	2-1747 seconds	Scale
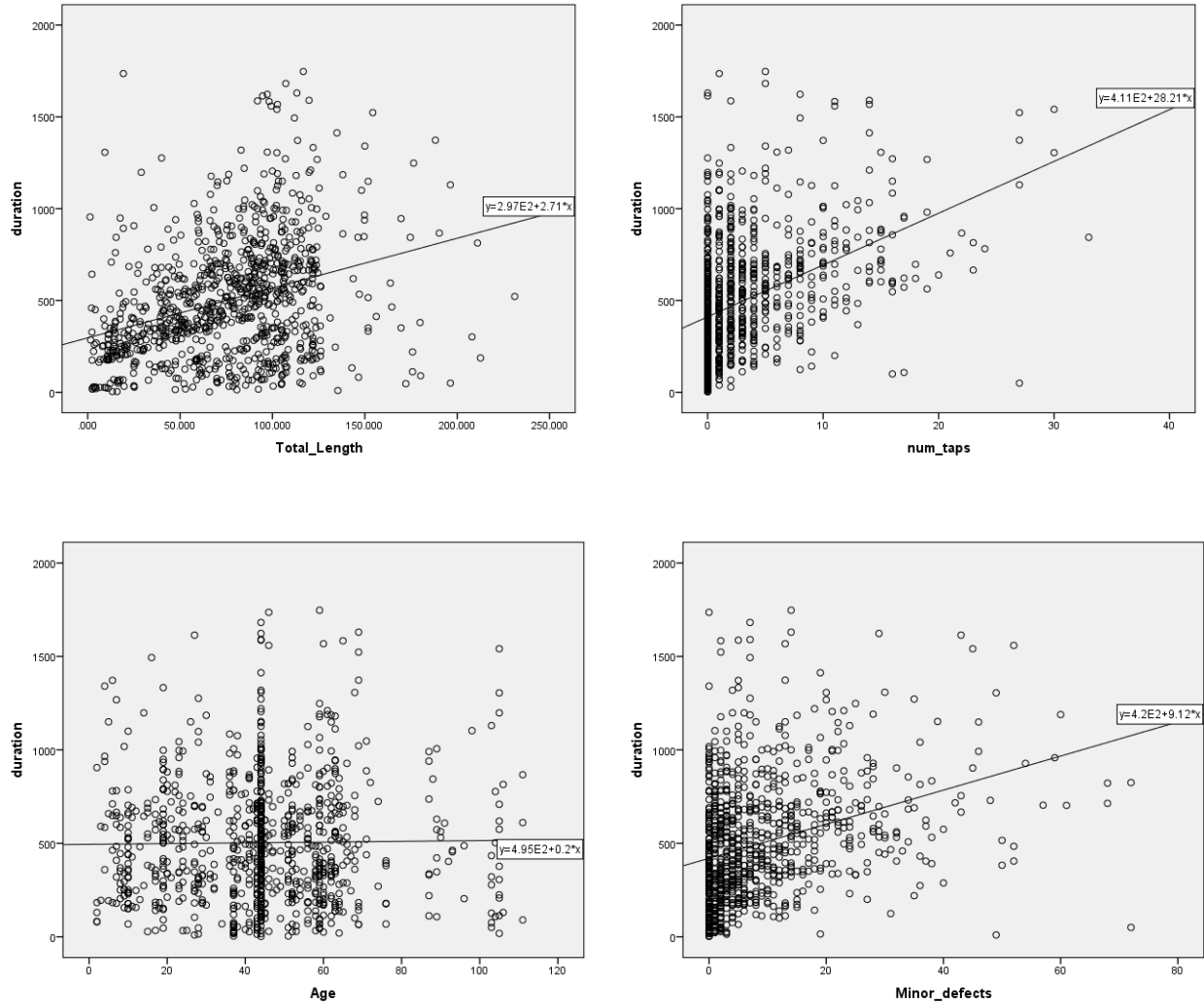
Figure 2.6: Scatter plot of some predicted variables to the recording duration.

2.4.2 Regression model

2.4.2.1 Preliminary linear regression

The stepwise linear regression model is developed by *SPSS* statistical software. Variables in Table 2.1 (except for the recording duration variable) are the input of the regression model, and recording duration is the dependent variable that will be predicted by the selected independent variables. Eight iterations are processed before we get the final results. The model summary of the regression analysis is presented in Table 2.2. In the final model, seven independent variables

are included in the regression model, which are the number of taps, minor defects, function-sanitary, total length, weather-snow, age, alley, and material-other. R square and adjusted R square are two key indicators to assess the performance of regression analysis, which has the perfect number of 1 and the worst number of 0 (Chughtai & Zayed, 2007). It is shown in Table 2.2 that the linear regression exhibits unacceptable performance in predicting the recording duration, with an R square of 0.305 and an adjusted R square of 0.298, which means that the predictors included in Model 8 can only explain around 30% of the variation in the recording duration. Table 2.3 shows the coefficient information of Model 8 in the final step of the stepwise regression analysis. The poor performance indicates that it is impracticable to use regression analysis to predict the recording duration directly, which leads us to the next step of benchmark extraction.

Table 2. 2: Model summary of the stepwise linear regression.

Model	R	R Square	Adjusted R Square	Standard Error of the Estimate
1	0.405[a]	0.164	0.163	283.468
2	0.472[b]	0.223	0.221	273.432
3	0.503[c]	0.253	0.250	268.344
4	0.522[d]	0.272	0.269	264.973
5	0.538[e]	0.289	0.285	262.055
6	0.544[f]	0.296	0.291	260.877
7	0.549[g]	0.301	0.295	260.172
8	0.552[h]	0.305	0.298	259.577

a. Predictors: (Constant), Number of taps
b. Predictors: (Constant), Number of taps, Minor defects
c. Predictors: (Constant), Number of taps, Minor defects, Function-Sanitary
d. Predictors: (Constant), Number of taps, Minor defects, Function-Sanitary, Total length
e. Predictors: (Constant), Number of taps, Minor defects, Function-Sanitary, Total length, Weather-Snow
f. Predictors: (Constant), Number of taps, Minor defects, Function-Sanitary, Total length, Weather-Snow, Age
g. Predictors: (Constant), Number of taps, Minor defects, Function-Sanitary, Total length, Weather-Snow, Age, Road condition-Alley
h. Predictors: (Constant), Number of taps, Minor defects, Function-Sanitary, Total length, Weather-Snow, Age, Road condition-Alley, Material-Other

Table 2.3: Regression results.

Variable	Coefficient	Standard eror	t-stat	p-value
(Constant)	311.877	30.419	10.253	0.000
Num of taps	21.779	2.007	10.849	0.000
Minor defects	6.735	0.838	8.035	0.000
Function: Sanitary	99.716	19.439	5.130	0.000
Total length	0.909	0.195	4.654	0.000
Weather: Snow	250.860	63.722	3.937	0.000
Age	-1.269	0.547	-2.320	0.021
Road condition: Alley	-80.221	34.519	-2.324	0.020
Material: Other	-142.300	65.888	-2.160	0.031

2.4.2.2 Benchmark extraction

As described in the methodology section, the RANSAC algorithm is employed to extract the standard CCTV recording process (which is a recording process for a pipe with no defects and no unpredicted situations occurring during the recording process, and the onsite-technologist controls the crawler in such a way that it travels at a constant speed). Inspection data of 1,557 pipe segments serve as the input of the RANSAC algorithm. Several parameters need to be calculated before running the algorithm. From Equation (2-2) we know that it is a 3D problem, with 2 independent variables and 1 dependent variable. Therefore, the number of initial points that need to be selected is 3 (see Figure 2.5). In order to determine the number of iterations (M) needed, we assume the inlier ratio λ as 0.5 and the probability (P), which is at least one of the objective functions built in all M iterations is constructed by 3 inliers (no outliers), as 99%. The minimum number of iterations can be calculated from Equation (2-6) as 35. Since the inlier ratio is unknown and assumed by the author, we can increase the number of iterations (10,000 iterations used in this research) to ensure that we have a higher probability to get the near-optimum solution. In this research, we use 10,000 iterations to run the algorithm considering the size of the dataset and the CPU processing time.

Another parameter that needs to be selected is the threshold (t). Four preliminary experiments have been performed, and the results are shown in Figure 2.7. For each part of Figure 2.7, the black points represent the inspection data for the 1,557 pipe segments, the black points with red marks are the inliers, otherwise, the other black points are the outliers. Note that all the inliers are classified as the benchmark for the sewer pipe CCTV recording process (standard recording process). The plane is the objective function (benchmark) derived from Equation (2-2) after running the algorithm. The thresholds for each of the four experiments are set as 30, 60, 90, and 120 seconds, respectively, which means for the inliers (points within the threshold), the variance between the data and the benchmark is smaller than the threshold. For example, in part a of Figure 2.7, for a sewer pipe with a length of 10m and 2 taps, the benchmark for the recording process of this pipe should be $2.6389 \times 10 + 37.7708 \times 2 + 67.9861 = 169.9 \, s$. Since the threshold is 30 s in this scenario, if the real recording duration is within the range $(139.9 \, s :$ $199.9 \, s)$, then it will be classified as an inlier by the algorithm, otherwise, it will be classified as an outlier. From Figure 2.7, it is clear that as the threshold increases, the number of inliers (red points) increases as well. In addition, the objective function is different as we choose a different threshold. Similar to the space cutting described in Figure 2.4, other information obtained from Figure 2.7 is that the plane (objective function) together with the threshold cuts the 3D space into 3 parts, which are benchmark space (around the plane within the threshold), upper space (upper of the upper threshold), lower space (lower of the lower threshold). For points in the upper and lower space, two separate multiple linear regression analyses are conducted to predict the recording duration for those pipe segments.

Recording Duration: Time = 2.6389L + 37.7708T + 67.9861

Recording Duration: Time = 2.3326L + 30.9045T + 81.8403
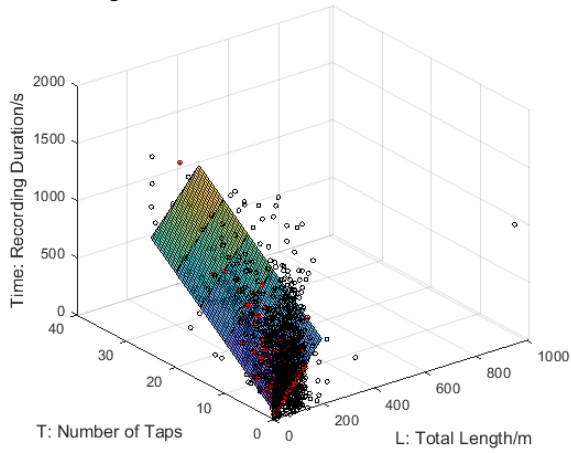
a. t = 30; Inliers = 352; Outliers = 1205
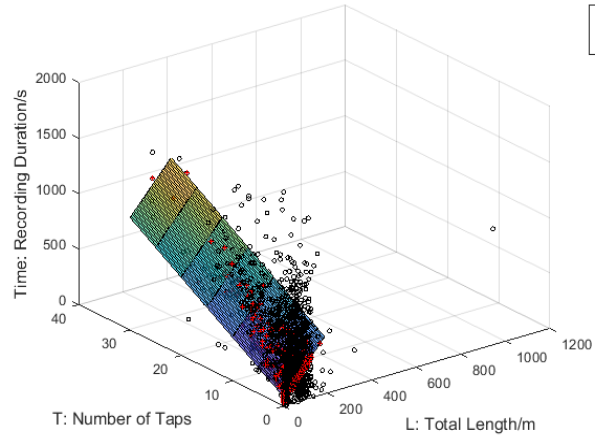
b. t = 60; Inliers =614; Outliers = 943

Recording Duration: Time = 2.2568L + 31.2032T + 118.8829

Recording Duration: Time = 2.4299L + 25.6986T + 125.1612
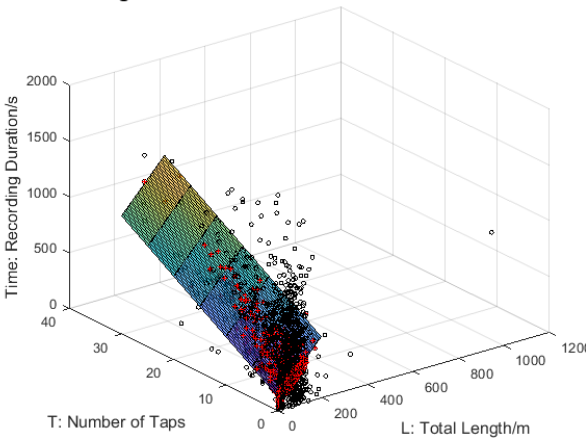
c. t = 90; Inliers = 791; Outliers = 766
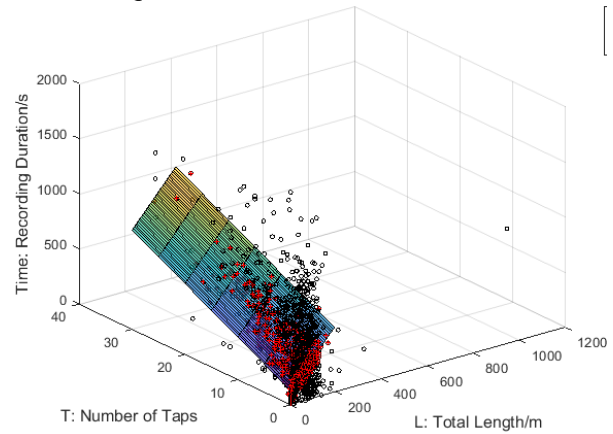
d. t = 120; Inliers = 938; Outliers = 619

Figure 2.7: Sample results of the RANSAC algorithm.

An appropriate threshold can reflect the characteristics of the dataset and it can also separate the dataset to reduce the points near the border (which means the upper and lower threshold, see the dashed line in Figure 2.4). A curve that illustrates the relationship between the threshold and the number of inliers is plotted in Figure 2.8. Note that threshold selection is a subjective process to some extent, therefore, statistical characteristics of the dataset and professional knowledge need to be considered at the same time in this process. As expected, from Figure 2.8 we can see that

the number of inliers grows as the threshold is increased. At thresholds within the range of approximately 50-70 s, the gradient of the curve starts to fall, which means there are fewer points near the border. In addition, considering the inspection results of our dataset, we choose 60 s as the threshold. This number may need to change according to different dataset collection times and different locations. For instance, if the datasets are collected in a city whose drainage system is newly built, the threshold may be smaller as most of the durations for the recording process would be close to the benchmark.
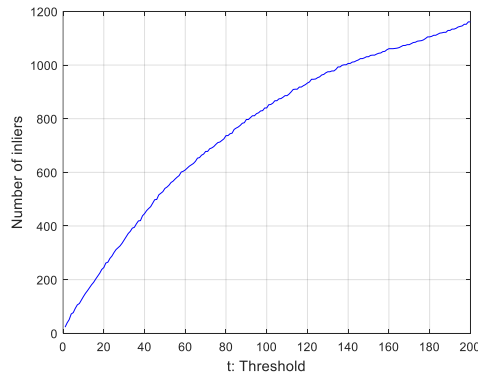
Figure 2.8: Influence of the threshold selection to the number of inliers.

The RANSAC algorithm is a nondeterministic algorithm (Nguyen et al., 2005). As is implied from the mechanism of the RANSAC algorithm, the results of the objective function derived from the algorithm are supposed to be different even for the same dataset. In order to feed the objective function (Equation (2-2)) into the simulation model, the distribution of parameters of the objective function in 500 runs of the algorithm is utilized. Figure 2.9 shows the distributions of each parameter in all the 500 runs, which has the mathematical expectation of E(a)=2.54 s/m, E(b)=28.03 s and E(c)=82.13 s. The distribution fittings are validated by the chi-square goodness of fit test under the significance level of 0.05. Some statistical interpretations can be made based on the mathematical expectation of each parameter of the objective function. For instance, the average speed of the crawler can be calculated as $1 \div 2.54 \times 60 = 23.6$ m/min, which is much

38

faster than the recommended speed in the PACP code (less than or equal to 9 meters per minute). Another finding is that the number of taps has a significant influence on the recording duration (the crawler spends about 28.03 s on each tap which includes the slowdown, inspection and acceleration process). The distribution of each parameter will be the input of the simulation model in the next step.
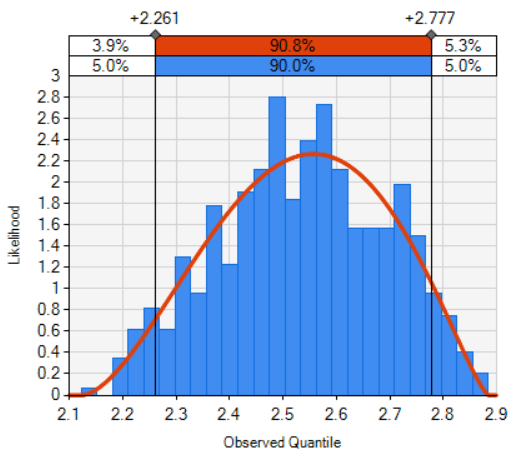
Parameters a: Beta distribution (α=2.759, β=2.326, Min=2.123, Max=2.884, Percentile$_{5th}$=2.261, Median=2.541, Percentile$_{95th}$=2.777)
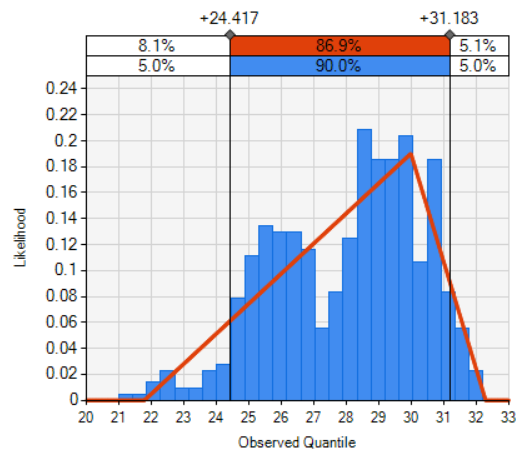
Parameters b: Triangular distribution (Min=21.776, Max=32.290, Mode=29.986, Percentile$_{5th}$=24.417, Median=28.465, Percentile$_{95th}$=31.183)
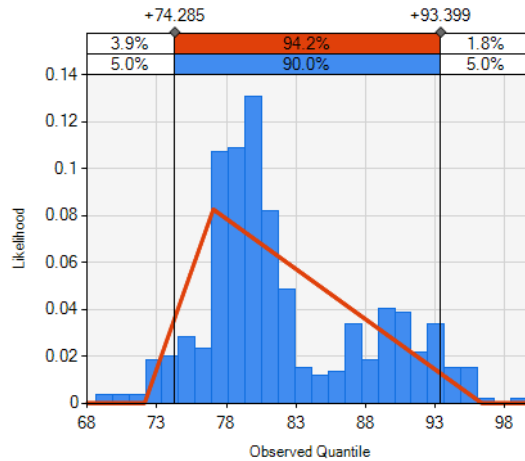
Parameters C: Triangular distribution (Min=72.117, Max=96.332, Mode=77.051, Percentile$_{5th}$=74.285, Median=80.281, Percentile$_{95th}$=93.399)

Figure 2.9 The distribution of each parameter in Equation (2-2) retrieved from the RANSAC algorithm.

2.4.2.3 Outlier linear regression

Points falling into the upper and lower spaces are outliers, and two linear regression analyses are performed for the modeling purpose. The stepwise method is used to select the appropriate variables as well. The input of independent and dependent variables is the same as the preliminary regression analysis described in Section 2.4.2.1. Results of the linear regression of lower space data and upper space data are presented in Table 2.4. The adjusted R square of the two models is 0.712 and 0.517 for data in lower space and upper space, respectively. F-test is performed on these two models with the F-stat of 86.335 and 86.763 for lower and upper data respectively and significance levels of both models are less than 0.05. Therefore, the overall performance is improved compared with the preliminary regression analysis including the whole dataset. The selected variables and their coefficients are listed in Table 2.4, which will be the input of the simulation model. From the p-value in the table, for the confidence level of 0.05, all the selected variables are acceptable, and the null hypothesis (all the coefficients in the model are equal to zero) is rejected. The Durbin-Watson test is conducted on the two regression models for data in the lower and upper space with a value of 1.971 and 2.091, which shows that there is no autocorrelation in residuals since the acceptable range is between 1.5-2.5 (Leung et al., 2000). The normality of the standardized residuals (which is an assumption for linear regression analysis that needs to be satisfied) of the two regression models is illustrated in the histogram and P-P plot in Figure 2.10 and Figure 2.11. Both of the histograms show that the standardized residuals in the statistical regression model satisfy the normal distribution (see part a of Figure 2.10 and part a of Figure 2.11). For the P-P plot of the lower data model (see part a of Figure 2.10), the plot is close to the diagonal, which means that the standardized residuals satisfy the

normal distribution, while the plot of the upper data is slightly worse but still near the diagonal (part b of Figure 2.11). Therefore, the normality of the standardized residuals is proven.

Table 2.4: Linear regression results for data in upper and lower space.

Model	Predictor variable	Coefficient	t-stat	p-value
Data in lower space	Constant	106.187	3.229	0.001
Adjusted R square =	Number of taps	23.113	18.921	0.000
0.712	Total length	0.927	9.275	0.000
	Height	0.124	3.515	0.001
	Function: Sanitary	65.360	3.944	0.000
	Material: Clay	-48.203	-2.883	0.004
	Season: Non winter	-65.291	-2.379	0.018
	Pre-cleaning method: No Pre-cleaning	-41.016	-2.312	0.022
Data in upper space	Constant	119.001	3.457	0.001
Adjusted R square =	Total length	3.908	14.187	0.000
0.517	Number of taps	25.170	11.382	0.000
	Minor defects	2.851	3.578	0.000
	Time of day: Day	71.426	2.980	0.003
	Material: PVC	-166.603	-4.466	0.000
	Pipe joint length	52.488	3.921	0.000
	Material: Other	-147.111	-2.192	0.029

For the sewer pipe recording process falling in the upper space, the duration is longer than the benchmark, while in the lower space, the duration is shorter than the benchmark. Both of the regression models use variables of total length and number of taps. However, the coefficient is different for the same variable, which may be caused by the differences in recording pattern for sewer pipes in the lower and upper spaces (such as the crawler traveling speed, time spent on each tap, etc.). For instance, according to the regression model, for those recordings falling in the lower space, the average duration of tap inspection was 23.113 s, while the duration is 25.170 s for the upper space data. These two regression models together with the benchmark model will be fed into the simulation model in the next part.
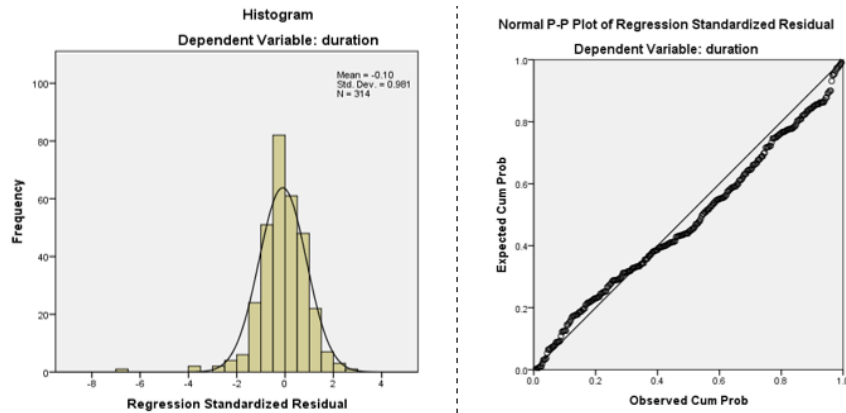
Figure 2.10: Residual analysis for the lower space data model.
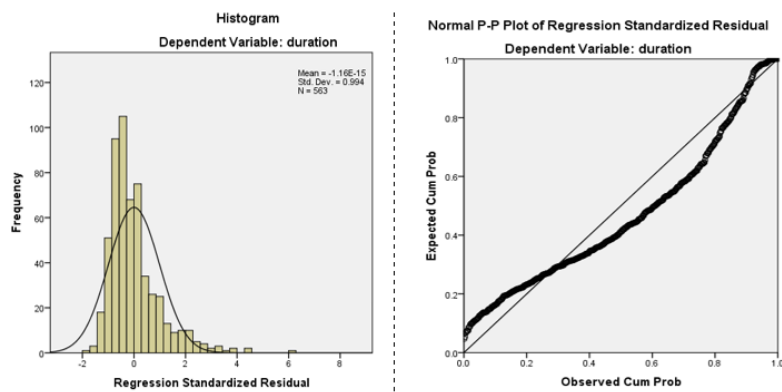
Figure 2.11: Residual analysis for the upper space data model.

2.4.3 Simulation model

The regression model is suitable for scheduling the recording process for a large number of sewer pipes. However, for a single pipe, the prediction may lead to inaccurate results. The regression models have many variables (7 variables for the lower space data, 2 variables for benchmark model, and 7 variables for the upper space data), which makes the implementation of the model complicated. Therefore, a DES model was built on the basis of the regression model to simplify the implementation of the regression model and predict the recording time for a large number of sewer pipes.

2.4.3.1 Model layout and element description

According to the regression analysis, the simulation model was built to mimic the CCTV sewer pipe recording process. The regression analysis contains three parts, as described in Section 2.4.2, which lead to the three-track layout of the simulation model (see Figure 2.12). The *create* element will create entities that represent the sewer pipe segment to be recorded. Along with the process of creating entities, the properties of the pipe segment (e.g., pipe length, the number of the taps) are assigned according to the historical data. To be more specific, the distribution of the total length of the pipe (Part a in Figure 2.13) and the number of taps (Part b in Figure 2.13) is embedded along with the process of creating sewer pipe segment entities. The subsequent *Capture_Crew* element can define the number of crews available for the inspection work, which can be defined by the user based on the real-life situation. In this case study, the number of crews is set to one crew to simplify the comparison process of the linear regression model, the RANSAC model, and the simulation model, which will be further discussed in the output analysis (Section 2.4.3.3).

After creating the entities, all the entities are separated into three parts (upper track, middle track, and lower track) based on the historical data corresponding to the upper space data, benchmark data, and lower space data. The probability of each part can be seen from the element *upper_benchmark_lower* probabilistic branch shown in Figure 2.12. Note that, for practical use of the schedule or for prediction purposes, the probability of each part can be updated with the increase in size of the dataset or can be adjusted considering the real-life situation (for instance, the probability of benchmark data can be higher if the scheduled inspection area is in a newly constructed community, otherwise for an older community, the probability for upper space data

may increase). After that, entities will go through each element on their own track to mimic the recording process with their unique situation.

The simulation of recording time for entities of pipe segments that travel along the middle track is straightforward with only one *execute* element to use the results from the RANSAC model. The distributions for each parameter (a, b and C) are shown in Figure 2.9. The total length and number of taps are created in the first step with the distribution shown in Part a and b of Figure 2.13. The formula executed in the *Benchmark* execute element is shown in Equation (2-8), which is developed based on Equation (2-2).

$$\text{Recording time=Beta}(\alpha=2.759, \beta=2.326, \text{Min}=2.123, \text{Max}=2.884)*\text{Normal}(\mu=63.449, \sigma=5.3919)+$$
$$\text{Triangular(Min}=21.776, \text{Max}=32.290, \text{Mode}=29.986)*\text{Beta}(\alpha=0.238, \beta=2.903, \text{Min}=0, \text{Max}=33)+$$
$$\text{Triangular(Min}=72.117, \text{Max}=96.332, \text{Mode}=77.051) \qquad (2\text{-}8）$$

The entities traveling along the upper track and the lower track will go through the two composite elements separately, with the detailed layout of the sub-simulation model shown in part B and C in Figure 2.12. Entities will go through each element on the basis of the regression model developed above. For instance, entities assigned to the upper track will go through an execute element to calculate the time spent on travel along the pipe, time spent on the taps, and routine set-up time, which is similar to Equation (2-8); however, the parameters will be adjusted according to Table 2.4. After that, the variables that have been identified as influencing the recording time will be considered. For example, for upper space data, minor defects will cause a 2.532 s delay in recording time on average. It should be noted that variables, such as an unknown number of minor defects, are not known before inspection; therefore, historical distributions can be used to generate the data. Such data can be adjusted based on the real-life situation (e.g. pipes in newly built neighborhoods and older neighborhoods may have a different number of minor defects), as mentioned before. After that, the entities will go through each element sequentially.

Sequentially, all the entities will go through *Duration_collection* element, which will collect the recording time for the purpose of statistical analysis. The *Counter 1* element can keep track of the total number of generated entities and ensure that all the entities generated have traveled through the simulation model, and the simulation process can be terminated by destroying all the entities with the *Destroy 1* element.
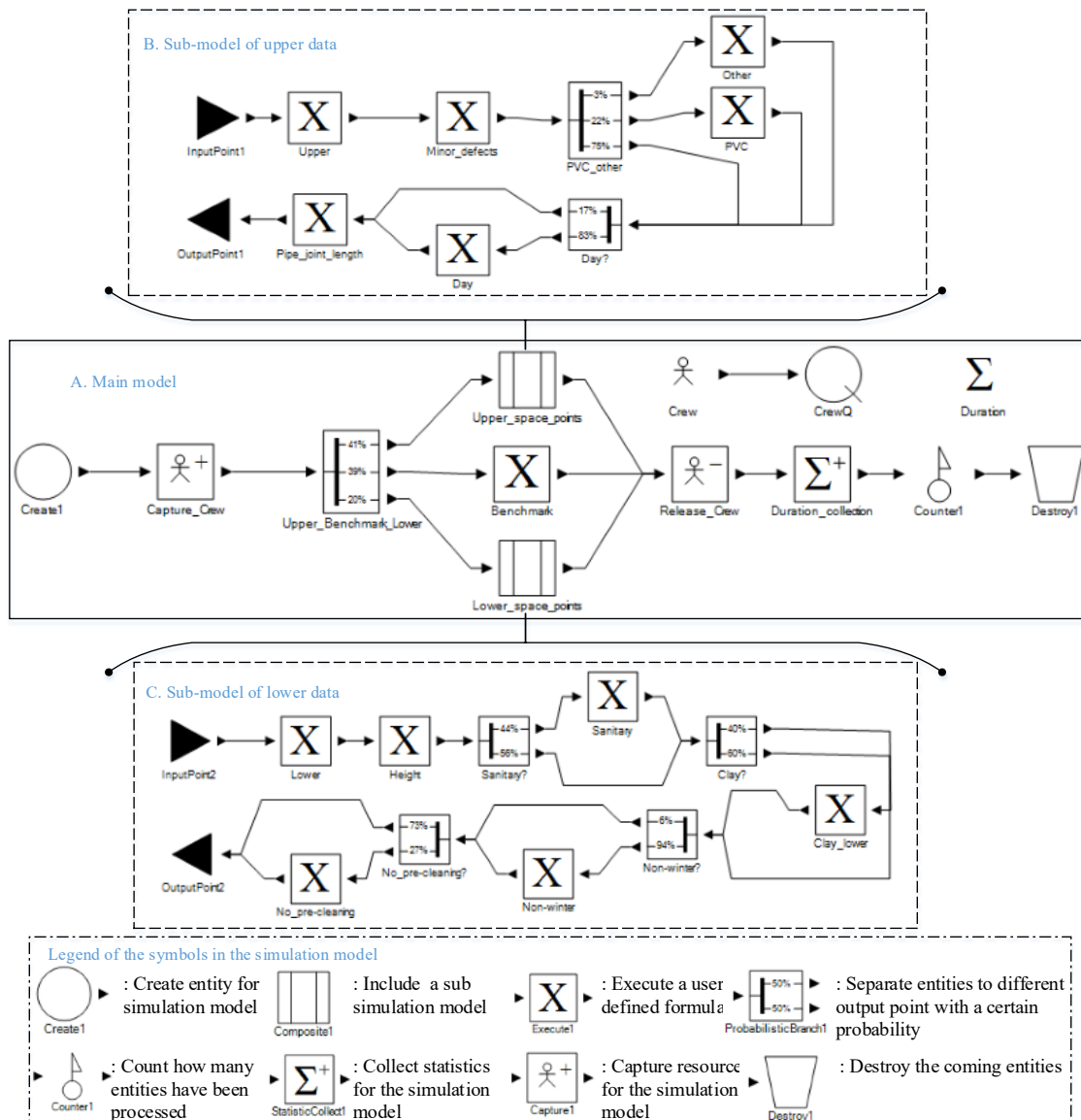
Figure 2.12: Simulation model layout.

45
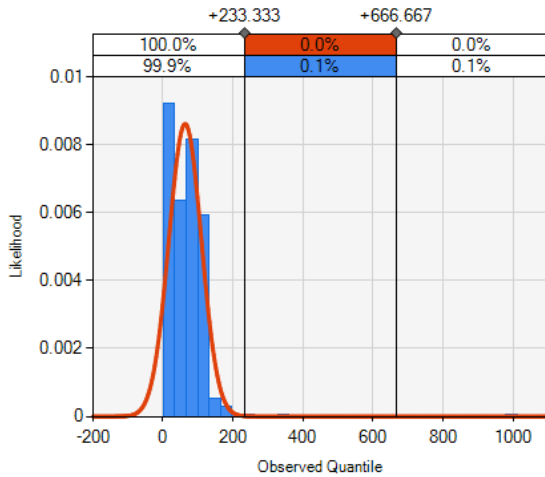
a. Total length, Normal distribution (μ=63.449, σ=5.3919)
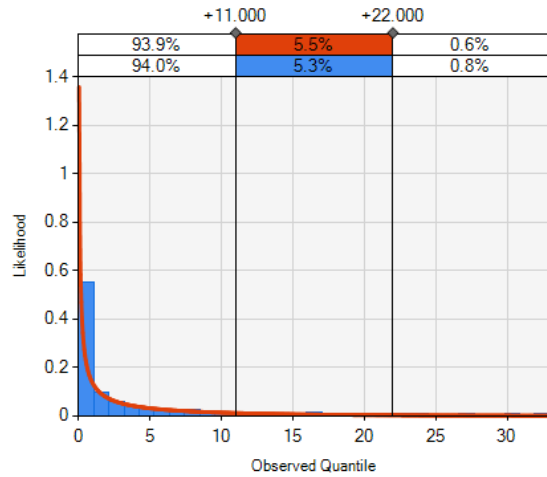
b. Number of taps, Beta distribution (α=0.238, β=2.903, Min=0, Max=33)
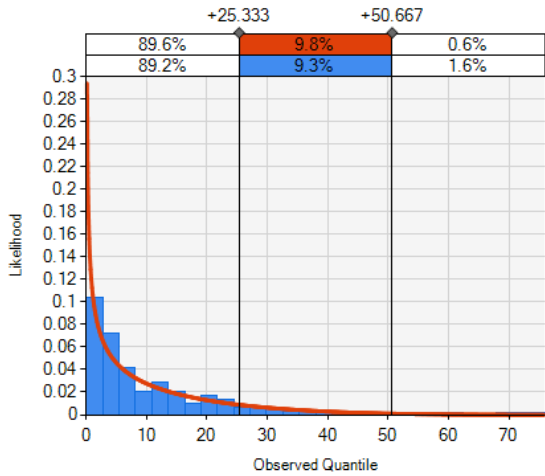
c. Minor defects, Beta distribution (α=0.553, β=3.734, Min=0, Max=76)
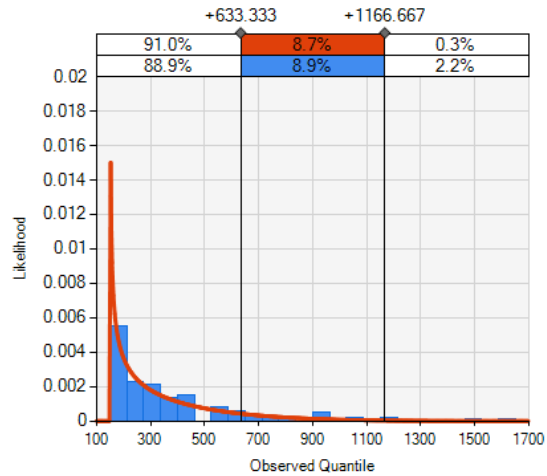
d. Height, Beta distribution (α=0.564, β=4.260, Min=150, Max=1650)

Figure 2.13: Distribution of variables.

2.4.3.3 Output analysis and model validation

The simulation model is validated by comparing the historical recording duration with the recording duration generated by the preliminary linear regression model, the RANSAC model, and the simulation model. The input for the three models is the same which are the variables retrieved from the historical data. In theory, the best prediction model will generate a dataset

(recording duration) that has similar statistical characteristics as the historical data for recording duration (referred to as the real duration in this section).

The input variables are the sub-group of variables shown in Table 2.1. Their selection is described in the above sections. The output of each model is shown in histograms in Figure 2.14 and as boxplots in Figure 2.15. The statistics of each output dataset are summarized in Table 2.5. Figure 2.14 shows three pairs of comparisons of the predicted recording duration generated from three prediction models (preliminary regression model, RANSAC model, and simulation model) with the real-life situation retrieved from the historical data.
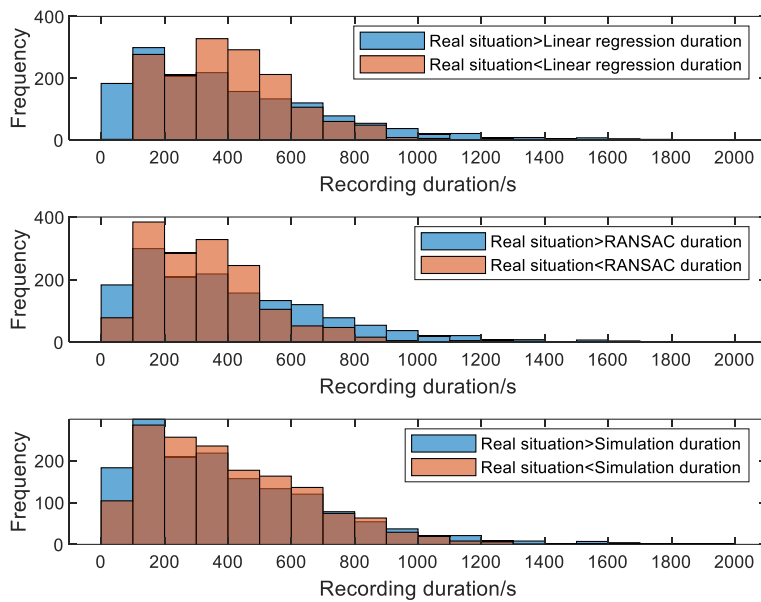
Figure 2.14: Comparison of the histogram.

The first histogram compares the video duration in the real-life situation with the predicted video recording duration by linear regression. There are more video recording durations (predicted by linear regression) that fall in the 300 s to 600 s range than the number of video durations in that same range in the real-life situation. The blue area is the part that cannot be explained by the

model; therefore, the linear model can not accurately predict the situations in which the duration of video recording time is between 0 and 100 s.

The second histogram compares the video duration predicted by the RANSAC model with the real-life situation. The height of the histogram of the RANSAC model is higher than the real-life situation in the 100-500 s range and lower than the real-life situation in the 500-1400 s range, which means that for an ideal situation (where no waiting time occurs during the recording process), there should be more instances where less time is spent on the recording process when compared to the real-life situation.

The last comparison pair in Figure 2.14 shows that the simulation model generates a dataset that is similar to the real-life situation, which has the smallest size of the unoverlapped area. However, between 200 and 700 s, the predicted recording time is slightly longer than the real-life situation, and the difference between the simulation and the historical data is small in comparison with the first two histograms shown in Figure 2.14. Therefore, as can be seen in Figure 2.14, the dataset generated from the simulation model is the most similar to the real-life situation.

The boxplot in Figure 2.15 shows the distribution of the recording duration dataset in another format. The boxplot can give information about the dataset by using 5 values: minimum, first quartile (Q1), median, third quartile (Q3), and maximum (Frigge et al., 1989). The blue box shows the range of the first quartile to the third quartile—which is called interquartile range (IQR)—while the red line in the middle is the median of the dataset. The solid black line in the upper and lower are called minimum and maximum, which are calculated as the Q1-1.5×IQR and Q3+1.5×IQR, respectively. The red cross markers in the upper are outliers identified by the boxplots. The boxplot can also show the tightness and symmetry of the dataset. It is shown in Figure 2.15 that the boxplot of the real-life recording durations is similar to that of the simulation

model with respect to the five values mentioned above. The size of the first blue box is most similar to the last boxplot, and the minimum and maximum values are also close, which means that the tightness of the datasets is similar. The two boxplots in the center represent the results of linear regression and the RANSAC model, which both have a smaller IQR, a lesser maximum value and a greater minimum value than those of the boxplots for the real-life situation and simulation. Some statistical values for these four datasets are summarized in Table 2.5, which shows results similar to those of the histograms shown in Figure 2.14, and the boxplots shown in Figure 2.15.

In general, the most definitive test for the validity of the simulation model is to measure how close the output from the simulation model is to the real-world observations (Law, 2008). Therefore, the effectiveness of the proposed simulation model is validated by comparing the CCTV recording duration outputted by the simulation model with the real situation. The prediction results of the RANSAC model and the preliminary regression model are served for the comparison purpose as well. The outputs from the RANSAC model represent the recording processes that follow the productivity retrieved from the benchmark of the historical dataset, therefore, deviations between the RANSAC and the real situation is inevitable. The preliminary regression model has an unacceptable performance (i.e., adjusted R square of 0.298), which determines that the regression model cannot predict the recording duration accurately. In the meantime, the histogram overlapping analysis (Figure 2.14), boxplot comparison (Figure 2.15), and statistics summarized in Table 2.5 show that the outputs from the simulation model are fairly close to the real situation. Overall, the predicted recording duration generated by the simulation model has the most similar statistical characteristics to the real-life recording duration in many

aspects, as described above, which means that the simulation model is validated by the historical data.
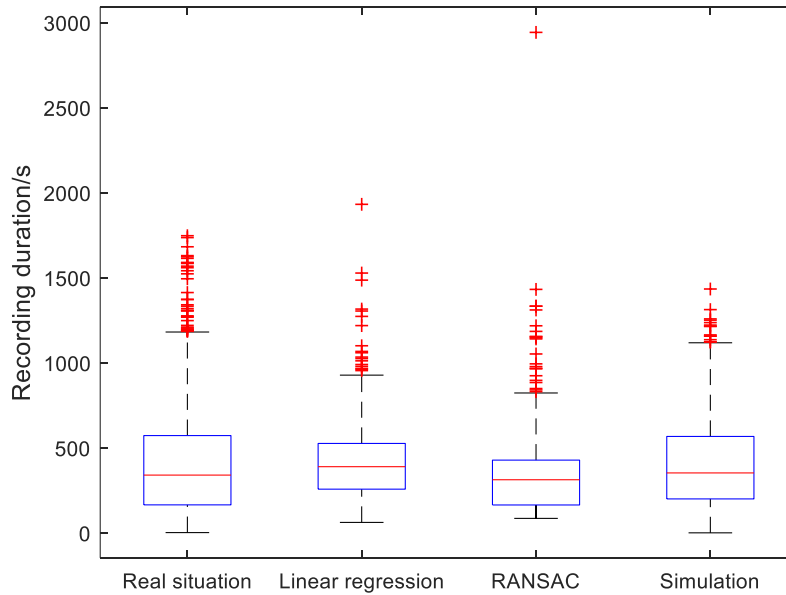
Figure 2.15: Comparison of the boxplots.

Table 2.5: Comparison of the statistics.

Statistics	Real duration (s)	Linear regression model (s)	RANSAC model (s)	Simulation model (s)
Minimum	2.000	61.895	85.590	0.617
Maximum	1747.000	1931.493	2942.219	1433.878
Mean	406.637	409.855	332.028	403.359
Standard deviation	301.618	201.091	205.292	250.049
Sum	633134.000	638143.730	516967.948	628030.163

2.5 Discussion

The simulation model integrated the benchmark extraction and outlier linear regression model to predict the video recording process duration for CCTV sewer pipe inspection purposes. The objectives of developing a simulation model are 1) to simplify the implementation of the multiple models developed to predict the recording duration with simple input and with flexible parameters that can be adjusted during implementation, and 2) to validate the prediction model

50

both visually and statistically. If we want to predict the video recording process duration for a large number of sewer pipe inspections with only the mathematical model proposed in this paper (RANSAC model and outlier linear regression model), the user needs to understand the mathematical model in order to select the proper variables, feed the right input value, separate the dataset by the RANSAC algorithm in order to get the outliers, etc. With the simulation model, the parameters can be adjusted dynamically, for instance, the probability setting of the probabilistic branch may be adjusted based on professional experience or according to the real-life condition of the sewer pipe(s) to be inspected. The output of the simulation model is also used to prove the validation of the proposed prediction model. It may not be accurate when predicting the CCTV video recording process duration for a specific pipe, but the simulation model can predict the total CCTV recording process duration on a larger scale (in terms of a weekly, monthly, or yearly inspection schedule). It is expected that a larger CCTV inspection dataset can improve the model in the future.

Essentially, to develop a simulation model is to build a representation of the real-world process. However, the modeling process always requires a certain level of simplification (abstraction) of the real-world system, but to keep the main features that represent the targeted system (Robinson, 2013). The simplification operation is also involved in this paper since many other potential influencing factors (e.g., depth of the pipe and bedding materials) are not included in the research due to data availability. Other unpredicted factors may also appear before or during the CCTV recording process to cause waiting time, such as unpredicted bad weather, equipment failure, etc. These issues are worth investigating from the perspective of risk measurement in future research. To be more specific, the risk measurement is to measure how much the performance of the DES is affected in the event that the simulation model does not include a specific factor.

In the engineering area, the RANSAC algorithm is widely used for image processing purposes (i.e., Barath & Matas, 2018, Rocha et al., 2011). However, this research proposed an innovative method using the RANSAC algorithm to separate the original dataset into a benchmark dataset and outlier datasets. The simulation model is developed based on the results from the RANSAC algorithm together with other modeling techniques (such as distribution fitting, regression analysis). The DES model developing procedure is innovative and the procedure can be applied for developing the DES models, where data separation is required to improve the performance of the DES model.

2.6 Conclusion and future works

This paper proposes a framework for modeling the productivity of the CCTV recording process for sewer pipe inspection purposes. The video recording process is an important part of the CCTV inspection process and is responsible for approximately 25% of the total time required for the overall CCTV inspection process. Modeling the productivity is the process of developing a model to predict the video recording process duration. A prior experiment in which a linear regression model was developed for the entire dataset was performed and proved to be ineffective at predicting the CCTV recording duration. After that, a prediction model containing two parts, namely benchmark extraction and outliers linear regression, is proposed. A machine-learning algorithm called the RANSAC algorithm is employed to extract the benchmark CCTV recording duration. The objective function retrieved from the RANSAC algorithm also divided the original dataset into three parts: benchmark data, upper space data, and lower space data. Then, a linear regression analysis is performed separately on each of the upper space and lower space datasets. Finally, a DES model is developed by integrating the RANSAC model and the two linear regression models to simulate the recording process. The results show that the

simulation can generate a dataset of CCTV recording process durations that has similar statistical characteristics to that of the historical dataset (real-life situation). The simulated total duration of the CCTV recording time is useful for planning the large-scale CCTV recording and assessment processes. The simulation model can not only integrate multiple models but also provide a simplified platform for users to predict CCTV recording process durations with fewer easier to understand inputs and adjustable parameters to adapt the model to different situations.

The benchmark of CCTV recording process of the city of Edmonton is extracted, which shows that the average moving speed of the crawler is 23.6 m/min, which is faster than the recommended speed in the PACP code, which is 9 m/min. On average, the crawler spends 28.03s on inspecting each tap and 82.13s on other operations, which may include the equipment adjustment and other routine processes. The influencing factors for the outliers are investigated to help to understand the CCTV recording process. For example, the non-winter season will decrease the CCTV recording time, while sanitary pipes are required more time for the CCTV recording operation than the stormwater pipe and combined pipe. In addition, the research proves that the RANSAC algorithm is efficient for data segmentation. The proposed procedure for developing the DES is described in detail and can be adopted for developing DES for various purposes.

Limitations of the research lie in the incompleteness of the input variables, which are retrieved from the database available for this case. Other potential valuable variables such as depth of the pipe and bedding materials are not available in this research. Future research may include more variables for the regression analysis in order to gain a better and more complete understanding of the impact factors for the CCTV recording process. Another future research direction may involve increasing the size of the dataset for the proposed model, which may, in turn, lead to

better performance in terms of predicting the CCTV recording process duration for a specific drainage area.

Chapter 3: A DEEP LEARNING-BASED FRAMEWORK FOR AN AUTOMATED DEFECT DETECTION SYSTEM FOR SEWER PIPES[2]

Preface

In order to conduct effective preventive maintenance, the health condition of the existing sewer pipes needs to be investigated, which is realized by the CCTV video assessment process (see Figure 1.2). This chapter aims at automating the defect detection process, which is conventionally conducted by pipe experts manually. Defect detection is stage one research for fully automating the video assessment process, and stage two research, which is the video interpretation algorithm and system development, is presented in Chapter 4. The major objective of this chapter is to identify and classify six types of defects (i.e., broken, hole, deposits, crack, fracture, and root) and one type of construction feature (tap) in CCTV videos automatically. The performance of the model is validated with a mean average precision (mAP) of 85.37%. Besides that, this chapter proposed a framework for developing a deep learning-based defect detector, and the framework focuses on streamlining the information and data flow, proposing patterns of input and output data processing. Therefore, future studies focusing on developing a defect detector could follow the proposed framework, in which major components, such as the deep learning object detection algorithm (i.e., YOLOv3 in this case study), can be replaced with the most advanced techniques in the future. The outputs of this chapter, which are labeled videos, are the inputs for the video interpretation algorithm proposed in Chapter 4.

3.1 Introduction

Drainage systems have played an important role in the development of human civilization since it has been associated with the increase in human longevity by improving sanitation, especially in the context of highly populated areas where hygiene is a key factor for avoiding large-scale epidemics (Meeker, 1971). Thus, municipalities devote large annual budgets and other resources to conduct preventive maintenance work for sewer pipes (Sinha & Knight, 2004).

A preventive maintenance practice aims at detecting as early as possible structural and operational deteriorations of sewer pipes in order to take the most appropriate corrective actions, e.g., repair, rehabilitation, or if the damage is extreme, replacement of the faulty sections. The purpose of maintenance work is to maintain the health of any infrastructure at a state where wide disruptions in operation are minimized, thus allowing a near-optimal level of service.

Defect examples are presented in Figure 3.1, and the defect images used in this research are selected based on the manual assessment results. It is important to note that although defects contained in each video are classified according to the description contained in a standardized code (e.g., PACP code), technologists' personal judgments during this phase may lead to some uncertainty in the final classification. For example, a crack may be easily confused with a fracture. The definition for the crack code in PACP is "the crack code is used where a break line is visible on the surface but is not visibly open; no gap is visible between the edge of a crack"; however, the definition for fracture is "a fracture is a break line that has become visibly open and a gap can be seen, although the sections of pipe wall are still in place and not able to move" (NASSCO, 2015). Therefore, the boundary of crack and fracture can be blurry in some cases, and their classification depends highly on the experience of the technologists.

Perhaps the most challenging part of the assessment phase is the need for technologists to watch every single video in its entirety even though defects are generally not frequent. In this respect, given that the time spent on classifying a defect is the time that adds value to the assessment process, it can be easily understood that a substantial amount of a technologists' time is wasted because of the infrequency of defects (Cheng & Wang, 2018). At the end of the video assessment phase, all the defects contained in the sewer sections of interest are identified, classified into different types, and labeled according to a standardized nomenclature (PACP code (NASSCO, 2015) is used in this paper). Different types of defects or construction features are recorded in detail, for instance, defects such as crack, fracture, broken, hole, deposits, and root are recorded; construction features such as tap are recorded as well. Each of these defects can have several subclasses (e.g., longitudinal crack and circumferential crack).

(a). Broken (b). Hole (c). Deposit

(d). Crack (e). Fracture (f). Root (g). Tap
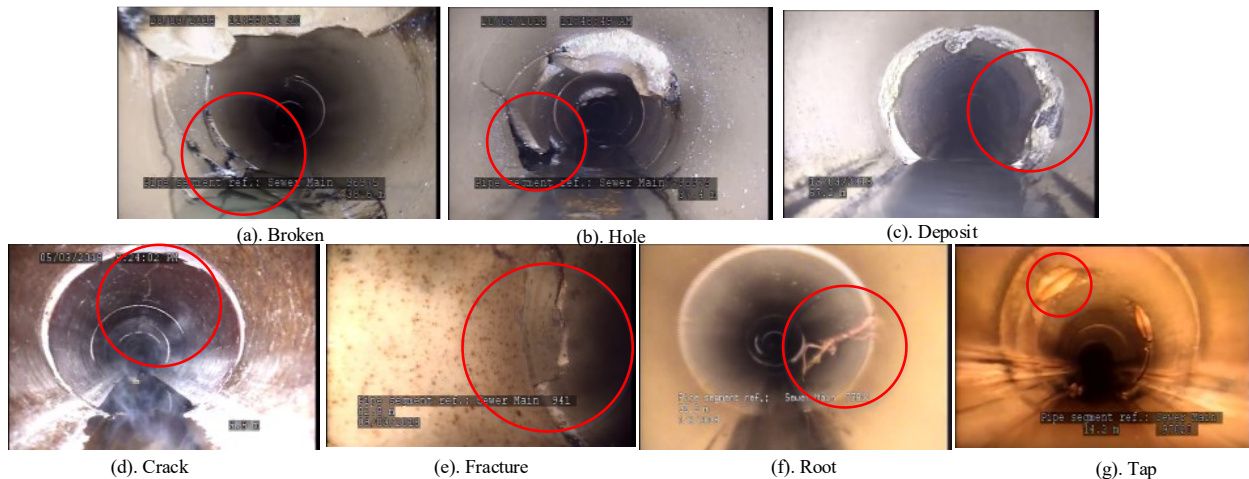
Figure 3.1: Examples of defects and tap.

This paper focuses on the detection of the aforementioned six types of defects and one type of construction feature in accordance with the available data. In addition, these types of defects are the most common problems that exist in sewer pipes and are considered important items by the municipal department that collaborated with this research team. The data and CCTV videos were provided by EPCOR Drainage Services based in Edmonton, Canada. A complete framework for

developing the automated defect detection system with a deep learning algorithm is proposed in this research. The input data, objects detector, and output data are described in detail. This framework sheds light on the system development process, which can streamline the input and output of the system. The proposed framework uses CCTV video as the input data instead of images as was does in previous research. The output data contains three parts, namely, labeled videos, labeled frames with defects, and defect information. Each of these can be used to serve the sewer pipe assessment process. The main part of the system is the deep learning-based object detector, which can be replaced with the advent of more advanced algorithms. In this research, YOLOv3 is used as the object detector.

3.2 Background

3.2.1 Sewer pipe maintenance and inspection

Within the engineering community worldwide, the implementation of proactive and preventive maintenance strategies for sewer pipes has gained an increasing amount of interest when compared to conventional reactive maintenance strategies (i.e., wait for a failure to occur then repair) given the important negative effects, i.e., financial, social, and ecological impacts, associated with unexpected failures (Fenner, 2000). However, many communities are still failing to carry out an effective maintenance strategy to mitigate the risk of largescale failures of sewer pipes that can lead to great inconveniences for residents and (extremely) high rehabilitation costs (Ariaratnam & MacLeod, 2002).

In the last few decades, sewer pipe maintenance strategies and inspection techniques have been developed to improve maintenance productivity and lower inspection cost. As for maintenance strategies, a common method is to classify the sewer pipes as critical and non-critical pipes. For example, Baah et al., (2015) have employed a risk matrix and a weighted sum multi-criteria

decision-matrix to develop a risk-based model to prioritize sewer pipes for a mid-sized city. Anbari et al. (2017) have used a probabilistic method and a weighted average method to compute the probability of failure and consequence of failure for sewer pipes in order to classify the sewer pipes into different classes from the perspective of risks. After the classification, a maintenance schedule can be made accordingly; for instance, sewer pipes in critical situations have a higher priority in the maintenance schedule.

Many inspection techniques have been developed with the development of technologies in recent years. Liu & Kleiner (2013) have summarized the state-of-the-art inspection techniques for pipe inspection. According to this research, Table 3.1 has been developed to summarize all the technologies investigated by current researchers. Pros and cons exist for each of the techniques for different types of sewer pipes. The adoption of a particular inspection technique depends on the unique situational factors of a given project, such as budget, existing professionals, equipment requirements, etc. (Duran et al., 2002; Liu & Kleiner, 2013; Wirahadikusumah et al., 1998). Among all the techniques, CCTV inspection is one of the most commonly used methods in practice due to several advantages inherent to this technique, such as fewer inspectors required, the safety of the inspectors is ensured, easy operation, etc. (Wu et al., 2015; Yang & Su, 2008). However, several factors can make CCTV inspection challenging, for instance, the lighting conditions inside the sewer pipes can deteriorate when the environment is foggy, the quality of the video highly depends on the operators' skills (Z. Liu & Kleiner, 2013), obstructions, such as deposits, may exist that can make the deploying of the CCTV camera impossible, and pipe sagging can, in some cases, lead to the camera being totally immersed in water (Olga Duran et al., 2002a).

Table 3.1: Summary of water pipe inspection technologies. (Liu & Kleiner, 2013).

Inspection method	Techniques
Visual inspection	CCTV inspection
	Laser scan
Electromagnetic methods	Magnetic flux leakage
	Remote field eddy current
	Broadband electromagnetic
	Pulsed eddy current testing
	Ground penetrating radar
	Ultra-wideband pulsed radar system: P-Scan
Acoustic methods	Sonar profiling system
	Impact echo
	SmartBall
	Sahara system
	Leak detection
Ultrasound methods	Guided wave ultrasound
	Discrete ultrasound
	Phased array technology
	Combined UT inspection
Radiographic methods	
Thermography methods	
Linear polarization resistance (LPR) of soil	
Soil characterization	
Pipe to soil potential survey	
Smart pipe	
Computer-aided approach: augmented reality	
Intelligent pigs and robotic survey systems	

Currently, in the vast majority of operational settings, CCTV on-site data collection is followed by a video assessment phase that is conducted manually by trained technologists. However, despite the trained eyes of the technologists, several factors can influence the outcome of the defect detection and classification operation, the most important of which are the quality of the video, the difference in the level of expertise between technologists, fatigue, and lack of attention and possibly the pressure to maintain a certain level of performance (Sinha et al., 2003). Considering this, computer vision-based techniques can contribute to the automation of the

assessment operation, improve productivity, and also help in reducing the human error of the CCTV video assessment operation.

3.2.2 Computer vision-based techniques for sewer pipe defect detection

Manual CCTV video evaluation is a tedious and time-consuming job for technologists that can be error-prone and highly dependent on their level of expertise (Mashford et al., 2010). Since the core component of CCTV video assessment is the analysis of video frames, i.e., images, for the purpose of identifying and classifying defects, the path towards automating the process will naturally rely on the multitude of algorithms that were developed in the last few decades for computer vision and image processing.

Pictures or video frames that record the inner situation of the sewer pipes are fed to computers for the purpose of automated defect detection. Table A-1 in Appendix A shows the summary of publications in selected journals regarding the topic of computer vision-based techniques for sewer pipe defect detection. It can be seen that several techniques have been developed by researchers regarding sewer pipe image processing for the purpose of classifying an image that contains defects from the general image (contains no defects). Moselhi & Shehab-Eldeen (1999,2000) proposed a semi-automated defect detection system using a three-layer back-propagation neural network. This system is fed with preprocessed images that are grabbed by a frame grabber from CCTV video. The feature information of defects (e.g., crack, joint displacement) is served as the input of the proposed neural network while the output is the classification results of the different type of defects. Sinha et al. (2003) further developed the defects classification method by integrating fuzzy logic theory with a neural network (namely, neuro-fuzzy network) to classify the defects of the sewer pipes. The morphological approach also employed to detect the image contains defects for the purpose of automated defect detection,

such as Iyer & Sinha (2006), and Mashford et al. (2010). Techniques such as support vector machine (SVM) (Mashford et al., 2008) and multilayer perceptron (MLP) (Duran et al., 2007) are also investigated and tested in the defect detection process.

In recent years, the rapid increase in computing power has led to the development of deep learning techniques. Object detection, as an important application of deep learning, has been gaining huge traction, with the application of pose estimation, vehicle detection, and surveillance, and so on. (Gandhi, 2018). Convolutional neural networks (CNN) are commonly used in the area of image classification (Krizhevsky et al., 2012). However, the standard format of CNN is ineffective when used without any modification in object detection problems since the input and output data is not in a fixed format. To be more specific, for the output of an object detection problem, processing one image will give the outputs of classification of detected objects (one or many) associated with one/many bounding boxes to locate the objects in the image. For the input, if we want to use standard CNN to detect the objects within an image, the first step is to find the regions of interest (RoI), which are the regions that contain an object that we want to detect. Then we need to classify this RoI into a certain class. However, to find all the RoI would take significant time and computation resources (Gandhi, 2018). Therefore, several algorithms, which are the extension of CNN, have been developed to solve the aforementioned problem, such as region-based CNN (R-CNN) (Girshick et al., 2014), fast R-CNN (Girshick, 2015), faster R-CNN (Ren et al., 2017), SPP-net (He et al., 2015), Single Shot Detector (SSD) (Liu et al., 2016), You Only Look Once (YOLO) (Redmon et al., 2016), etc. The major improvements of R-CNN, fast R-CNN, and faster R-CNN are their capacities to utilize innovative methods to find region proposals. For example, the R-CNN uses selective search method to extract around 2,000 regions from one image as the input for CNN (Girshick et al., 2014), while the fast R-CNN uses

convolutional feature map, generated by the input image, to find the RoI, which will feed to the R-CNN for detection purposes (Girshick, 2015). Faster R-CNN uses a region proposal network (RPN) to conduct the region proposals, which is much faster than fast R-CNN and R-CNN (Ren et al., 2017). These algorithms could classify several objects into different classifications and locate each object with a bounding box. The performance of such algorithms can be measured in multiple ways, such as mean average precision (mAP), intersection over union (IoU), and training speed, etc. (Liu et al., 2016).

Recently, some of these aforementioned efficient algorithms were employed in the automated sewer pipe defect detection research. For example, Kumar et al. (2018) proposed binary CNNs to classify images as containing or not containing a specific type of defect, which means in their research, a certain number of CNNs need to be trained to classify a certain number of types of defects. Cheng & Wang (2018) employed faster R-CNN model to detect four types of defects of sewer pipes (tree root intrusion, deposit, infiltration, and crack) automatically. Li et al. (2019) proposed a deep CNN with hierarchical softmax based on a vanilla Resnet18 network to form a two-level structure to automatically detect defects, such as deposit settlement, joint offset, broken, obstacles, water level stag, and deformation. The aforementioned research mainly focuses on processing images; however, in practice, the format of the inspection medium is video. The proposed framework focuses on processing CCTV video, with the output of a labeled video to label the defects with bounding boxes associated with the defect's name. In addition, the deep learning method (i.e., YOLOv3) employed in this research is superior to the previously used model in terms of processing speed, which will be discussed in the next section.

3.3 The Proposed system framework

3.3.1 Framework of the sewer pipe defect detection system

The overview of the proposed system framework is presented in Figure 3.2. CCTV inspection videos serve as the original data source, from which images containing defects and feature of defects can be derived. PACP code is utilized in this research as a major criterion for defining different types of defects since it is commonly used in North America. PACP code is also the standard used to classify the data used in this research. For the framework, some user-defined parameters are required, especially for the model training process. The object detection algorithm is YOLOv3 network in this research considering its many advantages (such as accuracy and processing speed). The major process includes image processing, model training, and video processing, which are connected in sequential order. The process will be discussed in detail in the next section. Outputs of the framework are labeled videos/images, frames that contain defects, and other information about defects associated with the CCTV video.
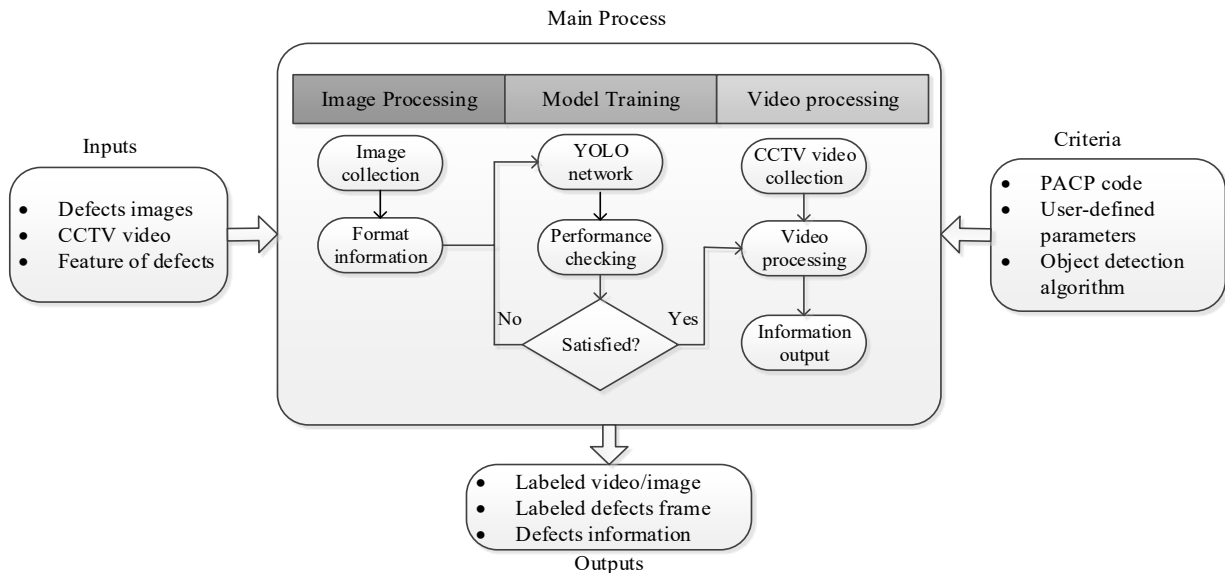
Figure 3.2: Overview of the automated defect detection system.

3.3.2 Input data preparation

The input images are collected from historical data that are recorded over time. In order to store the historical sewer pipe inspection information properly and for the convenience of later-stage processing, a database structure shown in Figure 3.3 is adopted in this research. Four databases are linked with each other through specific primary keys and foreign keys, and all the information is recorded during the CCTV footage assessment process. The inspections database records the basic properties of the sewer pipe and the media inspections database records the video information of the specific sewer pipe inspection. The conditions database records the defect information for the specific sewer pipe and is connected with the inspection database by the inspection ID. The media conditions database stores the video frames that contain defects and the corresponding defect ID (i.e., condition ID). With the structural database and the proposed frame extraction algorithm, video frames containing a specific type of defect can be extracted automatically.
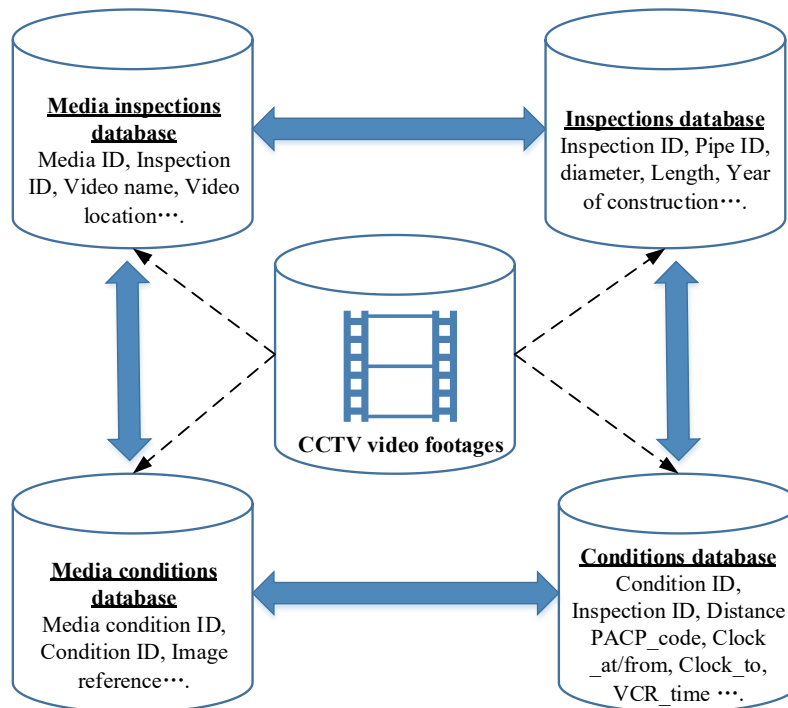
Figure 3.3: Input data schematic.

A procedure is proposed to prepare the input data for YOLOv3 network semi-automatically. The flow chart for this process is presented in Figure 3.4. The first step is to collect the original CCTV video from the database and also retrieve the corresponding inspection information, which is also recorded in the database. Note the CCTV video footage we are using to get the defect frames have been assessed by professional individuals according to the PACP code so that we can grab only those frames that contain defects. Then defects frames need to be extracted in the next step. A sub-process is shown in the right part of Figure 3.4 to show the detailed procedure. For each CCTV video, we need to find out the information recorded by the professional technologist related to each defect. More specifically, we need to find out the CCTV time in order to know where precisely the defects are in the CCTV video. Then, an automated frame grabbing algorithm developed in Python is proposed in this study for the purpose of grabbing multiple frames around that time.
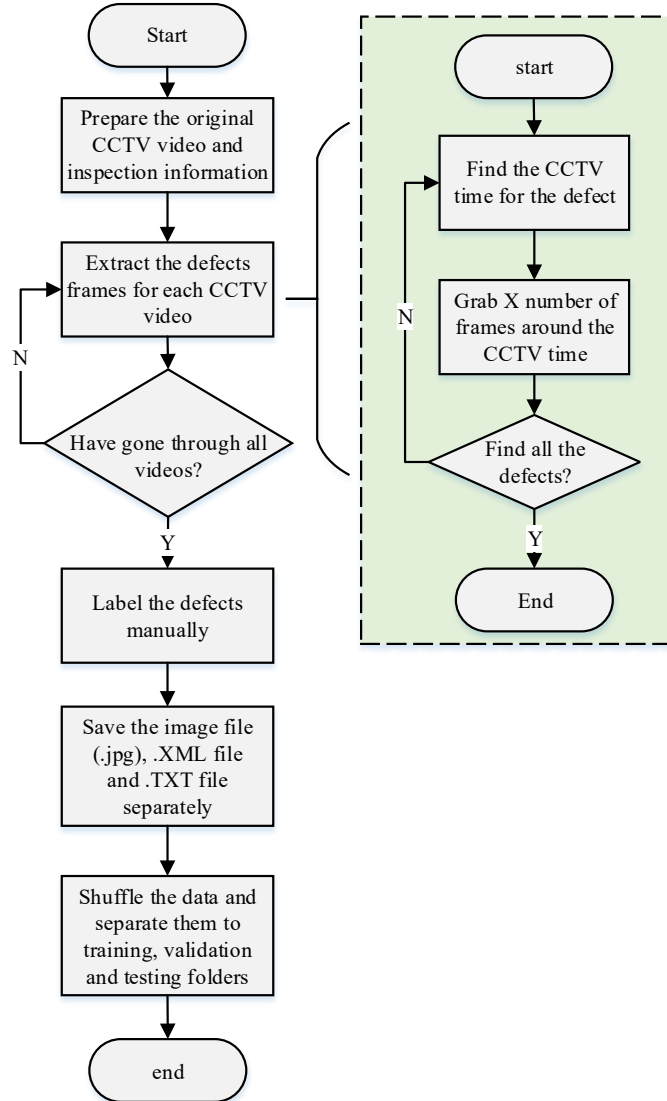
Figure 3.4: Flow chart of the input data preparation.

The pseudo-code for the automated frame grabbing algorithm is presented in Figure 3.5. For example, if the database shows that there is a crack at the video time of 45 seconds, the frame grabber will grab 60 frames (30 frames before the video time of 45 seconds, and 30 frames after). The number of frames grabbed in this step can be determined by the users. After all the defects are identified and the corresponding frames are extracted, the labeling process is processed successively. The labeling process is conducted manually by putting a bounding box for the specific defects, which serves as the ground truth in the objects detection algorithm. After

labeling the images, a .XML file is generated as a result, which contains defect information (e.g., the location in the image, the class) followed by an algorithm that can transform the .XML file to a .TXT file to align with the input data format for training the object detector. An algorithm is developed to shuffle the data and divide the data into the training group, validation group, and test group randomly with a predefined proportion. Save path is created to record the input data for the later training process.
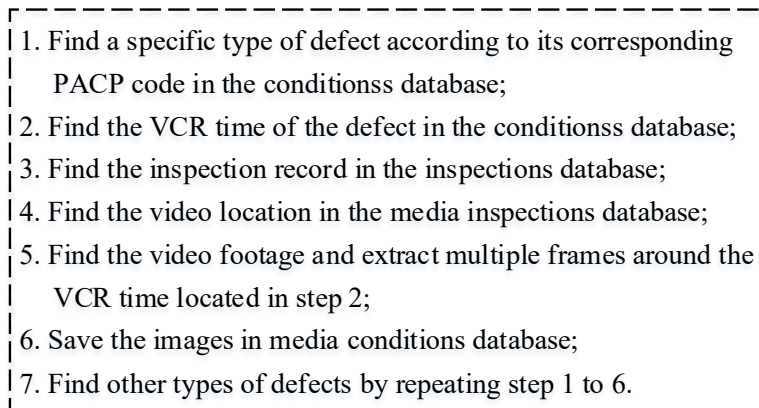
```
1. Find a specific type of defect according to its corresponding
   PACP code in the conditionss database;
2. Find the VCR time of the defect in the conditionss database;
3. Find the inspection record in the inspections database;
4. Find the video location in the media inspections database;
5. Find the video footage and extract multiple frames around the
   VCR time located in step 2;
6. Save the images in media conditions database;
7. Find other types of defects by repeating step 1 to 6.
```

Figure 3.5: Pseudo-code for extracting the video frames.

3.3.3 Implementation process of the framework

The implementation process can be divided into three sections, namely, image processing, model training, and video processing (see Figure 3.2 in process box). Three boxes depict the detail of each section in Figure 3.6. The red box represents the image processing section, which is described in section 3.3.2. The blue box presents the model training process: YOLOv3 network is employed as the object detector, which will be described in detail in section 3.4. Training and validation data serve as the input of the YOLOv3 network. A performance check is conducted on the test dataset. If the performance is satisfied (for example, when the measurements for accuracy and completeness exceed the user-defined parameters), the model structure (weights of the network) will be fixed and sent to the next stage, which is the video processing stage (the yellow box in Figure 3.6.). The first two stages mentioned above (red and blue boxes in Figure

68

3.6) are the foundation of the last stage. Once the detector has been trained, it can be used for the daily assessment of the sewer pipe CCTV video inspection. The output of the model contains three parts that will be discussed in the next section, namely, labeled video, frames containing defects, and documents of information pertaining to defects.
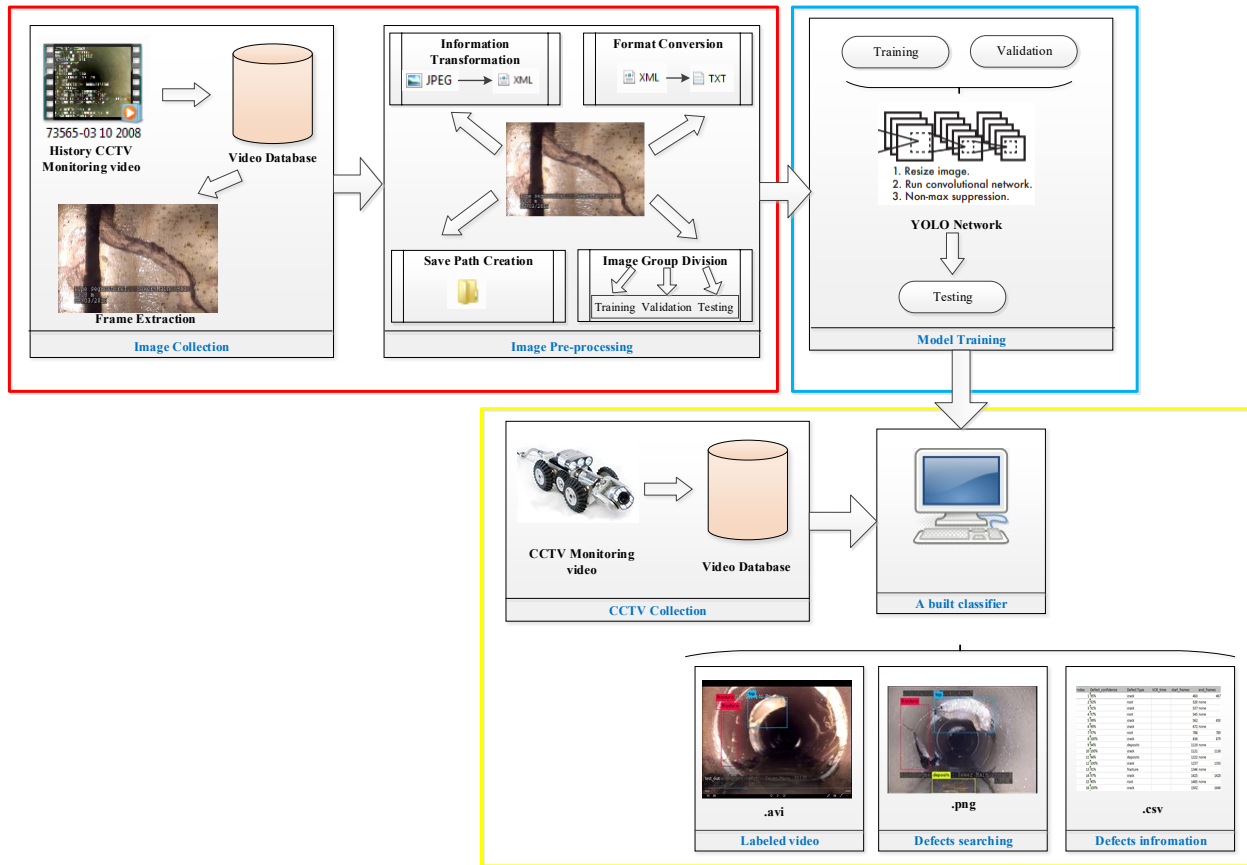
Figure 3.6: The framework of machine learning-based automated defect detection system.

3.3.4 Output of the framework

Different from the conventional manual assessment process, with this sewer pipe defect detection framework, the CCTV video will go through the system before being sent to the technologist. The outputs of the system are labeled videos and frames containing defects extracted from the video. As the video plays, the different types of defects are labeled with different colors and names. In addition, video frames containing defects can be derived from the

system. Then, the technologist can watch the labeled CCTV video to record the defect information. If the system is good enough, the technologist can trust the system and record the defect information as presented in the video. However, this kind of system cannot achieve 100% accuracy in practice for any machine learning techniques for now, which leads to the job of the technologist becoming the double-checking of the labeled video and the recording of the defect information instead of detecting the defects, classifying the defects, and recording the defect information manually. The CCTV video labeling and technologist assessment serve as a complement to each other. If the system labels the video incorrectly or misses some of the defects, the technologist can correct these errors while recording the defect information. Furthermore, the detecting and classifying of sewer pipe defects requires professional knowledge and concentration during the assessment process, thus the labeling of defects in the CCTV video can serve as a benchmark for the technologist.

The video frames containing defects can be exported by the system using the proposed defects searching function (see the yellow box in Figure 3.6) as well. There will be at least 30 frames that contain one defect that appears in the video for over 1 second (because our videos have 30 frames per second (FPS) property). All the defects detected by the system will be assigned a confidence number to represent the probability that the detected object is truly a specific defect. The process will be discussed in section 3.4.1. Multiple defect frames capturing the same defect should be processed with filtering and selecting approaches. Filtering is used to exclude the video frames detected by the system with a lower confidence number (which is a user-defined parameter that can be selected based on experiments). Selecting is used to choose the frame that contains defects with the highest confidence number in consecutive frames. The process will be conducted as follows. First, with all the frames outputted by the system containing defects, a

70

threshold should be set to filter the defect frames with low confidence numbers. Then, for the filtered frames with an acceptable confidence number, the specific frame with the highest confidence number will be selected to represent the defect shown in a consecutive sequence of frames that capture the same defect. For example, if the system gives information about a defect as having detected a fracture from video frames 1000 to 1100, the output will be one image of the fracture with the highest confidence number assigned by the system, instead of all the 101 frames. Finally, the technologist can examine the defect frame instead of watching the whole video in the case where a time-sensitive assessment is required, or in the case when the technologist wants to get a rough impression of the sewer pipe before conducting a more careful assessment process.

In addition, the system can provide defect information regarding a specific pipe in the format of a CSV file, which helps in the pipe evaluation procedure. The pipe evaluation can be processed according to the related code (such as PACP or WRc). This function will be developed in our future research.

3.4 The deep learning method for detecting the defects

3.4.1 Deep learning algorithm used in this research

This framework employed YOLOv3 (Redmon & Farhadi, 2018), which is the updated version of YOLO (Redmon et al., 2016) algorithm, as the defect detection tool. As an efficient object detection method, YOLO algorithms (e.g., YOLO, YOLOv2, YOLO9000, etc.) have been used in areas such as real-time detection of traffic participants (Ćorović et al., 2018), real-time fruit detection (Koirala et al., 2019), diagnosis system for digital X-ray (Al-antari et al., 2018; Al-masni et al., 2018), etc. However, it has not previously been used in sewer pipe defect detection problem. The major difference between YOLO and the region-based objects detection algorithm

(e.g., R-CNN, Fast R-CNN, Faster R-CNN.) is that instead of extracting region proposals for the classification process, YOLO only uses pre-defined grid cell to carry out the prediction, which will improve the detection speed significantly, making this algorithm realize real-time object detection. The speed of image processing for the different version of YOLO algorithms can vary from 45 FPS to 155 FPS a Titan X GPU (Redmon et al., 2016; Redmon & Farhadi, 2017, 2018). YOLOv3 is one of the state-of-the-art real-time object detection algorithms, whose performance is superior from the perspectives of speed and accuracy (Redmon & Farhadi, 2018; Tian et al., 2019).

Figure 3.7 shows the mechanism by which YOLO works (Kathuria, 2018a). There is a crack in the image that we want to detect by YOLO. The first step is to divide the image into an S × S grid; in this example, we use a 13 × 13 grid. The green box locates the crack of the sewer pipe inner wall. The center of the crack falls in the red grid cell, which will be used to predict crack in the late stage. In other words, the object in the green box is predicted by the red grid (which is the grid cell that the center of the object falls in). The red grid cell is the feature map that is responsible for detecting the crack in the image. For each grid cell, the model predicts B bounding boxes (B is the number of predicted bounding boxes for each grid cell, B equal to 3 in YOLOv3 model). In YOLOv3 model, the predicted bounding boxes are using anchor boxes, which are the priors generated by k-means clustering on training set bounding boxes (Redmon & Farhadi, 2017). The number of attributes of each predicted bounding box is equal to 5 + C where,

- 5 includes the box coordinates (t_x, t_y, t_w, and t_h) and objectness (p_o). The attributes are listed in the right part of Figure 3.7, where t_x and t_y are the offsets for the coordinates of anchor boxes: for example, if t_x=1, the anchor box should shift to the right for a distance of 1; t_w and t_h are the offset for the width and height of anchor box.

- C is the number of class of the predicted objects, for instance, we want to predict 7 types of defects in this research, the corresponding C is equal to 7. Therefore, we should predict 36 (calculated as (5+7) ×3= 36) attributes for each grid cell in this research (where B = 3, C = 7).
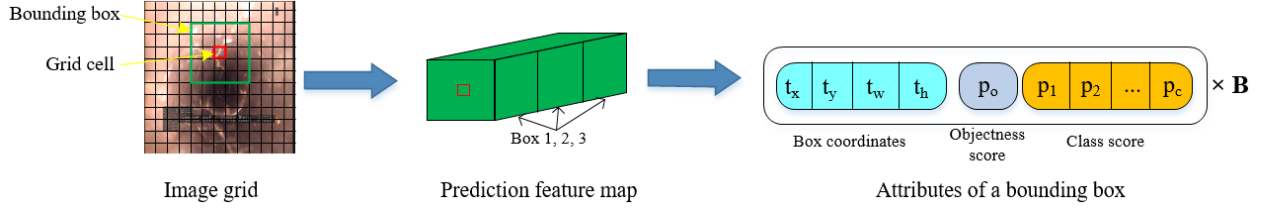
Figure 3.7: The YOLO detection system.
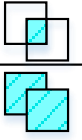
With the box coordinates, the predicted bounding box can be located as per the equation below (see Figure 3.8). The box width dot line is the anchor box while the blue box is the predicted bounding box (Redmon & Farhadi, 2017):

$$b_x = \sigma(t_x) + c_x$$
$$b_y = \sigma(t_y) + c_y$$
$$b_w = p_w e^{t_w}$$
$$b_h = p_h e^{t_h}$$

where b_x and b_y are the X-axis and Y-axis coordinates for the center of the predicted bounding box; b_w and b_h are the width and height of the predicted bounding box; p_w and p_h are the width and height of the anchor box; and c_x and c_y are the distance along the X-axis and Y-axis to the center grid cell upper-left corner. In Figure 3.7, p_o is the objectness score which presents the probability that the object is located inside the bounding box, and p_1 to p_c are the class scores that indicate the probabilities that the object belongs to a specific class. YOLOv3 uses an independent logistic classifier to conduct the class prediction instead of softmax function in order to label an object that belongs to multiple class (i.e. Woman and Person) (Redmon & Farhadi, 2018). The confidence for each bounding box is defined as (Equation (3-1)):

$$Confidence = \Pr(Class_i | Object) \times \Pr(Object) \times IOU_{pred}^{truth} = \Pr(Class_i) \times IOU_{pred}^{truth} \qquad (3\text{-}1)$$

where, $\Pr(Object)$ is a binary variable that indicates that if a bounding box contains an object or not; $\Pr(Class_i | Object)$ is a conditional probability that represents the probability that an object belongs to a specific class; and IOU_{pred}^{truth} is intersection over union that indicates the ratio of the area of intersection between ground truth and predicted bounding box over that of the union of the two boxes, which can be calculated as (Nakazawa et al., 2019):

$$\text{IoU} = \frac{\ \ }{\ \ }$$

The final bounding box in the image can be chosen with the non-maximum suppression (NSM) method (Rothe et al., 2015) in case of multiple predicted bounding boxes for one object. The loss function of YOLOv3 is defined in Equation (3-2) (Redmon et al., 2016; Tian et al., 2019) where, $Error_{coord}$ represents the error for bounding box coordinate error; $Error_{iou}$ represents the error for the IoU; and $Error_{cls}$ represents the class error. For a detailed mathematical description of the loss function, the interested reader may refer to Redmon et al. (2016). All the errors in Equation (3-2) penalize the difference between the prediction and the ground truth, which has the best value of zero. The overall architecture of the YOLOv3 is described in the next section.

$$Loss = Error_{coord} + Error_{iou} + Error_{cls} \qquad (3\text{-}2)$$
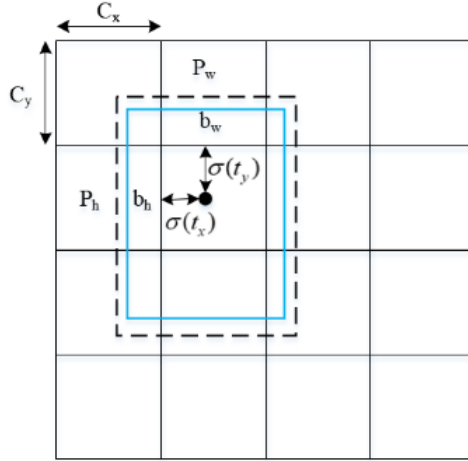
Figure 3.8: A bounding box with anchor box and location prediction.

3.4.2 YOLOv3 architecture

The overall architecture of the YOLOv3 is shown in Figure 3.9 (Ćorović et al., 2018; Kathuria, 2018b; Redmon & Farhadi, 2018). The detailed structure of the YOLOv3 employed in this research can be found in Table A-2 in Appendix A. There are 106 layers in total containing 75 convolutional layers together with residual blocks, detection blocks, and upsampling layers. Convolutional layers with a stride of 2 have been used to downsample the feature map instead of using pooling layers, which usually lead to low-level feature loss (Kathuria, 2018a). The major improvement of YOLOv3 compared with the former version of YOLO algorithm (e.g., YOLO, YOLOv2, YOLO9000) is that it has good performance when detecting small objects in the image due to the 3-scale prediction method. YOLOv3 predicts boxes at 3 different scales (e.g., 13×13, 26×26, 52×52) (Redmon & Farhadi, 2018), the detailed mechanism is described in section 3.4.2.3.
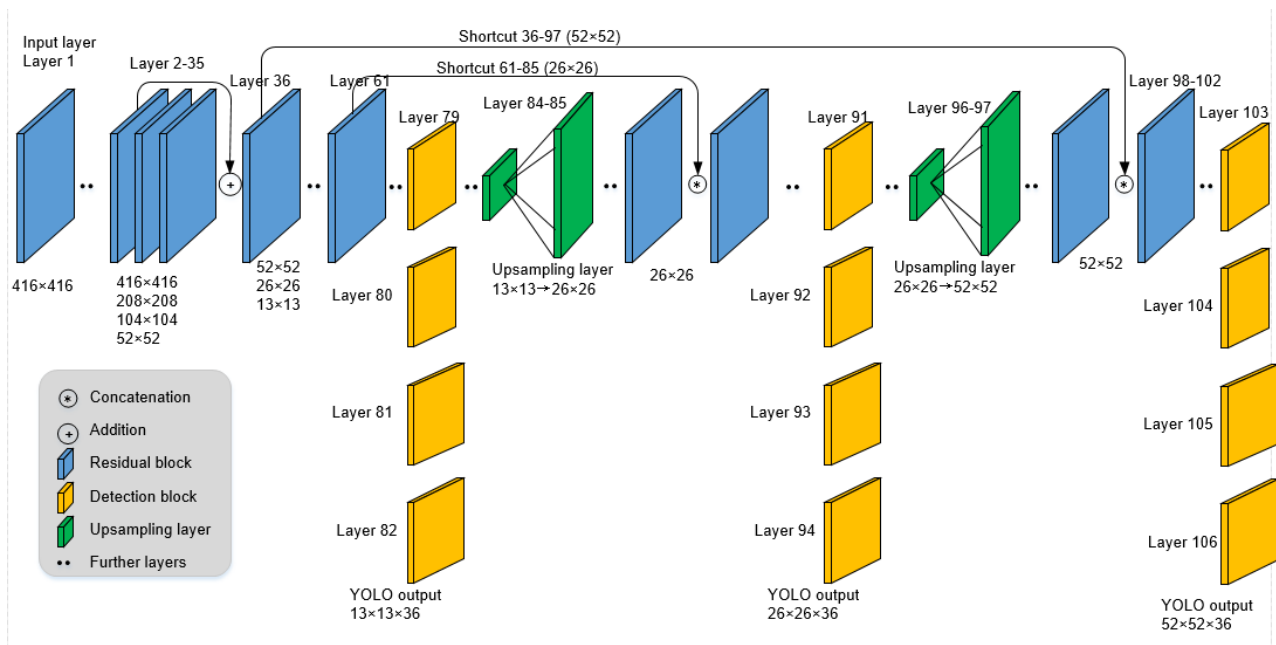
Figure 3.9: The architecture of the YOLOv3 network.

3.4.2.1 Convolutional layers

The major task of the convolutional layer is to perform convolution operations to extract important features from the preceding layers (Dung & Anh, 2019). The filters or kernels are used with predefined stride and padding settings to convolve around the input volume. The mechanism of convolution operation is well described by several researchers who are using CNN to detect the defects of sewer pipes (Cheng & Wang, 2018; Kumar et al., 2018). The difference for YOLOv3 convolutional layers is that it doesn't contain any pooling layers; as an alternative, it utilizes filters with a stride of 2 to downsample the features as mentioned above.

3.4.2.2 Residual block

In order to increase the accuracy of image classification, deeper CNN has been employed to carry out the image classification operation, taking advantage of the increase of computing power (Simonyan & Zisserman, 2015). However, as the number of layers increases, the accuracy drops in this process, which is due to the so-called degradation problem. In addition, the

76

vanishing/exploding gradients problem, which exists in the deep CNN training process, prevents the convergence of the algorithm (He et al., 2015a). In order to address the degradation problem, as well as the vanishing/exploding gradients problem, a deep residual network has been proposed by He et al. (2015a). It also has been deployed in YOLOv3 to increase the accuracy in such a deep network. The structure of the residual block is presented in Figure 3.10 (He et al., 2015a; Li et al., 2019). Relu function (f(u)=max(0,u)) serves as the activation function since it has been proven to be more efficient in classification problems (Nair & Hinton, 2010) and deep neural networks (Glorot et al., 2011) compared to other activation functions (e.g., sigmoid function). Input x could go through each convolutional layer step by step, and in the meantime, it can also take the shortcut path (curve line in Figure 3.10) to skip the convolutional layers. H(x) is the desired mapping function, which can be represented as F(x)+Wx. F(x) represents the residual mapping to be learned, W is a square to transform the dimension of the input x to matching with the output. He et al. (2015a) have found that it is easier to optimize these residual networks and get an increased accuracy for deep CNN. Therefore, several residual blocks are employed in YOLOv3 as an incremental improvement for the early version of YOLO network.
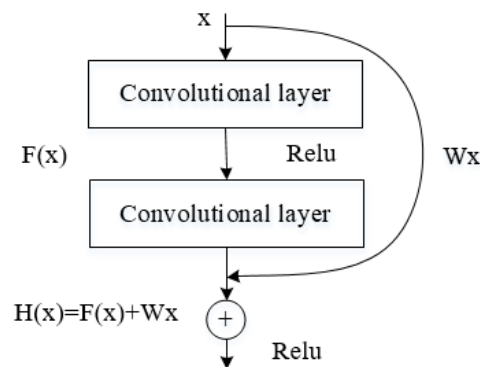
Figure 3.10: Residual block.

3.4.2.3 Detection block

YOLOv3 carries out the prediction operation at 3 different scales (e.g., 13×13, 26×26, 52×52) in order to efficiently detect objects of different sizes. The yellow squares in Figure 3.9 are the detection blocks, of which there are three sets. The number of predicted attributes for each bounding box is (5+7) ×3= 36 (see section 3.4.1). For each grid, YOLOv3 predicts three bounding boxes, thus the total number of the predicted bounding box for each image is (13×13+26×26+52×52)×3= 10,647. However, a certain number of objects need a certain number of bounding boxes (e.g. 3 bounding boxes for 3 objects) and the final results will show 3 bounding boxes instead of displaying the total 10,647 bounding boxes. Two steps are applied to reduce the redundant predicted bounding boxes (Kathuria, 2018a). Firstly, bounding boxes with the calculated confidence (see section 3.4.1) under a certain predefined threshold would be ignored. Followed by a non-maximum suppression (Neubeck & Gool, 2006) to choose the final bounding boxes for the multiple detection problem. In the detection layer, instead of using softmax function to perform the classification, YOLOv3 uses logistic regression as the classifier to classify the detected objects. Logistic regression is widely used in the binary classification problem (the general format see Equation (3-3)), which will be used in the proposed model to predict the probability that the detected object belongs to a specific class.

$$p = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + ... + \beta_m x_m)}}$$
(3-3）

where, x_i is the explanatory variable, and β_i is the coefficient for the model.

3.4.2.4 Upsampling layer

As mentioned above, the YOLOv3 detects objects at three different scales; for this purpose, downsampling and upsampling processes are applied to resize the images. Downsampling is

carried out by changing the stride of the filters, while the upsampling process is realized by upsampling layers (green layers in Figure 3.9) in YOLOv3 network. Upsampling is a deconvolution operation, which aims at enlarging the image to the desired size for later processing. The two upsampling layers resize the image from 13×13 to 26×26 then to 52×52. Bilinear interpolation is employed in the upsampling process, and the detailed information about bilinear interpolation and bilinear upsampling can be seen in the research of Guerzhoy (2015).

3.5 Case study

3.5.1 Data collection

The CCTV inspection data was recorded in Edmonton, Canada, by EPCOR Drainage Service. In addition to the video material, the associated CCTV video assessment results are also utilized to label the defects of the video frame before training the YOLOv3 network. The CCTV inspection and assessment was conducted under the PACP code. CCTV crawler is controlled by the technologist and travels smoothly along the sewer pipe unless the technologist sees severe defects (e.g., hole and broken) or any taps. Minor defects such as crack and fracture will not stop the crawler, meaning the time window for the technologist to detect such defects is small, which requires the concentration of the technologist at all times when evaluating the CCTV video. This is another benefit to using an automated defect detection system to reduce the work intensity for the technologist.

The dataset contains 3,664 images with 4,056 unique defects. The images were retrieved from 63 CCTV videos recording the clay sewer pipe followed the procedure described in section 3.3.2. The distribution of the different types of defects is illustrated in Figure 11. Six types of defects and one type of construction feature are the detection targets of this research, namely, broken, hole, deposits, crack, fracture, root, and tap. The definition and description of each type of defect

and construction feature are retrieved from PACP (2015). Each type of defect is nearly evenly distributed. The dataset is randomly split into three subsets, namely, training set, validation set, and test set, whose proportions are 75%, 10%, and 15%, respectively. The proportional division is determined based on previous similar research examples (e.g., Cheng & Wang, 2018; Li et al., 2019) and our dataset condition. We chose to keep more samples in the training data set to provide more features for training and also to improve the testing accuracy. The detailed information regarding the dataset we are using for training the model can be seen in Table 3.2. All the data are preprocessed as described in section 3.3.2.
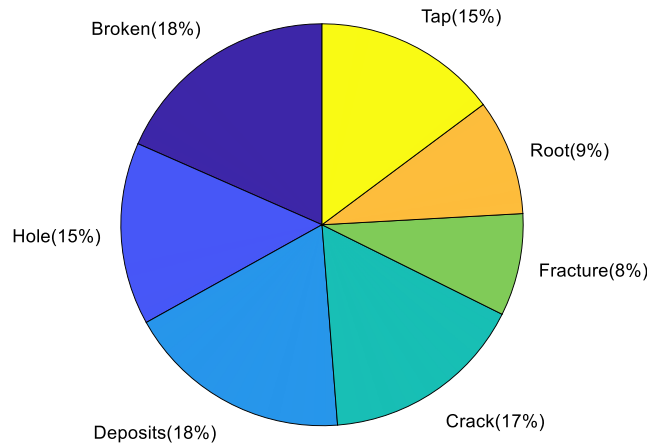
Figure 3.11: The distribution of each type of defect and construction feature of the data set.

Table 3.2: The number of samples used in each type of defect and construction feature of the data set.

	Training set	Validation set	Testing set
Broken	547	79	120
Hole	460	51	85
Deposits	556	74	106
Crack	495	77	94
Fracture	244	30	60
Root	294	35	49
Tap	449	55	96
Total	3,045	401	610
Percentage	75%	10%	15%

3.5.2 Deep learning model training and validation

Before we can use the model to label the CCTV video automatically, we need to train the object detection network first. During the training process, the training set and validation set are utilized as the input of the model. Throughout the training, we choose the parameters as per the standard YOLO model (a batch size of 64, a momentum of 0.9, and a decay of 0.0005) (Redmon et al., 2016). The value of loss function during the training process is presented in Figure 3.12. It took 17.5 hours in total to train 14,000 iterations. The final average loss is 0.233 with a learning rate of 0.0001. CCTV videos can be fed into the trained model to detect the defects automatically. The processing speed is around 33 FPS running on a GPU with RAM of 30000M. For example, for a test video with a length of 11.5 minutes (i.e. 20,701 frames), the total processing time is around 10.84 minutes. Figure 3.13 shows the speed for processing every frame of this video. Most of the frames can be processed at a speed of around 33 FPS; however, a small portion of the frames may require slightly more time for processing. Note that the processing speed is around 32.57 FPS on average for this video, given some extra time was spent on loading the trained weights of the YOLO model and other relative modules to run the algorithm.
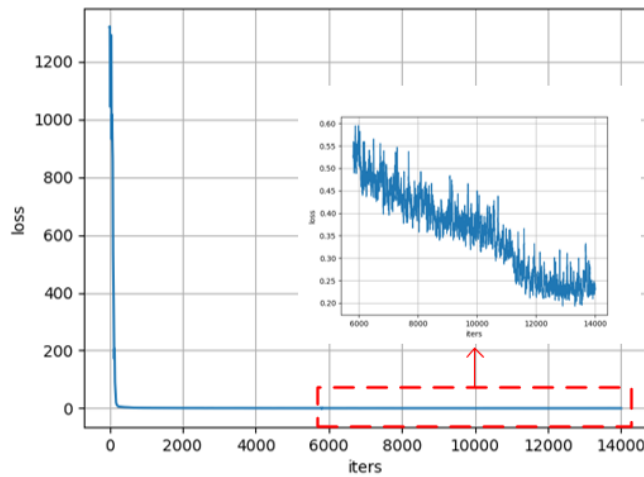
Figure 3.12: Value of loss during the training process.
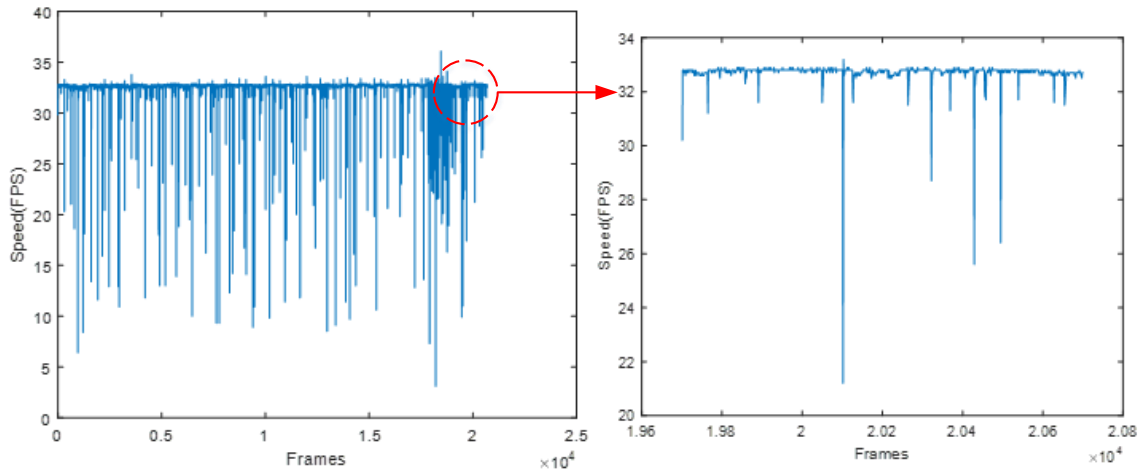
Figure 3.13: Example of processing speed for a CCTV footage.

The test set contains data that are new to the network and served the purpose of performance checking. Confusion matrix of the validation set and the test set are listed below (see Table 3.3 and Table 3.4) to describe the performance of the proposed object detection network. The bold numbers on the diagonal are the true positive detections. The closer the bold number to the total indicates better performance of the system. The definition of TP, FN, FP, TN can be seen in Table 3.5. In addition, F_1 is used to measure the performance of the classifier by calculating the harmonic mean of precision and recall, which has the best value at 1 and worst at 0 (see Equation (3-4) - (3-6)) (Hoang et al.,2018; Martínez-Rojas et al., 2015). The precision denotes the correct rate over all the detected objects, while the recall is the rate of completeness of detected objects among all the labeled objects. From Table 3.3 and Table 3.4, we can see that the system has better performance towards the defect types of broken and hole, whose F_1 is more than 0.9 both in the validation set and in the test set. From the test set, the system is bad at detecting crack fracture and root, whose F_1 is below 0.85. However, the lowest F_1 is still bigger than 0.8, which means the system can be said to have good performance in detecting defects of sewer pipes. The average F_1 scores are 0.876 and 0.882 for the validation set and testing set, respectively.

Table 3.3: Confusion matrix and F-measure of the validation set.

	Broken	Crack	Deposits	Fracture	Tap	Hole	Root	Miss	Total	Precision	Recall	F$_1$
Broken	**75**	0	0	0	0	3	0	1	79	0.987	0.949	0.968
Crack	0	**65**	2	0	3	0	0	7	77	0.878	0.844	0.861
Deposits	0	2	**60**	0	1	0	0	11	74	0.952	0.811	0.876
Fracture	0	1	0	**24**	0	0	1	4	30	0.889	0.800	0.842
Tap	0	5	1	3	**40**	0	1	5	55	0.889	0.727	0.800
Hole	1	1	0	0	0	**47**	0	2	51	0.940	0.922	0.931
Root	0	0	0	0	1	0	**27**	6	34	0.931	0.794	0.857

Table 3.4: Confusion matrix and F-measure of the test set.

	Broken	Crack	Deposits	Fracture	Tap	Hole	Root	Miss	Total	Precision	Recall	F$_1$
Broken	**111**	0	0	1	0	3	0	5	120	1.000	0.925	0.961
Crack	0	**78**	4	1	6	0	0	5	94	0.867	0.830	0.848
Deposits	0	2	**93**	1	1	0	0	9	106	0.921	0.877	0.899
Fracture	0	5	2	**46**	0	0	3	4	60	0.852	0.767	0.807
Tap	0	4	2	4	**79**	0	3	4	96	0.919	0.823	0.868
Hole	0	0	0	0	0	**82**	0	3	85	0.965	0.965	0.965
Root	0	1	0	1	0	0	**36**	11	49	0.857	0.800	0.828

Table 3.5: Definition of TP, FN, FP, TN.

		True condition	
		Positive	Negative
Predicted	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

$$\mathrm{Pr}\,ecision = \frac{TP}{TP+FP} \qquad (3\text{-}4)$$

$$\mathrm{Re}\,call = \frac{TP}{TP+FN} \qquad (3\text{-}5)$$

$$F_1 = 2 \times \frac{\mathrm{Pr}\,ecision \times \mathrm{Re}\,call}{\mathrm{Pr}\,ecision + \mathrm{Re}\,call} \qquad (3\text{-}6)$$

In order to measure the performance of the proposed algorithm, mAP has been calculated in this

research. The average precision (AP), as shown in Equation (3-5), is used to calculate the

average precision for different levels of recall (zero to one) for a specific class of object, while

the mAP is the average of all the classes. The equations to calculate the AP and mAP are presented in Equation (3-7) and Equation (3-8), respectively (Cheng & Wang, 2018; Everingham et al., 2009). The results of AP for each type of defect and mAP are tabulated in Table 3.6. The AP for crack and root is a little lower than that for other types of defects. The reason may be attributed to the sample size of some cracks and roots being too small compared to the sample size for other defects. However, the overall mAP is acceptable. A comparison of some recent research regarding the performance of defect detection is summarized in Table 3.7. Furthermore, the processing speed is one of the advantages of this proposed model, which enables the real-time processing capability for sewer pipe videos. This research focuses on seven types of targeted classes, which is more comprehensive than other research studies, which is another advantage. The accuracy is calculated based on a different dataset, by different metrics, and the number of classes is also different, which altogether makes it difficult to compare the accuracies directly. To sum up, the 85.37% mAP in this research is considered acceptable, but there is room for improvement in future research.

$$AP = \frac{1}{11} \sum_{r \in \{0,0.1,...,1\}} \max_{\tilde{r}:\tilde{r} \geq r} p(\tilde{r}) \qquad (3\text{-}7)$$

$$mAP = \frac{1}{N_{class}} \sum AP_i \qquad (3\text{-}8)$$

Table 3.6: Results of AP and mAP for the proposed system.

Class	Broken	Crack	Deposits	Fracture	Tap	Hole	Root	mAP
AP	96.27%	77.25%	81.26%	83.66%	88.38%	94.53%	76.24%	85.37%

Table 3.7: Performance comparison of different defects detection methods in recent research.

Reference	Defect type	Methods	Speed	Dataset sample size	Accuracy
This present study	Broken, hole, deposits, crack, fracture, root, tap	YOLOv3 algorithm	33 FPS	4,056	mAP of 85.37%
Hassan et al., (2019)	Longitudinal defect, debris silty, joint faulty, joint open, lateral, surface damage	CNN	N/A	47,072	F1 score of 96.33%
Li et al., (2019)	Deposit settlement, joint offset, broken, obstacles, water level stag, deformation	deep CNN and hierarchical classification approach	N/A	16,225	F1 score of 64.8%
Kumar et al., (2018)	Root intrusions, deposits, cracks	deep CNN	N/A	12,000	Accuracy of 86.2%
Cheng & Wang, (2018)	Root, crack, infiltration, deposit	Faster R-CNN	9.434 FPS	3,000	mAP of 83%

3.5.3 System implementation

After the model training and the performance checking, the object detector is ready to label the CCTV video. The video recorded on-site will be sent to the off-site office for assessment purposes. With the automated defect detection system, the video will be processed by the computer first before sending to the assessment technologist. Figure 3.14 shows some examples of the screenshot of the output videos of the defect detection system. Different types of defects are labeled with a different color box associated with the class name. Assessment technologists could watch through the video with the bounding boxes and defect names displayed at the same time. Technologists' knowledge and experience could correct some of the missing defects or wrongly detected defects. With the performance checking conducted above, the good performance of the system will ensure that technologists spend less time and energy on the video assessment job. Furthermore, the system will also work as a benchmark for the technologists. The video assessment requires high concentration over the whole process; therefore, the labeled defects work as a reminder for technologists when they may be carelessly distracted or not

familiar with a certain type of defect. In addition, the defect detection system could also work as an education system for training new technologists.
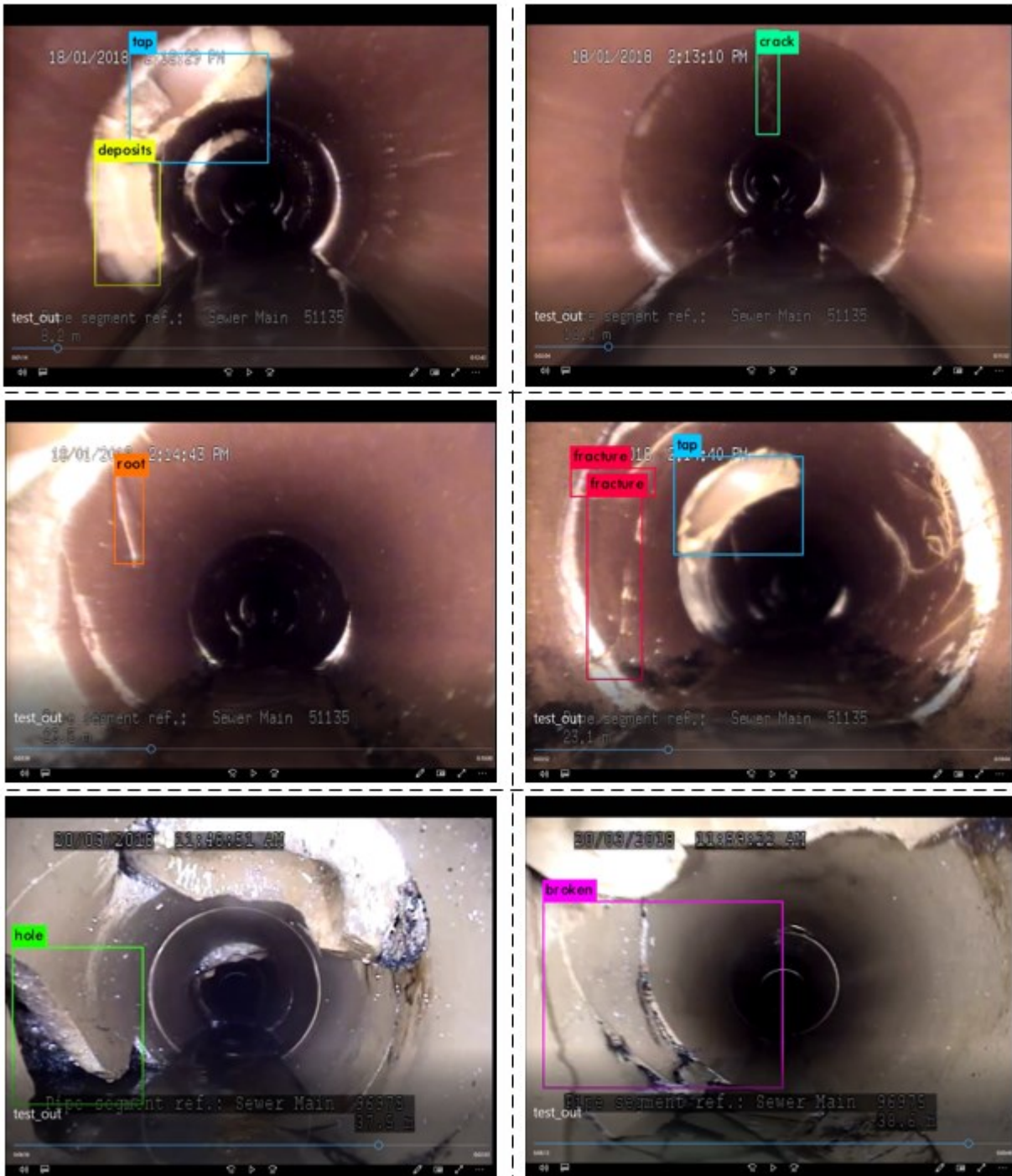
Figure 3.14: Identified defect types.

In addition to labeling the CCTV video, another output of the system is the labeled video frames that contain defects. Figure 3.15 shows an example of output frames containing defects. The left part is the output video frame, while the right part is the associate labeled CCTV video. The technologist can assess the extracted frames in succession. The confidence threshold is set at 90% (which is fixed after several experiments to get a relatively small number of frames). In the implementation example, for a CCTV video with a video time of 11.5 minutes, the total number of frames containing defects is 6,921. However, after filtering and selecting the frames, the final number of output frames is 126. The sample information of this example can be seen in Table 3.8, which records the defect type, confidence, the serial number of start to end frames, and the video time of start to end frames. For example, five consecutive video frames, which are from 463 to 467, contain defects of type crack with the highest confidence of 96%. This defect appears on the video from 15.43 to 15.57 seconds. The frame with 96% confidence can be outputted to represent this particular defect (crack) in the sewer pipe. For time-sensitive video assessment jobs, the technologist could simply check these 126 images to get a rough idea of the condition of the sewer pipe instead of watching through the whole video, which takes at least 11.5 min in the best-case scenario. Of course, evaluating only the selected frames would be an inaccurate and incomplete method by which to get the full information of a sewer pipe. However, it can be improved by designing a suitable selecting method, which is the next stage of our research. With further research, the pipe condition assessment results can be calculated from the output CSV file. However, this function needs more information rather than Table 3.8, such as the condition grade of each defect.

Figure 3.15: An output video frame that contains defects.

Table 3.8: Sample information of output frames.

Index	Defect confidence	Defect Type	Start frames	End frames	Video time starts (Seconds)	Video time ends (Seconds)
1	95%	Crack	463	467	15.43	15.57
2	92%	Root	520	520	17.33	
3	91%	Crack	537	537	17.90	
4	97%	Root	545	545	18.17	
5	99%	Crack	562	650	18.73	21.67
6	99%	Crack	672	672	22.40	
7	97%	Root	788	789	26.27	26.30
8	100%	Crack	836	879	27.87	29.30
9	94%	Deposits	1,119	1,119	37.30	
10	100%	Crack	1,121	1,138	37.37	37.93
11	94%	Deposits	1,222	1,222	40.73	
12	100%	Crack	1,237	1,350	41.23	45.00
13	91%	Fracture	1,346	1,346	44.87	
14	97%	Crack	1,425	1,428	47.50	47.60
15	90%	Root	1,465	1,465	48.83	
16	100%	Crack	1,502	1,646	50.07	54.87
17	98%	Deposits	1,539	1,599	51.30	53.30
18	93%	Fracture	1,613	1,613	53.77	
19	98%	Deposits	1,631	1,632	54.37	54.40
20	100%	Crack	1,790	1,822	59.67	60.73

3.6 Discussion

One of the contributions of this research is the customization of a state-of-the-art object detector to enable the real-time processing capability for the sewer pipe defect detection system, which had not been realized in previous research. The real-time processing capability makes it possible for the framework to use CCTV videos as input and to produce labeled videos as output. Another contribution is that the proposed system focuses on streamlining the information and data flow in the course of developing an automated real-time defect detection system with a deep-learning algorithm. Note that the framework not only focuses on the implementation of the YOLOv3 algorithm, it also emphasizes the data input and output process. In terms of input, a database manipulation process aimed at getting the input data efficiently is proposed as an element of the framework. As for the output, besides the labeled videos, frames containing defects are also provided automatically as outputs for the convenience of pipe condition assessment. In addition, CCTV videos can be translated into tabulated text format by taking advantage of real-time video processing capability, which had not been realized in previous research. This research aims to improve the efficiency and accuracy of the described CCTV assessment process by applying an automated defect detection system that is based on deep learning techniques. The framework presents the input data, object detector, and output data in a detailed manner and can serve as a guide for the development of a similar system.

The YOLOv3 network outperforms other similar algorithms in processing speed and accuracy according to the experiment conducted by the developer (Redmon & Farhadi, 2018). The major difference between YOLO and other region-based object detection algorithms is that YOLO uses a fewer number of fixed regions for detecting the object instead of using selective algorithms to output region proposals, which ensures that the YOLO network is faster in the object detection

process. In a YOLOv3 network, three different scales of object detection are performed in one detection process, together with residual block and other techniques to improve the accuracy of the network. The aforementioned reasons motivate us to choose it as the classifier in the proposed detection system. In this research, the model training is conducted by a remote supercomputer and the CCTV video can be processed by supercomputer or windows (preferred with GPU). The processing object is video in this research instead of images, therefore, the outputs are labeled videos and frames containing defects. The labeled videos and the assessment technologists serve as complements to each other, which is discussed in section 3.3.4. The F_1 score of the proposed defect detection system is greater than 0.8, and the mAP is 85.37%, which mean that the performance of the system is desirable. The output frames can give the technologist a quick overview of the condition of the sewer pipe; however, this function still needs to be refined in the next stage of our research.

In the future, with real-time video processing capability, the proposed system can be migrated to the CCTV collection platform, which can facilitate the on-site operator in the video collection process. For example, with the reference label shown on the monitors, the on-site operator (even with little experience discerning each type of defect) could stop the CCTV crawler at every targeted defect (which will be labeled automatically) to undertake a detailed inspection of the defect and use the CCTV crawler to record the condition of it.

3.7 Conclusion and future works

A framework to develop an automated defect detection system is proposed to label the CCTV video and output video frames that contain defects. YOLOv3 network is employed as the object detector in this research due to the advantages from the perspective of speed and accuracy. The model is trained and tested using a data set that contains 4,056 samples with six types of defects

(broken, hole, deposits, crack, fracture and root) and one type of construction feature (tap). The performance of the model is validated with an average F_1 score of 0.876 and 0.882 for the validation set and testing set, respectively. In addition, the mAP is 85.37%. Video frames containing defects are the other outputs of the system, which can provide an overview of the sewer pipe condition in a time-saving manner, compared with watching through the entirety of the CCTV video. The original contribution of this paper is twofold: 1) a state-of-the-art object detector has been customized to enable the real-time capability for the sewer pipe defect detection system; and 2) a framework has been proposed for developing an automated real-time defect detection system focusing on the data and information flow.

The function of output video frames needs to be improved in the next stage of the research. The objective is to extract a proper number of frames to represent the condition of the sewer pipe. The desired outcome would be a small number of output images that contain all the defects of the sewer pipe. In addition, the output of defect information in the CSV document needs to be enhanced by including more information (e.g., condition grade of the defects) in order to realize the automation of pipe assessment process. Another improvement needed is to improve the size of the training dataset to try to cover as many features of each type of defect as possible, which will lead to improving the performance of the model.

Chapter 4: AUTOMATION FOR SEWER PIPE ASSESSMENT: CCTV VIDEO INTERPRETATION ALGORITHM AND SEWER PIPE VIDEO ASSESSMENT (SPVA) SYSTEM DEVELOPMENT[3]

Preface

A defect detector has been developed in Chapter 3, which could process the raw videos collected on-site, and output with labeled videos, where pipe defects are labeled with different bounding boxes and the names of the defects are indicated above the bounding box (see Figure 3.14). However, the defect information is still embedded in the labeled video. Manual operations are still needed to record the types and locations of labeled defects to the maintenance database. To improve the level of automation of the CCTV assessment, this chapter presents the stage two research for fully automating the assessment process, which is an extended study of developing a defect detector (presented in Chapter 3). A novel video interpretation algorithm for sewer pipes (VIASP) is proposed to use the labeled video as the input in order to extract the useful information from the video, with the final output being the sewer pipe assessment report in textual format. In addition, a prototype of a sewer pipe video assessment (SPVA) system is developed to integrate all the previously developed automated functions into a user-friendly software program. The implementation of the proposed SPVA shows how the developed automation techniques can fit into the daily workflow of sewer pipe assessment work. The sewer pipe assessment reports generated by the SPVA record the health conditions of existing sewer pipes, which serve as the foundation for deterioration model development (see Figure 1.2), which is the next stage of preventive maintenance, and that is the main objective in Chapter 5.

---

[3] A version of this chapter has been submitted to the *Automation in Construction*.

4.1 Introduction

As the health condition of sewer pipes deteriorates as the pipes age, preventive maintenance must be carried out on a regular basis to keep the sewer pipes functional at all times (Fenner, 2000; Yang & Su, 2009; Yin et al., 2020b). CCTV inspection, which is the main techniques for preventive maintenance in this study, contains two main processes: on-site video collection and off-site video assessment (see Figure 4.1).
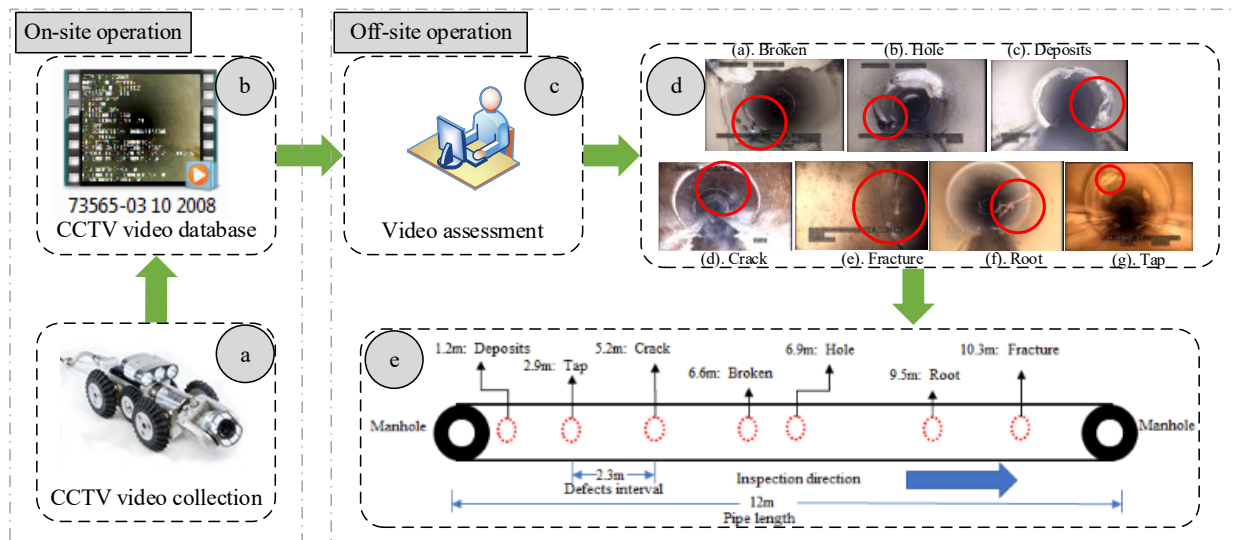
Figure 4.1: CCTV inspection workflow for preventive maintenance of sewer pipes.

The on-site crew is responsible for collecting CCTV videos of sewer pipes based on the work order issued by the maintenance planning office, as indicated in Parts (a) and (b) of Figure 4.1. Then the off-site assessment can be proceeded (Part (c) of Figure 4.2). As shown in Part (d) of Figure 4.1, examples of the defect types (and construction features) found in CCTV inspection video footage and their labels can include broken, hole, deposits, crack, fracture, root, and tap (construction feature). Based on the assessment results from the technologists, the sewer pipe inspection database can be constructed, where inspection information is stored for future maintenance planning purposes. The inspection information is presented graphically in Part (e) of Figure 4.1, which includes the type of the defect (construction feature) and its location

93

(distance to the start manhole in meters). In practice, the information pertaining to the defect is stored in a tabulated format in an Excel file or other database format (e.g., Microsoft Access). Table 4.1 shows a tabulated example of Part (e) in Figure 4.1. In reviewing the current literature, defect detection, as shown in Parts (c) and (d) in Figure 4.1, can be accomplished automatically by various methods, of which deep learning techniques have been explored extensively (e.g., Cheng & Wang, 2018; Li, Cong, & Guo, 2019; Yin et al., 2020c). However, to get the final output of the CCTV inspection as shown in Table 4.1, further operations are needed. For example, technologists must still record the detected defect type and its location into an Excel file. This research aims at taking the automation of the CCTV inspection process a step further by interpreting the CCTV video and compiling the output information in a tabulated format as shown in Table 4.1, which is to automate the process from Part (b) to Part (e) in Figure 4.1. To accomplish this, several challenges need to be overcome, and these are listed below.

1. In order for a defect detection algorithm to be considered effective, it must be outstanding in terms of both accuracy and processing speed. A highly accurate algorithm is the foundation for accurate video interpretation and a fast processing speed is what enables the algorithm to process video in real-time. The defects of the pipes will be labeled automatically by this defect detection algorithm, which will generate labeled videos as an output.

2. An interpretation algorithm must translate the information from labeled videos to tabulated textual information. Note that, to date, none of the research available in the literature claims that defect detection could achieve an accuracy of 100%. In fact, for certain images (i.e., video frames), the signal to noise ratio can be very low, which makes accurate detection and classification of sewer defects very challenging and at times

94

erroneous. This problem makes the idea of translating the information from labeled video directly impossible. This issue will be further discussed in Sections 4.3 & 4.4.1.

3. A text recognition algorithm must be able to extract the textual information contained in the video frame to identify the location of each detected defect or feature.

Many research efforts have been devoted to solving Challenge 1, which is to develop defect detection algorithms that are capable of detecting the defects within the video frames. State-of-the-art defect detection algorithms are reviewed in Section 4.2. In addition, our research group also developed a defect detection framework based on the algorithm called You Only Look Once (YOLO) (Yin et al., 2020c). The present study is built on the framework that we developed previously. There are also numerous studies targeting Challenge 3 (e.g., Hassan et al., 2019), which are trying to recognize the text information within the CCTV video frames. This present study takes advantage of the existing text recognition technology developed by Microsoft to test our proposed software. Challenge 2 has not previously been investigated in a thorough manner and is the main target of this study. The proposed algorithm to solve Challenge 2 is described and validated in Section 4.4. Existing literature pertaining to Challenges 1 and 2 are reviewed and summarized in Section 4.2.

Table 4.1: Example of the inspection information of a sewer pipe.

Defect type	Location
Deposits	1.2 m
Tap	2.9 m
Crack	5.2 m
Broken	6.6 m
Hole	6.9 m
Root	9.5 m
Fracture	10.3 m

4.2 Background

Recognized for its ability to process high-dimensional natural raw data (e.g., image, speech), deep learning has made substantial breakthroughs in many areas (e.g., image recognition, speech recognition, natural language processing) for the development of artificial intelligence (Lecun et al., 2015). With the development of deep learning technology, computer vision-based image processing has been adopted in various areas, such as object detection, face recognition, action and activity recognition, human pose estimation, etc. (Voulodimos et al., 2018), to improve the working productivity or level of automation. Object detection is used, in the context of defect detection for sewer pipe videos, to label the specific objects (e.g., crack, fracture, holes, etc.) that are typically identified by technologists, and further to automate this tedious and time-consuming process. Various frameworks have been adopted to facilitate object detection, among them, convolutional neural network (CNN) is the most widely used (Krizhevsky et al., 2012). In recent years, several frameworks built on the original CNN structure have been aimed at improving the performance of the CNN, such as region-based CNN (R-CNN) (Girshick et al., 2014), Fast R-CNN (Girshick, 2015), SPP-net (He et al., 2015), Single Shot Detector (SSD) (W. Liu et al., 2016), YOLO (Redmon et al., 2016), Faster R-CNN (Ren et al., 2017), Mask R-CNN (He et al., 2020), etc. These various algorithms sought to improve detection accuracy and detection speed. Interestingly, these advanced object detection algorithms that have been investigated and developed by computer scientists are well adapted to solving practical problems occurring in construction engineering and management, such as construction safety and personnel monitoring, resource tracking, activity monitoring, surveying, as-is modeling, and inspection and condition monitoring (Martinez et al., 2019).

For the specific area of defect detection that is CCTV videos for sewer pipes, several studies have been carried out in the past few years. For example, Cheng & Wang (2018) have adopted Faster R-CNN for the defect detection of sewer pipes and have achieved an mean average precision (mAP) of 83% for detecting roots, cracks, infiltration, and deposits; Kumar et al. (2018) have employed a deep CNN to detect root intrusion, deposits, and cracks by using a dataset that contains 12,000 images, and reaching an accuracy of 86.2%; and Li et al. (2019) have used deep CNN and hierarchical classification approach to classify deposit settlement, joint offset, broken, obstacles, water level sag and deformation within the sewer pipes. Hassan et al. (2019) have used a CNN together with text recognition to generate defect reports based on keyframes (i.e., the frames that contain targeted defects) of the CCTV video, and Yin et al. (2020c) have proposed a deep learning-based framework to realize real-time defect detection for CCTV videos of sewer pipes, where YOLOv3 (Redmon & Farhadi, 2018) is used as the defect detector. Information/data flows for such a framework have been streamlined, which could be a benchmark for future defect detection system development. Among them, all the studies focus on processing specific frames of the video, which is different from the video interpretation process. Note that video interpretation is different from processing every frame of the video, since there are further operations (e.g., noises need to be excluded since the deep learning algorithms cannot achieve one-hundred percent accuracy) need to be conducted before outputting the video information. The challenges and mechanism of CCTV video interpretation will be described in Section 4.4.1.

Video is a type of media that is much more information-intensive compared to images; however, fewer research efforts have been devoted to the processing (e.g., classification) of videos directly (Karpathy et al., 2014). Although a video is essentially a stack of images, the information

delivered in video format is different from static images since there is more contextual information. In the area of construction, research efforts have been made in terms of retrieving information from on-site surveillance videos. For example, Son et al. (2019) have proposed a real-time warning system to prevent potential collisions between workers and heavy equipment based on real-time surveillance videos, where Faster R-CNN is used to detect the workers and 3D estimation is used to estimate the distance from workers to the equipment; Roberts & Golparvar-Fard (2019) proposed an innovative method to track and analyze earthmoving activities based on videos using multiple machine learning algorithms, such as CNN, hidden Markov model, Gaussian mixture model, and support vector machine; and Gong & Caldas (2011) proposed an intelligent video interpretation system for the productivity analysis of construction operations. The abovementioned studies show that various types of valuable information can be retrieved from videos; however, in the sewer pipe maintenance process, few studies have focused on collecting information from CCTV videos directly, which is essential in the context of automating the process from Part (b) to Part (e) as shown in Figure 4.1. Thus, the present research focuses its efforts on this research gap, which is to take the automation a step further than image processing-based defect detection to interpret the CCTV video and translate the information from video format to text format. The next section will describe the proposed method in detail.

4.3 Methodology

Figure 4.2 shows the proposed methodology to automate sewer pipe assessments based on CCTV videos. The inputs are the videos collected on-site, shown as Part (a) of Figure 4.2, that show the inside condition of the sewer pipes. The intent of the methodology is twofold: 1) to generate textual information, which is in Excel in this case as shown in Part (g) of Figure 4.2,

that records the essential information for sewer pipe assessment purposes, and 2) to develop user-friendly software that includes all the developed functions. For this purpose, several operations need to be conducted from Part (b) to Part (h) of Figure 4.2. The videos must be processed by a defect detector, shown in Part (b) of Figure 4.2, to label the defects that appear in the CCTV video. In the present research, we use the defect detector developed by Yin et al. (2020c), which is based on YOLOv3. The adopted defect detector has the ability to accomplish detection in real-time, wherein the video processing speed is around 33 frames per second (FPS). The performance of the defect detector is also good in terms of accuracy (e.g., mAP of 85.37%) in comparison with the performance results of other similar research studies, which can be found in more detail in the study by Yin et al. (2020c). The developed defect detector is able to detect defects including broken, hole, deposits, crack, fracture, root, as well as taps, which are a construction feature of sewer pipes.
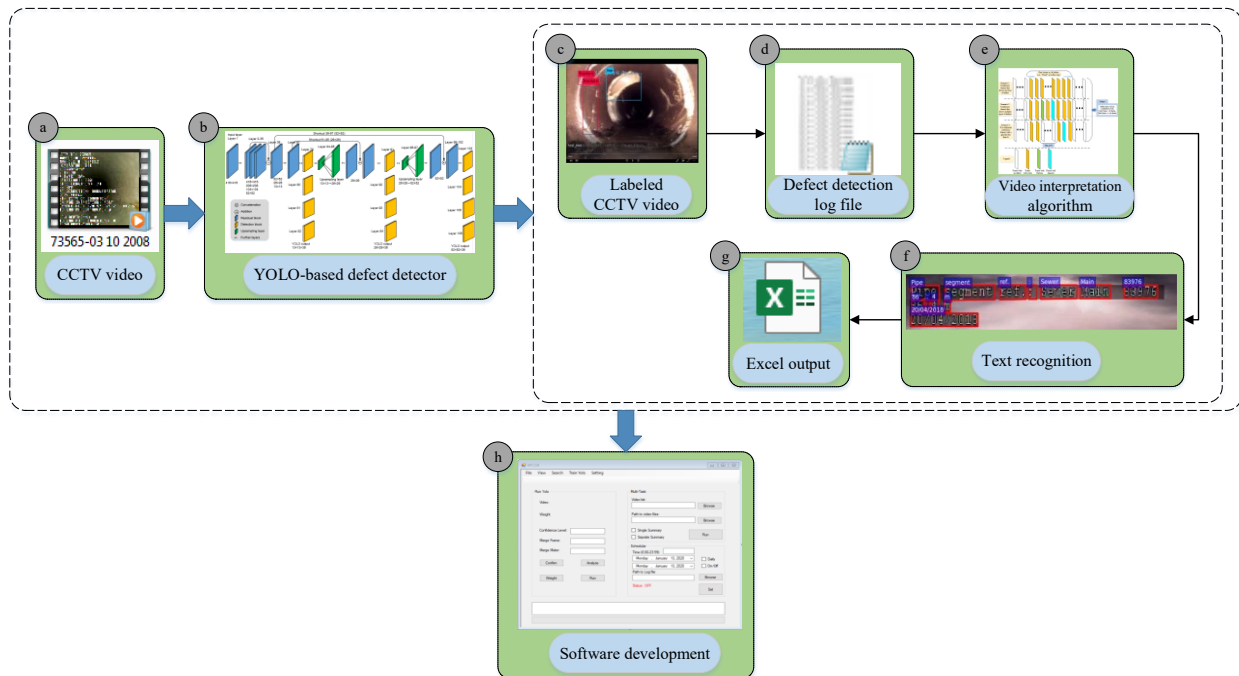
Figure 4.2: The methodology for the automation of sewer pipe assessment based on CCTV video inspection.

After being processed by the defect detector, the defects within the video will be appropriately labeled with bounding boxes surrounding each of these, and the labeled video, shown in Part (c) of Figure 4.2, is the output, which is typical of computer vision-based defect detection studies for sewer pipes. This research takes this a step further by proposing a video interpretation algorithm for sewer pipes (VIASP) that summarizes the findings of the defect detector in a format that can be useful to practitioners. Figure 4.3 shows a screenshot of a defect detection log file. The frame index, which shows the number of frames, is highlighted in the red box. For each detected defect, the defect type and the confidence level (expressed as a percentage) of the defect detection are also presented below the frame index. The confidence level indicates how much confidence the defect detector has when assigning a specific feature to a predefined category (i.e., deposits in this example) (Yin et al., 2020c). From the log file shown in Figure 4.3, it is clear that the defect detector finds a deposit from frame 1395–1400, and another deposit on frame 1403. There are no defects detected on frames 1340 and 1341. Note our tested video has a property of 25–30 FPS, which means 1 frame will only appear for less than 0.04 s in the video. There is a high degree of probability that these are the same deposits that appear in the video from frames 1395–1403. Some of the possible reasons for the discontinuity at frames 1340 and 1341 are listed below:

1. No defects appear in these two frames. However, it is nearly impossible for such a small amount of time (i.e., approximately 0.08 s) pass between two individual defects since the camera is travelling at a speed of approximately 9 m/min (15cm/s), which is the recommended speed from PACP code. Note that the actual traveling speed of the camera truck in our case study is even faster than 9 m/min, achieving at 23.6 m/min, according to one of the research that conducted by our research group (Yin et al., 2020a).

2. The perspective from which the defect is viewed by the camera changes as the camera moves forward. In this case, these two frames with different camera angles may cause the defect detector to miss the defect, which could occur because these defect or features are unfamiliar to the defect detector at those particular camera angles.

3. Mistakes made by the detector. As mentioned above, automated defect detection cannot ensure 100% accuracy, and there is always the possibility that the detector will make some mistakes of by either missing some defects, wrongly classifying a defect (e.g., classify a crack as a root), or raising a false alarm (i.e., reporting a defect on the frame when there is not a defect present).

4. The quality of the video changes during this short period of time. The quality of the video can be influenced by many factors such as lighting conditions, water that submerges the camera, mist or fog due to the temperature difference between inside the pipe and the ground, etc. Therefore, it is difficult to ensure the video footage is captured under the exact same conditions from beginning to end for each sewer pipe.

Given that so many factors can cause noise that hinders the performance of automated defect detection from CCTV videos, it is practically impossible (with currently available technology) to develop a perfect defect detector that makes no mistakes when analyzing video frames. To output the detection results frame by frame will generate noises, duplications, or false alarms in this case. Therefore, a VIASP is proposed to process the original output from the defect detector in a log file format (see Figure 4.3) into output in tabulated format (textual) that contains information that is useful for the sewer pipe assessment department. The VIASP is described in more detail in Section 4.4.

```
 cvWriteFrame                             cvWriteFrame
[2J[1;1H                                  [2J[1;1H
FPS:32.4                                  FPS:32.5
Objects:                                  Objects:

there is a defect_frame!!!               there is a defect_frame!!!
1395 <- frame with defect index          1399 <- frame with defect index
deposits                                 deposits
: 97%                                     : 98%

 cvWriteFrame                             cvWriteFrame
[2J[1;1H                                  [2J[1;1H
FPS:32.5                                  FPS:32.5
Objects:                                  Objects:

there is a defect_frame!!!               there is a defect_frame!!!
1396 <- frame with defect index          1400 <- frame with defect index
deposits                                 deposits
: 94%                                     : 66%

 cvWriteFrame                             cvWriteFrame
[2J[1;1H                                  [2J[1;1H
FPS:32.6                                  FPS:32.5
Objects:                                  Objects:

there is a defect_frame!!!                cvWriteFrame
1397 <- frame with defect index          [2J[1;1H
deposits                                  FPS:32.7
: 97%                                     Objects:

 cvWriteFrame
[2J[1;1H                                   cvWriteFrame
FPS:32.5                                  [2J[1;1H
Objects:                                   FPS:32.6
                                          Objects:
there is a defect_frame!!!
1398 <- frame with defect index          there is a defect_frame!!!
deposits                                  1403 <- frame with defect index
: 98%                                     deposits
                                          : 66%
 cvWriteFrame
[2J[1;1H
FPS:32.5
Objects:
```

Figure 4.3: Example of the log file with defect detection information.

After the log file generated by the defect detector is processed by the VIASP, a text recognition

function, or optical character reader (OCR), is needed to extract the textual information from the

video frame that indicates the location to determine the distance of the detected defect from the

starting manhole, as illustrated in Part (e) of Figure 4.1. This function has been investigated by

Dang et al. (2018) by using Tesseract OCR to extract the text information (which included words

in Korean that record the defect type and numbers that record the defect distance) from the

CCTV video for sewer pipes. In their research, both the distance information and the defect type

are included in the text information of the CCTV video, which differs from the video being

processed in the present study since only distance information is included in the video footage under study and the defect type is determined by technologists working at the off-site assessment office (see the maintenance workflow presented in Fig.1). Since there is no defect type information presented in text format in the video frame, instead the defect is automatically detected and labelled by the defect detector. In the present study, we use an OCR developed by Microsoft Azure (which is a cloud computing platform) to detect the text information within the video frame. The adopted OCR has a detecting accuracy of 90% on our dataset based on a prior experiment with 184 randomly sampled video frames. In this prior experiment, the mistakes largely occurred in one specific video where the color of the text was very close to the colour of the background, the text in these frames may have been missed by the OCR. This problem can be improved by detecting the missed frames for multiple times or replacing them with their adjacent video frames. With this implementation mentioned above, the detecting accuracy could increase to approximately 95%. Note that the text recognition function is functionality that could be integrated into the proposed sewer pipe video assessment (SPVA) system. Future developers are free to use any functional OCR (e.g., Tesseract, Microsoft Azure, etc.) in the development of their software. In the present study, we chose to use Microsoft Azure since this OCR has been well investigated by previous researchers and useful tools have been developed and adopted to extract the text information contained within CCTV videos for sewer pipes (Dang et al., 2018). The present research will focus on the development of VIASP and the overall SPVA system. With all the functions mentioned above, as shown in Parts (a)–(f) of Figure 4.2, the information extracted from each step will be integrated and formatted as output in an Excel file, as per Part (g) of Figure 4.2.

The final step is to package all the functions from Parts (a)–(g) of Figure 4.2 into a user-friendly software, which is the proposed SPVA system. In addition to integrating all the functions including defect detection, video interpretation, and text recognition, the development of SPVA also takes into consideration the maintenance department's management process to develop several useful functions in our case study. Based on the SPVA system developed for our case study, other developers could easily customize the system to fit their particular needs. The SPVA system is described in detail in Section 4.5.

4.4 CCTV video interpretation

4.4.1 CCTV video interpretation algorithm

A non-perfect deep learning algorithm could cause problems when trying to analyze the raw output of the defect detector. As in the example mentioned in the methodology section, any mistakes made by the algorithm could cause noise in the output. Theoretically, frames 1395–1403 could be considered as one defect as mentioned above (see Figure 4.3); however, the two discontinuous frames make the situation more complicated. Instead of outputting one defect from frames 1395–1403 as one deposit, the raw output will be "deposit from 1395–1400", and "another deposit on frame 1403". In fact, one defect typically does appear in the video for approximately 1–4 seconds as the camera is moving forward, which will result in approximately 25–120 frames that contain the same defect. There is a possibility that the non-perfect defect detector makes several mistakes for these continuous frames due to the reasons previously discussed in the methodology section. If a random number of mistakes occur at random frames in a continuous series of frames, the one defect will be cut into little pieces, instead of showing as one defect that appears in continuous frames. For example, Figure 4.4 presents an example of the consequence of this noise that can then be found in the raw output of the defect detection log file.

There are cracks that are detected in frames 100–108, 110, 113–140, one fracture on frame 109, one deposit on 112, and no defects on frame 111. For a video with a frame rate of 25 FPS, each of those frames in Figure 4.4 will appear in the video for 0.04 s. In fact, considering the short time period during which these video frames appear on the screen, it is likely that from frame 100 to frame 140, there is only one defect, i.e., crack. Figure 4.4 illustrates an extreme example that would occur very rarely considering the accuracy of the defect detector that we employed in the present study. Note that even if the defect detector has an excellent detection accuracy, the unpredictable underground conditions during video collection process will decrease the performance of the defect detector to some extent. Possible factors that influence the performance of the defect detector are mentioned in the methodology section, such as a change in video quality or change in camera angle, etc.
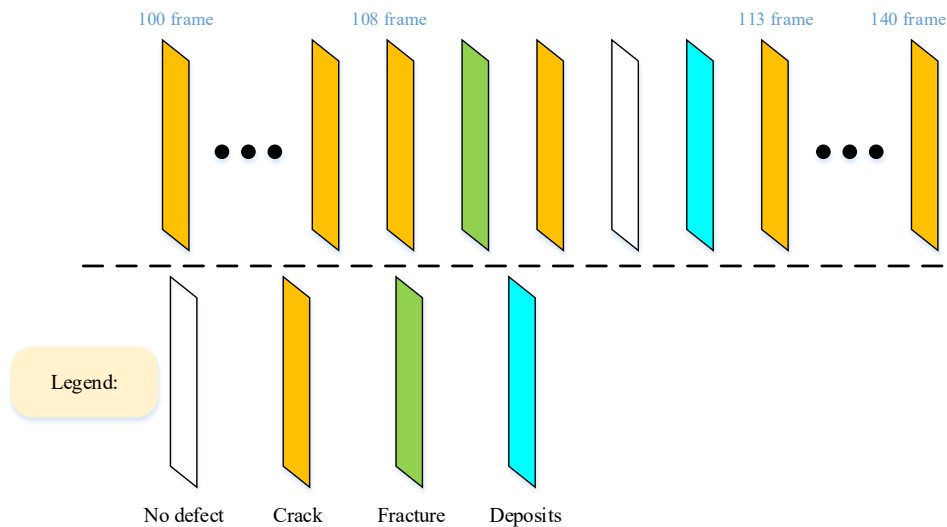
Figure 4.4: An example of continuous frames with noises generated by the defect detector.

Several common scenarios found in the labeled CCTV videos are shown in Figure 4.5. Scenario 1 is the ideal situation where continuous frames are detected with only one type of defect, which can easily result in output indicating there is a defect, which is crack in the case of Scenario 1 in Figure 4.5, from frame i to $i+n$. Scenario 2 is more complicated than Scenario 1 since there is

noise in the series of frames, which could result in output indicating multiple different defects as described in Figure 4.4. Scenario 3 is the most complicated situation in which a small number of no defect frames (or frames that contain defects with a low confidence level) together with some noise are included in the series of continuous frames from frame i to $i+n$. The gap (frames with no defects) could split the continuous frame cluster into smaller pieces, which will increase the level of difficulty for interpreting the video. In an ideal situation, all the scenarios presented in Figure 4.5 should result in the same output, which is that there is a crack from frame i to $j+n$. In order to accomplish this, VIASP is proposed to exclude the noise and merge frames that show the same defect.

Before describing the VIASP in detail, the concept of the defect frame cluster (DFC) is proposed:

- For a certain number of continuous frames, defects (that could be of any type) are detected in almost all these frames. If the frame gap (g) between those adjacent frames that do contain defects is less than a predefined parameter G (see Figure 4.5), then consider the continuous frames that come before and after the gap together as one cluster.

Based on the proposed concept, the three scenarios presented in Figure 4.5 contain one defect frame cluster (DFC) each. The output of the VIASP is the information about this DFC, such as defect type, the average confidence level of the DFC, and start frame and end frame of the DFC. The defect type for the DFC is the defect type that appears most frequently in one DFC (majority rule). For example, for a DFC with 100 frames, of which 95 are detected with a crack and 5 are noise frames (either with other types of defects or are frames with no defects), then the defect type for this DFC is determined as crack. The average confidence level of this DFC is calculated based on those frames that show the same type of defect within the DFC.
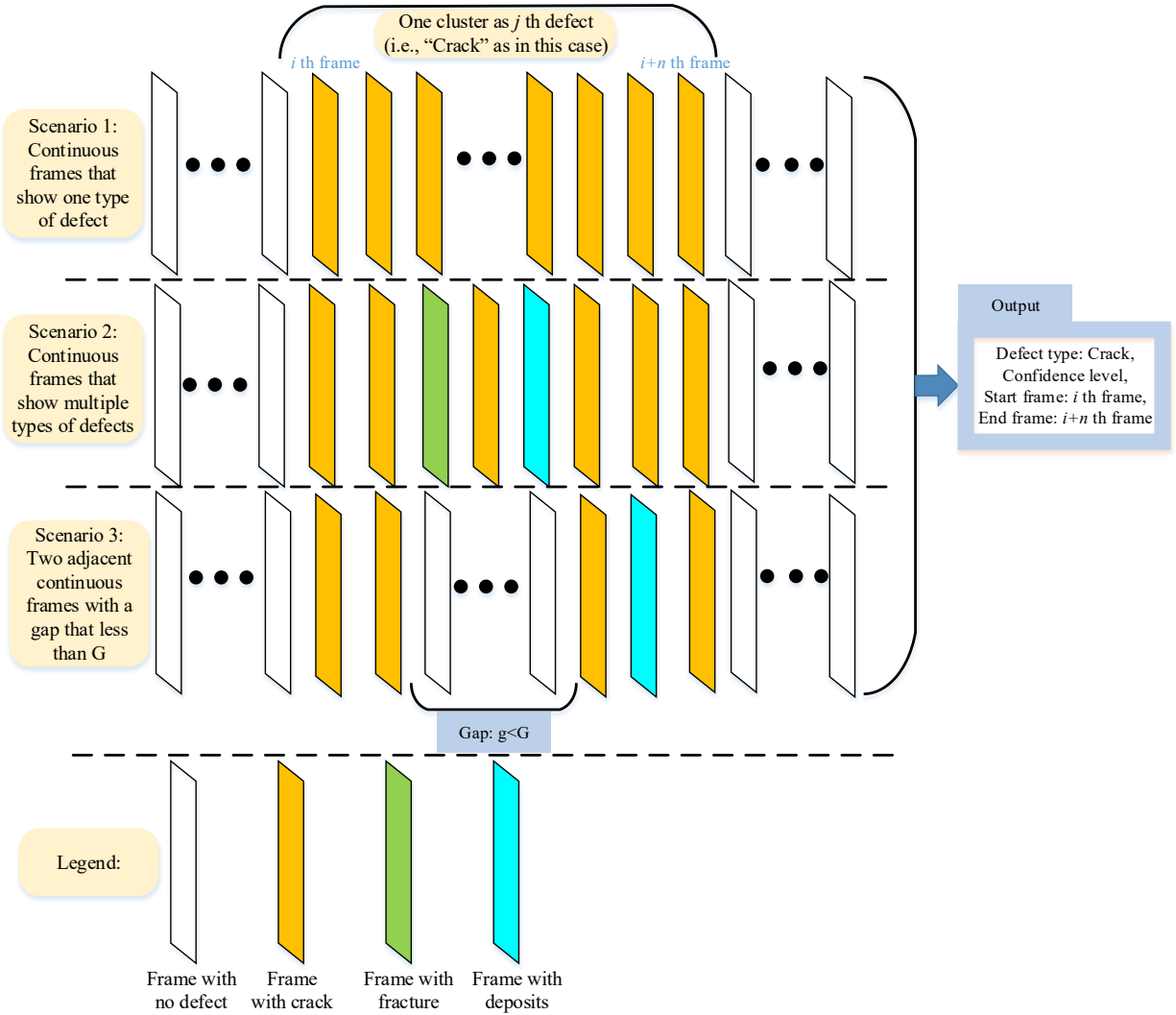
Figure 4.5: Common scenarios of labeled frames in CCTV video.

The pseudo-code for the VIASP is presented in Figure 4.6. The algorithm starts by defining the initialized variables and the human predefined variables. The initialized variables include the counting variable for frames (i) and the counting variable for the DFC (j). The predefined parameters are merge gap (G) (see G in Figure 4.5), which is used for determining whether discontinuous frames are within the former DFC; confidence level C_1, which is used to filter out all the frames that are detected with defects and have confidence levels lower than C_1; confidence C_2, which is used to filter out all the DFCs that have confidence levels lower than C_2. The determination of these predefined parameters is discussed in Section 4.4.2. The log file

generated by the defect detector needs to be loaded to VIASP to extract the information about the

frames' indexes, detected defect types, and confidence levels for the defects (Step 2 of Figure

4.6).

1. Start:
    a) Initialized variables: i=1 (ith frame), j=1 (jth defects);
    b) Predefined variables: merge gap (G), confidence 1 (C$_1$) and confidence 2 (C$_2$);
2. Read the log file generated by YOLO;
3. Assign an index (T$_i$) for all the frames:
    a) If there is no defect: T$_i$ =0;
    b) If there is a defect, then:
        i)   If the confidence level is smaller than C$_1$ (c<C$_1$), then T$_i$ =0;
        ii) If the confidence level is bigger than C$_1$ (c>C$_1$), then: Broken: T$_i$ = 1; Hole: T$_i$ = 2; Deposits: T$_i$ = 3; Crack: T$_i$ = 4; Fracture: T$_i$ = 5; Root: T$_i$ = 6; Tap: T$_i$ = 7;
4. While i <= total frames:
    a) Find the 1st T$_i$>0, and count it as the 1st defect, j=1;
    b) Find the DFC (within the predefined merge gap G) that include this frame;
    c) Output defect type for jth DFC
        i)   If there is T$_i$=7 among the DFC, then define the type of jth DFC as "Tap"
        ii) If there isn't T$_i$=7 among the DFC, then define the type of jth DFC as the most frequent defect type (DT$_j$) among the cluster;
    d) Calculate the average confidence level for all the defect= DT$_j$ in this cluster as the confidence level (CL$_j$) of jth DFC;
    e) If the confidence level (CL$_j$) of jth DFC is smaller than C$_2$, then discard it; Otherwise, keep it;
    f) Record defect type, confidence level, start frame, and end frame for jth defect;
    g) Find the next T$_i$>0, and j=j+1;
5. Format the output in the EXCEL file;
6. End

Figure 4.6: Pseudo-code for the VIASP.

After Step 2, the first-round confidence level filtering (see Step 3 of Figure 4.6) aims to filter out

all the frames that have a lower confidence level. For the frames with no defects and the frames

with low-confidence-level defects, the VIASP assigns an index $T_i = 0$ to them, which means

there are no defects detected on the ith frame. The results of the frames will be assigned with an

index of T_i according to different types of defects. For example, if there is a crack detected on

$i\,th$ frame, then a value of 4 is assigned, i.e., $T_i = 4$.

After assigning an index of T to each frame of the video, the next step is to find the DFC of the video (see Step 4). To find the DFC, an important parameter, merge gap (G), is needed to determine whether the next frame with a defect after a gap (a gap of continuous frames with no defects) is within the DFC or not (for a graphical representation, see Figure 4.5). After a continuous series of frames with no defects, the number of frames within this gap are counted, and if the count number is lower than G, then the next frame is within the DFC (j^{th}), otherwise the frame with defects is in the next DFC ($(j + 1)^{th}$).

After finding all the DFCs in the video, the type of each DFC needs to be determined based on the majority rule mentioned above. One exception is that if the DFC includes any taps (any frames for which $T_i = 7$), then the DFC type will be determined as a tap. The reason for this stems from the CCTV recording process for tap (regulated by PACP) in our case study: For a normal CCTV recording process, the crawler that is equipped with a camera moves at a constant speed within the sewer pipe unless it encounters a tap (or very severe defects), in which case the crawler will stop and move the camera around to conduct a detailed inspection of the tap (Yin et al., 2019). Figure 4.7 shows the process of inspecting tap A, where the crawler needs to stop beside the tap and turn the camera to the tap to conduct a detailed inspection. This detailed inspection of each tap will result in a much longer recording time for a tap than other type of defect, which leads to a larger number of video frames for each tap (e.g., 250 frames for a 10 s inspection). The higher the number of frames, the higher the possibility of there being false detected frames. On the other hand, when the camera rotates to face toward the tap, different views of the tap (from different camera angles) could mislead the defect detector, which will also result in some noise during the relatively long duration of the inspection of the tap. In addition, the content within the tap can cause false alarms (those defects not belonging to the pipe) by the

defect detector. Note that the interest in this work is the main line only, not what happens in secondary lines linking the main sewer line to private property. Thus, the defects within the tap are usually not recorded by the technologist in the sewer pipe assessment. For example, the defects within tap A shown in Figure 4.7 will not be recorded in the manual sewer pipe assessment report; however, if there are images of defects within tap A, the defect detector will detect them, and a record of these defects also will be shown in the log file. If the camera remains for a longer period of time at the tap to record a defect located within the tap, then there will be higher number of frames of that defect type as compared to tap in the DFC. In this case, the majority rule will not work. Therefore, if taps are detected in any frames within a DFC, then with high confidence we can say that this DFC is a tap. The experiments in Section 4.4.2 show that this special rule has not resulted in any mistakes in terms of categorizing taps as other defects and vice versa.
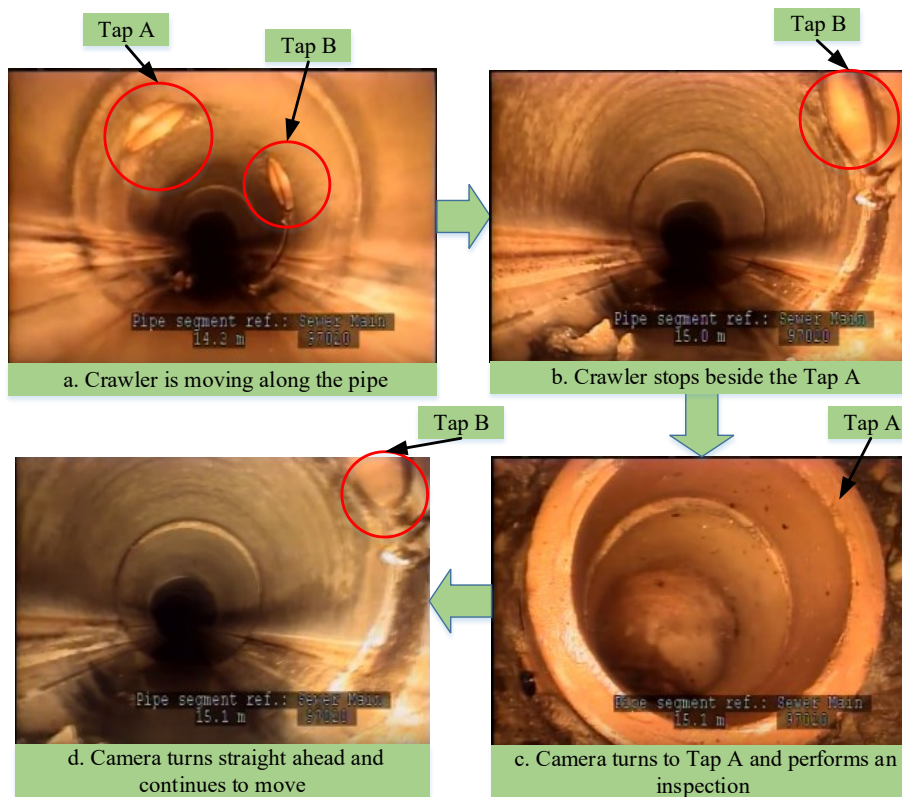
Figure 4.7: CCTV recording for tap within the sewer pipe.

Once identified as a DFC, the confidence level of the DFC needs to be calculated using the rule specified in Step 4-d of Figure 4.6. The following step is the second-round confidence level filtering process to filter out any DFC with a confidence level lower than C_2 to avoid duplications. If the DFC survives after this filtering, then the information pertaining to this DFC (defect type, confidence level, start frame, and end frame) needs to be outputted. The abovementioned steps will repeat until all the frames within the video are checked and all the DFCs are determined. The final step is to format these outputs and write them to an Excel file. The next section will describe how to determine the appropriate predefined parameters (i.e., C_1, C_2, and G) to obtain optimal output from VIASP and how the performance of VIASP is validated.

4.4.2 Experiment and validation

4.4.2.1 Objective function construction

Three parameters are of vital importance to VIASP: C_1, C_2, and G. These parameters are the key variables used to find the appropriate DFCs. Based on the logic of the algorithm, if C_1 or C_2 increase, then there will be fewer DFCs in the video since a higher confidence filter will block more frames or DFCs with a lower confidence level. The same result could occur if G increases since a higher value for G could make each DFC include more frames, which would result in fewer DFCs in the final output. In an ideal situation, with optimal values for C_1, C_2, and G the output of the VIASP should cover all the defects within the sewer pipe and without any false alarms, which means, if the manual assessment conducted by a technologist is one-hundred percent accurate, a perfect VIASP would produce output that exactly matches the manual assessment. However, it is difficult to achieve one-hundred percent accuracy by either method (manual assessment or VIASP) due to various reasons. For example, there exists an inconsistency in the manual assessment method since different technologists may have slightly

different criteria according to which they categorize those defects that are closely related (e.g., crack and fracture). As for the VIASP, the imperfect defect detector is the reason the VIASP does not have perfect accuracy. Therefore, the objective is to select values for C_1, C_2, and G that make the output of the VIASP as close to the manual assessment report as possible since the manual assessment is the only existing dataset against which the results can be compared and validated. To realize the objective, a searching method needs to be carried out to find the optimal combination of these three predefined parameters. To match the output of the VIASP with the manual assessment report directly is difficult at this stage, and the reason for that is described in the following section. If matching the defects outputted by the VIASP and the assessment report is impossible (or cost-prohibitive), we can match the number of DFCs outputted by the VIASP and the number of defects in the manual assessment report to facilitate finding the optimal predefined parameters, C_1, C_2, and G. Therefore, our objective function for the searching algorithm is to minimize the difference between the number of the DFCs outputted by the VIASP and the quantity of defects in the manual assessment report. The searching algorithm for finding the predefined parameters is described in the following section (see Section 4.4.2.2). The performance of the VIASP with the selected predefined parameters is also tested in our case study to ensure the algorithm is accurate (see Section 4.4.2.3).

The comparison of the output of the VIASP with the manual assessment report cannot be conducted automatically at this stage since the location of the defect identified by the VIASP is different from the location indicated in the manual assessment report. For example, if there is a defect at 2 m from the starting manhole, the technologist conducting the manual assessment will wait until the defect has just disappeared from the screen before recording the location since this is the location at which the defect is right beside the crawler. However, the VIASP will output

the starting frame and the end frame of the DFC representing the same defect. The location (in meters) shown on the starting frame is different from the real location of the defect since the meters indicated on that frame is ahead of the real location of the defect. As for the end frame of the DFC, this also varies from the real location in certain circumstances. As the camera is moving, the defect will disappear from the screen gradually, and during the process, the perspectives at which the defect is viewed keep changing, which may lead to the end frame of the DFC not aligning with the very last frame before the defect disappears from the video completely. In addition, sometimes two adjacent defects are very close to each other, which makes it difficult to discriminate one from the other. In summary, it is difficult to match a defect found by way of the VIASP with the defect found via manual assessment automatically. Therefore, the present study proposes to compare the number of defects identified by VIASP with the number of defects identified by technologists conducting a manual assessment instead of matching the output of the VASP with the manual assessment report directly. The objective function is presented in Equation (4-1).

$$\text{Obj: } \min f(C_1, C_2, G) = Num_{DFC}(C_1, C_2, G) - Num(defects_{manual}) \quad (4\text{-}1)$$

where $f(C_1, C_2, G)$ is a function used to calculate the difference between the number of defects identified by VIASP and the number of defects contained in the manual assessment report. $Num_{DFC}(C_1, C_2, G)$ is the number of defects identified by VIASP with the predefined parameters of $C_1, C_2,$ and G. $Num(defects_{manual})$ is the number of defects in the manual assessment report for our testing dataset, which is a constant number once we select a dataset.

The objective function is subject to Equations (4-2) – (4-6):

$$Num_{DFC}(C_1, C_2, G) - Num(defects_{manual}) \geq 0 \quad (4\text{-}2)$$

$$0 \leq C_1 \leq 100 \quad (4\text{-}3)$$

113

$$0 \leq C_2 \leq 100 \qquad\qquad (4\text{-}4)$$

$$0 \leq G \leq 300 \qquad\qquad (4\text{-}5)$$

$$C_1 - C_2 \leq 0 \qquad\qquad (4\text{-}6)$$

Equation (4-2) shows that the $Num_{DFC}(C_1, C_2, G)$ is greater than $Num(defects_{manual})$, which means that the number of DFCs identified from VIASP should not be less than the number of defects in the manual assessment report. The reason to set this constraint is that the risk of false alarm is lower than the risk of missing a defect, thus, the VIASP tries to cover as much of the defects as possible, while outputting fewer false alarms. The constraints for C_1 and C_2 range from a no-confidence threshold (confidence level equal to 0) to the highest confidence threshold (confidence level equal to 100). The merge gap (G) is set to 0–300, which is 0–12 s of video time considering the frame rate propriety of the video (25–30FPS), which means that the DFC could have a duration in the range of 0–12 s. The constraint for G is manually set to be a larger range to include more possibilities for the next step (optimization algorithm). In addition, a large range for G facilitates the inclusion of a larger DFC, such as the DFC for a tap. Equation (4-6) shows that C_1 must be less than or equal to C_2, otherwise C_2 is meaningless, since if C_1 is higher than C_2, then the confidence level for every DFC must be higher than C_2, which means C_2 cannot filter out any DFCs. With the objective function and constraints listed from Equations (4-1)–(4-6), an optimization algorithm is developed as described in the following section to find the combination of $C_1, C_2,$ and G that leads to an optimal solution of $\min f(C_1, C_2, G)$.

4.4.2.2 Optimization algorithm

Optimization, which is central to decision making in various areas (such as engineering and economics), aims to find the best solution from among many possible alternatives (Chong & Zak, 2013). To find the best solution of C_1, C_2, G from all the possible alternatives, an optimization

114

algorithm called simulated annealing (SA) is employed in this study. The SA algorithm, first proposed by Kirkpatrick et al. (1983) in the category of random searching algorithms (Zabinsky, 2011), is developed to solve complicated combinatorial optimization problems (Jung et al., 2016), and is known for preventing the algorithm from getting stuck in a local minimum point in the searching process (Chong & Zak, 2013). As we can see from Figure 4.8, the global minimal point is A, while B is the local minimizer. If the random starting point is C, the naive random searching algorithm (Chong & Zak, 2013) will move towards point B and stick with it as the minimal point, since it is hard for naive random searching algorithm to climb over the local maximum point D. While the SA algorithm is designed to climb out of the local minimum point (point B in Figure 4.8) and to try to find the global minimum point (point A of Figure 4.8) as the result. The interested reader may consult Chong & Zak (2013) for a detailed description of the mechanism underlying the naive random searching algorithm and SA. The SA algorithm has been successfully employed in many optimization problems, for example, Paya et al. (2008) applied SA to solve multi-objective optimization problems of concrete frame design; Zeferino et al. (2009) used SA for wastewater system planning; and Hackl et al. (2018) employed SA in planning restoration programs for transportation networks. A detailed description of the SA algorithm is provided in the following section.
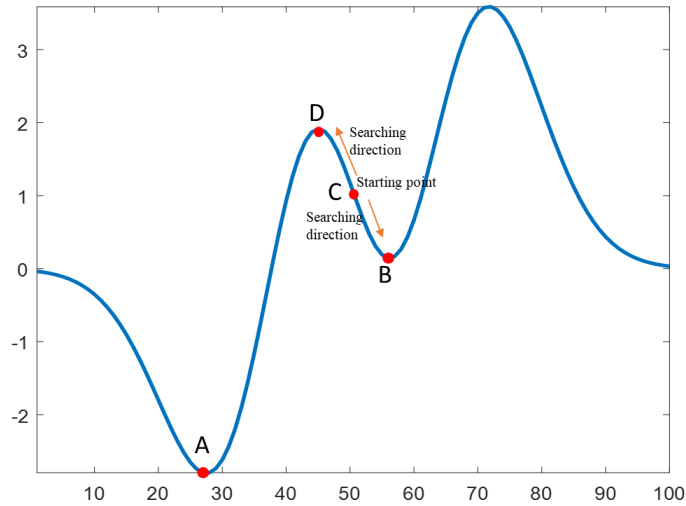
Figure 4.8: 2D example of a local minimizer and global minimizer.

The pseudo-code for the SA to find the minimal point for $f(x)$ is presented in Figure 4.9. Step 1 is to randomly sample a point from Ω, which is the domain of the function, as the starting point from where the algorithm starts its searching process, which is to sample a point from its neighborhood $N(x^{(k)})$. Step 3 defines the rule for whether to accept a candidate point or not. If a candidate point is a better point (the value of the objective function is smaller), then accept the point as the next point. If the candidate point is a worse point (the value of the objective function is larger), then the algorithm will accept the worse point with a probability of $p\left(k, f\left(z^{(k)}\right), f\left(x^{(k)}\right)\right)$, which is calculated by $\exp\left(-\frac{\Delta E}{T_k}\right)$, where ΔE is the energy difference between the candidate point and the original point, T_k is the temperature of the kth iteration. As the number of iterations increases, T_k decreases, which ensures that the probability of accepting a worse candidate point will decrease as the algorithm runs to a higher number of iterations, which is a process that mimics the physical annealing process for hot metal at high temperature changing to low temperature (Hackl et al., 2018; Metropolis et al., 1953). This rule can help the algorithm avoid getting stuck in one local minimum point, and always provides the algorithm

116

with the possibility to move to the global minimum point. After each iteration, the best-so-far point is recorded in Step 4. Stop criterion is checked in the next steps to determine whether to stop the algorithm or return to Step 2 to start another iteration.

Minimize $f(x)$
Subject to $x \in \Omega$
1. Set $k = 0$; select an initial point $x^{(0)} \in \Omega$
2. Pick a candidate point $z^{(k)}$ at random from $N(x^{(k)})$.
3. If $\Delta E = f(z^{(k)}) - f(x^{(k)}) < 0$:
    Accept the candidate point as the next point, $x^{(k+1)} = z^{(k)}$
  Else:
    Generate a random number $R \in (0,1)$
    If $R \le p\left(k, f(z^{(k)}), f(x^{(k)})\right) = \exp\left(-\frac{\Delta E}{T_k}\right)$:
        Accept the candidate point as the next point, $x^{(k+1)} = z^{(k)}$
    Else:
        Reject the candidate point as the next point, $x^{(k+1)} = x^{(k)}$.
4. Record the best-so-far point:
  $x_{best}^{(k)} = \begin{cases} x^{(k)}, if\ f(x^{(k)}) < f(x^{(k-1)}) \\ x^{(k)}, otherwise \end{cases}$
5. If stopping criterion is satisfied, then stop.
Step 6. Set $k = k + 1$, go to Step 2.

Figure 4.9: Pseudo-code for the SA.

Parameter selection is determined according to several factors, such as previous studies, the property of the problem, and experiment performance. Since the SA algorithm is a heuristic algorithm, its design does not have only one true solution. The design of such an algorithm is problem-specific. In our case study, parameters (or the structure) of the SA algorithm, such as neighborhood selection, cooling schedule, initial point, and stopping criteria are determined based on the factors mentioned above. Validation of the results of the optimization algorithm is described in Section 4.4.2.3. Neighborhood selection is mainly a problem-specific choice (Hackl et al., 2018), which requires, to some extent, subjective judgment that considers the particular situation related to the question itself. In the present problem, which considers the feasible range of three predefined variables (as mentioned in Section 4.4.2.1), the neighborhood is sampled

from integer numbers from $[-5, 5]$, since all three variables are integer numbers in our case. There are many previous studies focused on the cooling schedule, such as Hajek (1988) and Siddique & Adeli (2016). The method used in the present study was proposed by Johnson et al. (1991) to define the initial temperature T_0, and the minimum temperature at the last iteration (n) T_n. The method has been adapted in many studies, such as Zeferino et al. (2009). With the initial acceptance rate p_0, and the last acceptance rate p_n, T_0 and T_n.can be calculated as $T_0 = -\frac{1}{\ln(p_0)}$ and $T_n = -\frac{1}{\ln(p_n)}$. We choose a typical initial acceptance rate of 0.8 and final acceptance rate of 0.001, which means the probability for accepting a worse candidate point is higher at the early stage and much lower at the final stage. T_0 and T_n can be calculated accordingly. The decay factor that controls how the temperature changes at each step can be calculated as $\left(\frac{T_n}{T_n}\right)^{\frac{1}{n-1}}$. The next iteration temperature can be calculated as the current temperature multiplied by this decay factor, which simulates the annealing process in the field of metallurgy. For the initial point, we use our best educated guess $[C1, C2, G]_0 = [20, 30, 30]$, which is informed based on the experience of watching CCTV videos and reviewing their raw log files generated by the defect detector. A fixed number of iterations (100) is used as the stopping criteria, and we can always increase the number of iterations based on the output of the algorithm, and the best-so-far point will be recorded during the whole process.

The implementation of the SA with the objective function and constraints described in Section 4.4.2.1 is presented in Figure 4.10. Five videos with 98 defects are used as the test dataset in this study. The objective function keeps decreasing with several small fluctuations during the process and finally reaches a stable range around 0. The values of C_1, C_2, and G during the searching process are plotted in Figure 4.10 as well. The overall trend is contrary to the objective function,

as the three variables keep increasing and finally stay stable at a certain range. The minimum value of the objective function is 4, which means the number of defects identified by the VIASP is 4 more than that of the manual assessment report. The best point is $[C_1, C_2, G]_{best} = [42,80,85]$, which means that the optimum solution generated by the SA is that C_1 is equal to 42, C_2 is equal to 80, and G is equal to 85. These three values generated by the SA for the predefined parameters are then used in the VIASP to generate an assessment report automatically, and the performance of the VIASP using these predefined variables is tested in the next section.
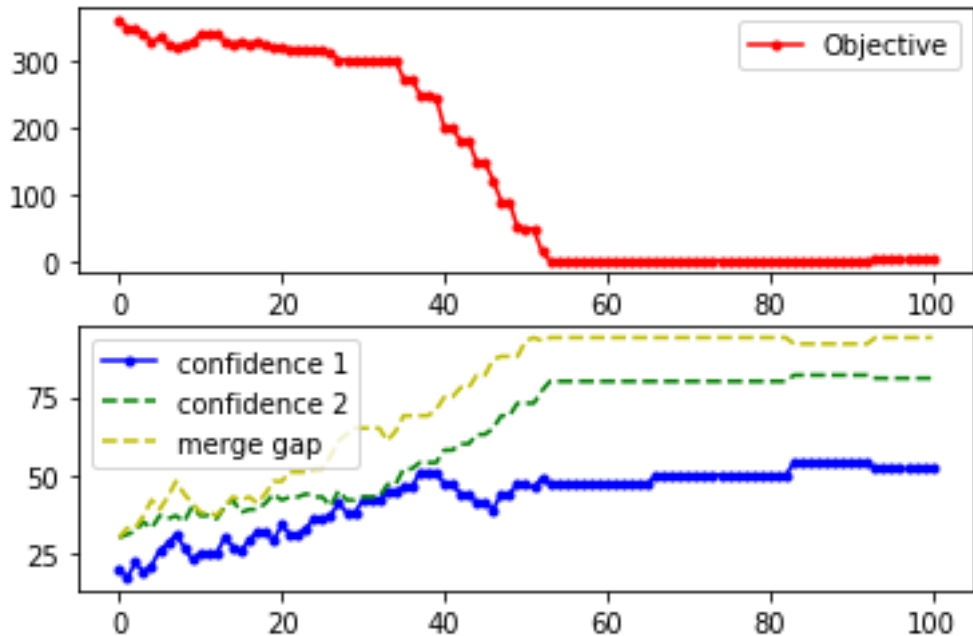
Figure 4.10: The results of the SA.

4.4.2.3 Validation and analysis for VIASP

The objective of the optimization algorithm (i.e., SA) is to find the optimal solution for three predefined variables required by the VIASP, which are $[C_1, C_2, G]_{best} = [42,80,85]$. With these three variables and the log file generated by the defect detector, the video assessment report can be generated automatically by VIASP. Figure 4.11 shows the results of the comparison of the output of VIASP and manual assessment results. The matched defects are shaded with a green

119

background in Figure 4.11 and linked with the two-way arrows. The VIASP provides more detailed results than the manual assessment report, for example, the confidence level is not included in the manual assessment report. There is a small difference in terms of the distance (shown in meters) between the two results. The VIASP generates two distances for one defect: start distance and end distance. However, there is only one distance identified by technologists and recorded in the manual assessment report. The start distance is always ahead of the distance indicated by manual assessment result, which occurs because once the defect appears on the screen, the defect will be labeled. The end distance is relatively close to the distance indicated in the manual assessment report, although there are small differences. Three false alarms are reported by the VIASP, and two cracks are missed by the VIASP in this example.

❏ VIASP Result

No.	Type	Confidence	Start Frame	End Frame	Start time	End time	Start diatance	End distane
1	Deposits	81.47	643	659	00:25.7	00:26.4	6.4	6.4
2	Root	89.71	703	719	00:28.1	00:28.8	7	7.1
3	Tap	88.66	2208	2284	01:28.3	01:31.4	19.5	20.1
4	Tap	93.92	2834	2957	01:53.4	01:58.3	20.2	20.7
5	Deposits	83.41	3907	3986	02:36.3	02:39.4	24.8	25.4
6	Deposits	95.99	4074	4156	02:43.0	02:46.2	26.2	26.9
7	Tap	95.59	4600	4691	03:04.0	03:07.6	30.6	31
8	Deposits	85.22	5037	5045	03:21.5	03:21.8	31.2	31.2
9	Tap	93.74	5243	5323	03:29.7	03:32.9	31.7	32.1
10	Tap	95.17	6356	6425	04:14.2	04:17.0	36	36.6
11	Tap	97.6	6932	6956	04:37.3	04:38.2	37	37.2
12	Fracture	95.44	6957	6974	04:38.3	04:39.0	37.3	37.3
13	Deposits	86.58	7512	7615	05:00.5	05:04.6	37.6	38.4
14	Tap	91.12	7618	7665	05:04.7	05:06.6	38.4	38.8
15	Tap	91.31	8354	8404	05:34.2	05:36.2	39.3	39.6
16	Hole	83	8407	8415	05:36.3	05:36.6	39.7	39.7
17	Deposits	86.88	8443	8482	05:37.7	05:39.3	39.7	39.7
18	Broken	81.66	8598	8678	05:43.9	05:47.1	39.8	39.8

❏ Manual assessment Result

Defect type	Distance
Root	7.2
Tap	20.1
Crack	20.6
Tap	20.7
Deposits	25.5
Deposits	27
Tap	31.1
Tap	32.2
Tap	36.6
Tap	37.4
Fracture	37.5
Deposits	38
Tap	38.7
Crack	39
Hole	39.6
Tap	39.7

TP	Matched	
FP	False alarm	
FN	Miss	

Figure 4.11: Example of comparison between VIASP and manual assessment report.

The performance of the VIASP is tested with five CCTV videos including 98 defects identified in the manual assessment report in total. The performance of VIASP is tested by comparing the output of VIASP with the manual assessment report. The comparison is conducted manually since the matching of the defects from VIASP with the manual assessment report automatically

is challenging at this stage (as described in Section 4.4.2.1). F-measure, which is a widely used metric to test accuracy (Martínez-Rojas et al., 2018; Sasaki, 2007), is used in this study to measure the accuracy of VIASP. The definitions of F-measure are presented in Equations (4-7)– (4-9).

$$Precision = \frac{TP}{TP+FP} \qquad (4\text{-}7)$$

$$Recall = \frac{TP}{TP+FN} \qquad (4\text{-}8)$$

$$F_1 = 2 \times \frac{Precision \times Recall}{Precision+Recall} \qquad (4\text{-}9)$$

where TP is the true positive result, FP is the false positive, FN is the false negative. Based on the equations, precision denotes the correct rate over all the detected objects, while recall is the rate of completeness of detected objects among all the manual results (Yin et al., 2020c). A detailed description of F-measure can be found in a study authored by Sasaki (2007). Overall, for the selected five videos with 98 defects, the precision for VIASP is approximately 0.77 and the recall is approximately 0.73, which gives an F_1 score of 0.75. There are areas for potential improvement in future research in terms of accuracy for the proposed algorithm. Several examples of how real-life conditions could influence the performance of the VIASP are summarized below:

1. The inconsistency of the manual assessment results influences the value of the metric that measures the performance of the algorithm (i.e., F_1 score in this study). The manual denotation of some small defects, such as deposit, shows some inconsistency compared with the VIASP. For example, as is shown in Figure 4.12a and Figure 4.12b, for two similar deposits, one technologist recorded one as a deposit; however, a similar deposit was not recorded as a deposit by another technologist (or perhaps even by the same

technologist). In addition, in some cases, small defects may be ignored or overlooked by the technologists (as shown in Figure 4.12c, which is the No.1 defect in Figure 4.11), but the same defects are detected by the defect detector because the VIASP has a consistent standard for identifying the defects. This issue has a negative impact on the F_1 score in our case, if the impact of the manual assessment method's inherent inconsistencies could be accounted for, the measured performance of the algorithm would improve.

2. Some of the defects (e.g., deposits) within a tap are identified and outputted in the final report by the VIASP; however, these defects do not need to be recorded in the manual assessment report. For instance, Figure 4.12d shows an example of deposits within a tap that were identified by the VIASP, but were not recorded in the manual assessment report. In practice, these defects are not counted as defects for the inspected sewer pipe because they are in tap. We developed a rule to avoid this situation (see Step 4-c of Figure 4.6), which works for the majority of cases; however, a certain number of false alarms are still made by the VIASP. In fact, the false alarms for deposits (No. 8 and No. 17) in Figure 4.11 are deposits within a tap.

3. Some small defects are detected and labeled in the CCTV video; however, some of them are excluded from the final output of VIASP because these detected frames show a low confidence level, as shown for example in Figure 4.12e. The crack is labeled by the defect detector successfully, but the confidence level of the specific frame or confidence level of DFC that represents this crack is lower than the confidence threshold that we defined. During the trade-off process of choosing the confidence level and the merge gap, a relatively higher confidence level is selected that results in the detection recall of some

of the small defects being sacrificed to eliminate a large number of duplications in the final report.

4. Obstructions within the sewer pipe that hinder the moving of the crawler will cause false alarms since the CCTV camera needs to rotate to different angles to check how the obstruction is blocking the path, which will generate many abnormal views for the defect detection algorithm. An example of the crawler being blocked by obstructions is presented in Figure 4.12f. The operator is moving the camera to face the crawler's wheel to check the condition of the obstruction. The defect detector wrongly detects the wheel as a hole in this case. It is difficult to detect anything preciously based on these views since these views are not included in the training dataset, which means that they have never before been seen by the defect detection algorithm. A potential direction for improving the performance of the VIASP would be to include more abnormal views of CCTV videos in the training of the defect detector in order for the defect detector to be able to discern these images, which will in turn improve the performance of the defect detector, and further to improve the performance of VIASP.

5. The quality of the video may change significantly due to human factors (improper operation such as moving too fast, not using lighting properly, etc.) or non-human-factors (water sag or smoking within the pipe) during the recording process. Figure 4.12g shows the camera being submerged in water sag, which will cause discontinuous frames if these frames are within the middle of a DFC or cause false alarms. Figure 4.12h shows the blurred video frame caused by unknown equipment problems.

In summary, taking into consideration how these factors influence the performance of the VIASP, the F_1 score of 0.75 is considered acceptable for the proposed VIASP. Future efforts

could be made to improve the performance of the algorithm including enhancing the performance of the defect detector as mentioned above.

| a. Deposits identified manually | b. Deposits has not identified manually | c. Deposits identified by VIASP not manually |

| d. Deposits within a tap | e. Exclude by VIASP but recorded manually | f. Obstructions within the sewer pipe |

| g. Camera under water | h. Low quality video |
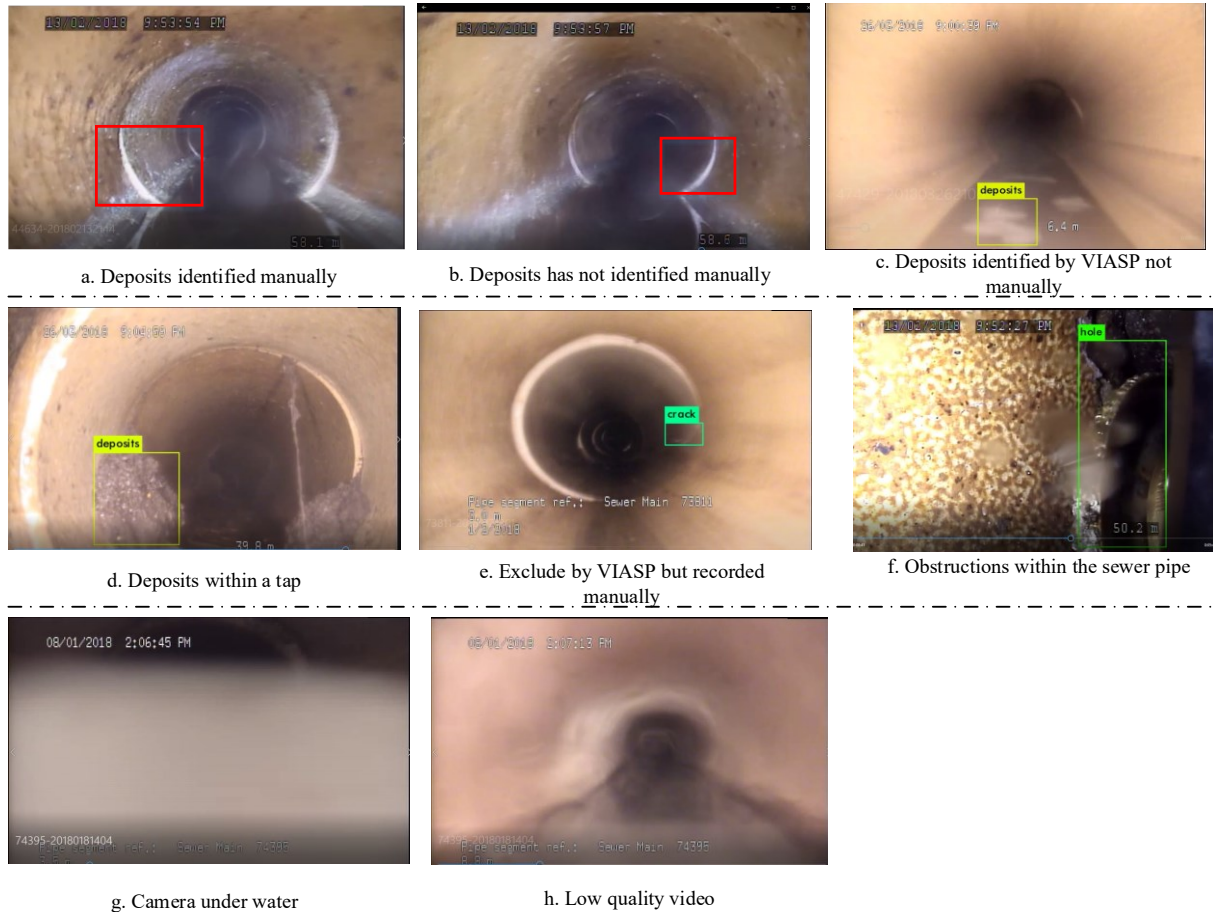
Figure 4.12: Examples that influence the performance of VIASP.

4.5 SPVA system prototype development

With the developed defect detector and VIASP, the present research develops a sewer pipe video assessment (SPVA) system prototype to pack all the developed functionality into a user-friendly software package to facilitate the daily operations of the sewer pipe maintenance department. Figure 4.13 presents the interaction and workflow between the on-site and off-site parts of the overall process undertaken by a sewer pipe maintenance department. As is shown in the existing information flow in Figure 4.13, the maintenance office can generate work orders for CCTV

collection based on the maintenance schedule, which is developed based on the historical CCTV video assessment reports. CCTV collection crews will collect the CCTV videos, while in the meantime, log files recording daily works are generated along with the CCTV collection processes. With the log files, the assessment office can know which videos are most recently collected and still need to be assessed. Then, each of the newly collected videos is watched from beginning to end by a technologist. The assessment reports are generated while watching these videos. With the proposed SPVA system, the assessment process could be completed automatically (see the updated information flow in Figure 4.13). The log file that records the daily work completed and the CCTV videos there were collected can be fed into the SPVA system directly. The SPVA system will find the targeted CCTV video and complete the assessment process automatically (defect detection, video interpretation, text recognition, writing to Excel, etc.). The generated report then can be sent to the assessment technologist for review and further analysis.
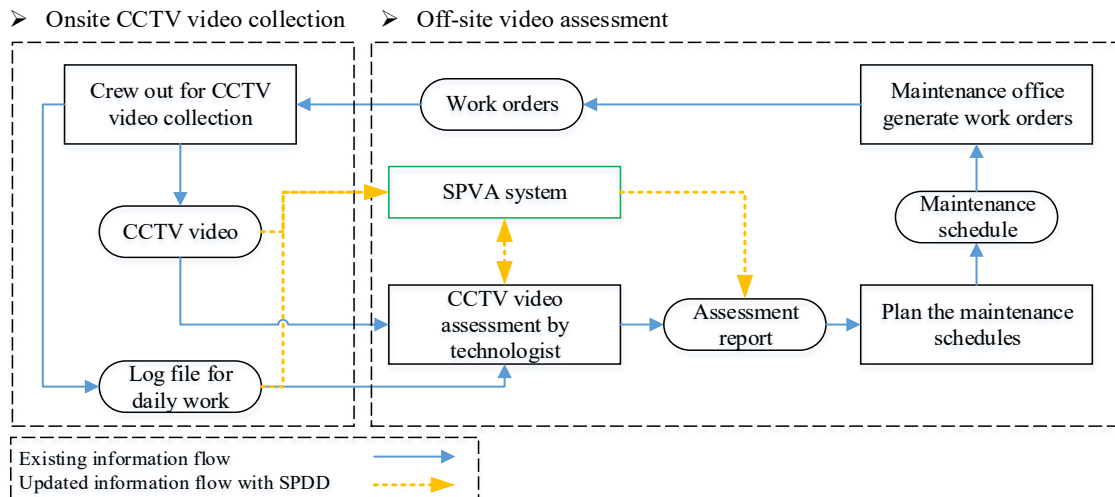
Figure 4.13: Interaction and workflow between on-site and off-site working processing.

A prototype of the SPVA system, whose user interface (UI) is shown in Figure 4.14, is proposed in this research. Part A of Figure 4.14 is the menu bar, which includes *File*, *View*, *Search*, *Train*

YOLO, and *Setting* (see Part A on the right of Figure 4.14 for the submenus within each menu category). The user can open a specific CCTV video that needs to be processed from the *File* menu. After the processing of the video is complete, the user can choose to save the results in Excel format or Database format (e.g., Microsoft Access database), or both. The *View* function generates a summary of the process results. For example, information such as the total number of videos processed, or the total number of each type of defect can be viewed here. The *Search* function gives the user the option to search a specific type of defect within a specific location of the processed sewer pipe. The *Train YOLO* function facilitates the retraining process for the defect detector (i.e., YOLO in this case) by helping the user to select the image files and annotation files (e.g., .TXT file and .XML file) to replace the old versions of image files and annotation files. *Setting* function provides a portal to link the SPVA with Microsoft Azure for executing the text recognition function. Users could also set the saving path for the labeled video, and the generated Excel (or database) file.

Part B of Figure 4.14 is the run YOLO module, where parameters such as confidence level (C_1, C_2), merge gap (G), weight (different versions of previous trained YOLO), could be set and adjusted by the user. The video number of the video currently being processed is showing after the label of *video*. Several parameters can be adjusted and the *Confirm* button loads the parameters that were changed by the users: for example, the *weight* button is for selecting a previously trained YOLO and loading its weight to the software, the *Run* button is to start the video processing, and the *Analyze* button can analyze the log file generated by YOLO using VIASP, instead of going through the process of running YOLO and then running the VIASP.

The *multi-task* module is shown in Part C of Figure 4.14, where users can process a folder of videos instead of running a single video in module B. The *Video selector* helps to filter the

videos by date since the video database is would typically contain a large number of files, thus, filtering could help users find the targeted videos more efficiently. The *schedular* module (Part D of Figure 4.14) is a feature that can schedule the assessment of the videos automatically based on the uploaded CCTV inspection log files. The CCTV inspection crews will upload the log file once they are finished their daily work, and the log file is where such information as collecting time, media storage location, sewer pipe number, uploading time, etc., is recorded. The *schedular* module can identify those videos that have been uploaded but have not yet been assessed. Then, according to a set schedule (e.g., 6 am, from March 1 to March 30, daily), the *schedular* automatically starts the software and processes the newly uploaded videos according to the inspection log file. Assuming all the video processing jobs could be done within two hours, then, the assessment technologists could see the assessment report at 8 am when they arrive at the office. Therefore, this *schedular* feature could save time in terms of the amount of work the technolgists must do when they arrive at the assessment office.

Note that this is a proposed prototype based on our case study, which aims specifically at facilitating the work processes in our case study. With the developed core functions (defect detection, video interpretation, etc.), the features in terms of application could always be adjusted for other cities' municipal departments. This prototype is simply a showcase to demonstrate how to integrate the developed machine learning-based automation tools into real-life sewer pipe maintenance work. Admittedly, there is room for improvement in terms of software development; however, the backbones (e.g., defect detector, VIASP) of the SPVA system are generalized tools that could improve the automation of sewer pipe assessment work.
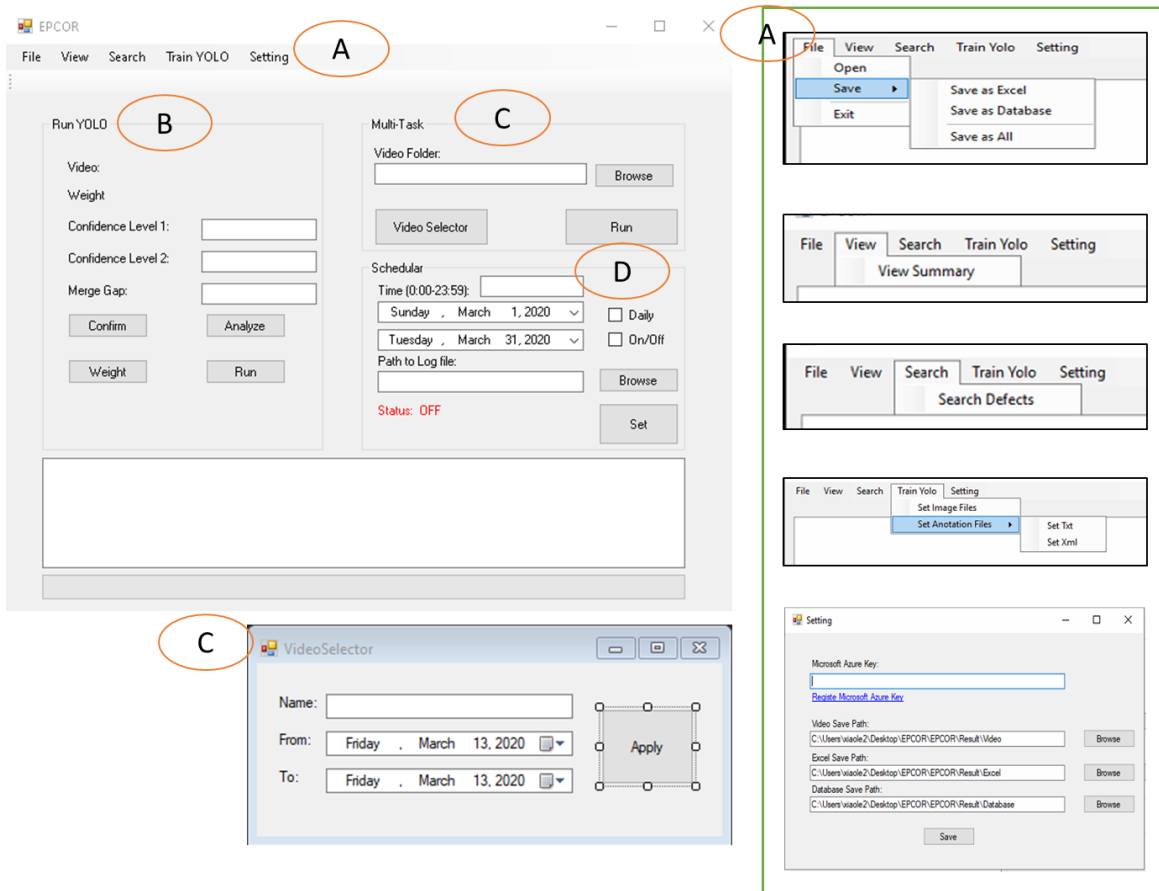
Figure 4.14: User interface of the SPVA system.

4.6 Discussion

A highly accurate defect detector is the foundation for the VIASP as well as the SPVA system, since the more accurate the defect detector is, the more accurate the VIASP will be. The accurate output of the VIASP enables the SPVA system to improve the level of automation of the overall sewer pipe assessment process. This study employed an excellent defect detector based on the YOLO algorithm; however, improvements could be achieved with more efficient and accurate algorithms in future research regarding defect detection. The performance of VIASP could be improved as a result of using a better defect detector; therefore, developing a more advanced defect detector is an important research direction that requires continuous efforts in the future.

Multiple prior experiments were conducted before finalizing the structure of the VIASP since the development process starts with a thorough understanding of the targeted problem and then considers the estimation of the performance of the algorithm for different versions of the VIASP during the algorithm design process. The algorithm development process is a trial and error process; lessons can be learned from all the previous efforts. In the future, the performance of VIASP could be improved in various ways, for example, by adding more filters to filter out the false alarms that cannot be filtered out at this stage (e.g., defects within the tap), or by finding the appropriate confidence level for each type of defect instead of using one confidence level for all the defects, as was done in this study.

This research offers four contributions to the body of knowledge. 1) This research proposes an automated sewer pipe video assessment framework that includes defect detector, video interpretation algorithm, text recognition function. This present study provides a framework that future similar system development can follow. 2) This study proposes a novel video interpretation algorithm for sewer pipes (VIASP), which is tested and validated in this study. An optimization algorithm is employed to find the optimal human-defined parameters for the VIASP. 3) The authors of the present study develop a prototype for the sewer pipe video assessment (SPVA) system, which demonstrates how the developed automated functions (defect detector, VIASP, etc.) can be deployed in a real-life sewer pipe maintenance scenario considering the workflow of the daily operations. Note that the VIASP can improve productivity, but at this point it is not a replacement for the technologists since there are mistakes made by the defect detector. In this case, risk management should be considered in the implementation of this kind of system. 4) Overall, the research improves the level of automation in sewer pipe assessment

work. Consistency is a by-product of the automation of the sewer pipe assessment system since the defect detector and the VIASP can output assessment results in-line with a constant standard.

4.7 Conclusions

Sewer pipe assessment is the foundation on which an efficient maintenance plan is developed since the health condition of the sewer pipe is assessed and recorded as the basis for decision making in this process. In order to improve the level of automation of this CCTV-based sewer pipe assessment process, various previous studies have developed different versions of automated defect detectors for labeling the defects that appear in the video with specific names according to a particular nomenclature. However, the labeled videos or video frames require further processing to extract useful information from the CCTV videos and to automatically generate the assessment reports. The present research aims at improving the automation for sewer pipe assessment in terms of CCTV video interpretation and SPVA system development. The VIASP is designed to identify the frame clusters that could represent the defect in the sewer pipe. The VIASP features confidence level filtering, and the merging of frame gaps with no defect or discontinuous frames. Three key predefined parameters ($[C_1, C_2, G]_{best} = [42,80,85]$) for VIASP are determined by an optimization algorithm (i.e., SA). The performance of the VIASP is tested by comparing the output of the VIASP with the manual assessment results. The testing shows that the VIASP could realize an F_1 score of 0.75. Potential factors that could affect the performance of the VIASP are analyzed in detail, which provides insights for potential ways to improve the performance of VIASP in future works. In order to show how the previously developed automated functions (e.g., defect detection, video interpretation, text recognition) aimed at facilitating the sewer pipe assessment could fit into the everyday workflow of sewer pipe maintenance work, a prototype of an SPVA system is proposed to demonstrate the overall

process. Backbone components (i.e., defect detector, VIASP, text recognition algorithm) of the SPVA are replaceable whenever more advanced techniques are available. Overall, this research could serve as a framework for future similar system development.

Chapter 5: DATA-DRIVEN BI-LEVEL SEWER PIPE DETERIORATION MODEL DESIGN AND ANALYSIS[4]

Preface

To inspect every pipe in a city annually would be a costly and labor-intensive endeavor due to the significant number of sewer pipes underground. In practice, sewer pipes in bad (good) health conditions should be assigned a high (low) priority for maintenance. A deterioration model is an effective tool for prioritizing sewer pipe maintenance since it is able to predict the future conditions of these pipes. The deterioration model is developed based on a large amount of historical data recording the health condition of sewer pipes (which is usually recorded in the pipe assessment reports, which is the major focus of Chapters 3 and 4). This chapter proposes a bi-level deterioration model to predict the condition of sewer pipes at a neighborhood level and an individual level. The neighborhood-level prediction is used to facilitate the maintenance schedule at the city scale, while the individual-level prediction is used to identify the sewer pipes with the highest risk of failure so that maintenance operations can be scheduled to keep them operating at an acceptable level of service. The deterioration model is important in the preventive maintenance of sewer pipe systems since it could provide evidence for decision-making in the next stage inspection schedule (i.e., CCTV collection in this case), which is the main topic of Chapter 2 (see Figure 1.2). The Chapter 5 not only developed a bi-level deterioration model for the targeted city but also proposed a framework that can be generalized for municipal

---

[4] A version of this chapter has been published in the journal, *Automation in Construction*, as follows: Yin, X., Chen, Y., Bouferguene, A., and Al-Hussein, M. (2020). "Data-driven bi-level sewer pipe deterioration model: Design and analysis." *Automation in Construction*, 116, 103181. https://doi.org/10.1016/j.autcon.2020.103181. It also has been reprinted with permission from Elsevier.

departments located in other cities that aim to develop their own deterioration model for sewer pipe systems.

5.1 Introduction

Sewer pipes are a critical component of urban infrastructure, not only because of their important functionality, which aims at transporting wastewater and stormwater to treatment plants or water streams in the vicinity of urban centers but also because of the cost required for their construction and repair (Ana & Bauwens, 2010; Hawari et al., 2018; Wirahadikusumah et al.,2001). In order to be operational at all times, the pipes of the drainage system need to be structurally sound and free of debris or deposits since failures (structural and/or operational) can lead to flooded streets and buildings (Tran et al., 2008). Deterioration, a problem caused by aging or other factors, may prevent the sewer pipe system from maintaining an adequate level of service and increases the probability of pipe failure (e.g., blockage or collapse), which would lead to major repair or replacement work. Therefore, pipe failure caused by deterioration will not only lead to impacts on the environment but also lead to large repair costs (Hassan et al., 2019a). According to a 2016 report, in Canada, approximately 35% of wastewater and 23% of stormwater infrastructures are in very poor to a fair physical condition requiring $78 billion and $31 billion for their replacement, respectively (CIRC, 2016). Therefore, a proper maintenance strategy should be adopted to keep sewer pipes operating at an adequate level of service, to extend their service life, and to decrease the annual investment required for replacement (Elmasry et al.,2018).

Inspection and assessment are the main activities for the preventive maintenance by assessing the condition of the sewer pipe and further assigning different priorities for pipes in different conditions (Elmasry et al., 2019; Hawari et al., 2017). However, planning an effective preventive

maintenance schedule requires an accurate deterioration model to predict the current and future condition of the pipes and prioritize those pipes in a worse condition so they can be maintained in a timely manner (Baik et al., 2006). Essentially, a deterioration model for sewer pipes is a prediction model that uses certain influencing factors to predict the condition of the pipe. Thus, several mathematical prediction models have been adopted in the study of this problem, such as the regression model (Chughtai & Zayed, 2008), Markov chain model (Micevski et al., 2002).

In conducting a literature review on the topic of deterioration models for sewer pipes, the author found many studies pertaining to predicting the future condition of sewer pipes at the individual sewer pipe level (e.g., Baik et al., 2006; Ana et al., 2009, etc.), but few at the neighborhood-level specifically, even though a neighborhood-level prediction model could be an important tool for facilitating city-scale maintenance planning. Therefore, a neighborhood-level deterioration model is proposed in this research. Furthermore, in previous sewer pipe deterioration models (which are summarized in the following section), the researchers have, for the purpose of improving prediction accuracy, included as many input variables as possible when developing their models. However, including more variables means the data collection process requires more effort. In addition, some of the variables included in previous deterioration models may be inaccessible by other cities that intend to develop their own deterioration models using an existing method. Therefore, reducing the number of input variables could be economically and technically beneficial for municipal departments when it comes to developing such a deterioration model. Thus, to develop the individual-level deterioration model this study proposes integrating a backward variable elimination process with a neural network (NN), which is an efficient tool for solving classification problems.

This research proposes a novel bi-level deterioration model, taking advantage of both the rapidly increasing power of modern computers and a large amount of available historical data to predict the condition of sewer pipes at a neighborhood-level and the level of the individual sewer pipe. The neighborhood-level prediction could help with the scheduling of the maintenance (e.g., CCTV inspection) by prioritizing the neighborhoods. For example, those neighborhoods predicted as having sewer pipes in an overall worse condition will be assigned a high priority for maintenance, while other neighborhoods would have a lower priority. The other motivation for the neighborhood-level prediction is that the results of the prediction can serve as input for the individual-level prediction model since the neighborhood prediction results are representative of the sewer pipe condition in the immediately surrounding area. The individual prediction model will be constructed with a pattern recognition model developed by the NN.

Several coding systems could be used to rate the condition of the sewer pipe, such as the Pipeline Assessment and Certification Program (PACP) (NASSCO, 2015), and Manual of Sewer Condition Classification (MSCC) (Water Research Centre (WRc), 2013). Generally, the coding system is used to assess the CCTV inspection videos of sewer pipes. The defects existing in the pipe will be coded according to one of the coding manuals mentioned above, and finally, a score can be calculated to represent the overall condition of the sewer pipe (Hawariet al., 2018; Yin et al., 2019). In this research, one index called the likelihood of failure (LOF) from PACP would be adopted to determine the condition of the sewer pipe. According to PACP, LOF is defined as a numerical value that provides an overview of the pipe condition based on the probability of failure (NASSCO, 2015). The LOF index was also used in previous research studies to measure the condition of sewer pipes (e.g., Arthur et al., 2009; Hahn et al., 2002). The LOF has the best value of 1 (meaning that there are no defects in the pipe segment) and the worst value of 5

135

(meaning that the pipe has a significant number of severe defects). The detailed description of the LOF index is shown in Table 5.1 (Ehret, 2011). The sewer pipe with low LOF usually will be assigned with a low priority in the maintenance schedule, while high LOF corresponds to a high priority in the maintenance schedule. Thus, the LOF serves as a risk-based reference to measure the health condition of the sewer pipe. In this research, the LOF will be the target index which will be predicted by several selected predictors.

Table 5.1: Description of LOF.

LOF grade	Description
1	Failure unlikely in foreseeable future
2	Pipe unlikely to fail for at least 10–20 years
3	Pipe may fail in 10–20 years
4	Pipe will probably fail in 5–10 years
5	Pipe has failed or will likely fail within 5 years

5.2 Related works

5.2.1 Deterioration model for underground pipeline

Several deterioration models have been developed over the years. In order to provide accurate prediction results, researchers try to determine the most appropriate mathematical model to mimic the deterioration process of the sewer pipe. Table 5.2 summarizes sewer pipe deterioration models developed in previous research. It is shown in the table that the regression model and the Markov Chain model are two models most commonly used in the previous research. The regression analysis aims at determining the relationship between the sewer pipe condition and the selected predictors. For example, Chughtai & Zayed (2008) developed a deterioration model using multiple regression techniques. The research predicted the structural grade and operational grade of the pipe for various pipe materials. Ana et al. (2009) used a logistic regression model to predict a binary outcome variable. They classified the five classes of sewer pipe inspection results, where class 1 is very good, and class 5 is very poor, into two categories: the good state

136

(classes 1–3) and the failed state (classes 4–5). By using this approach, they can predict whether a pipe is in a condition that needs rehabilitation or replacement. Younis & Knight (2010) developed a probabilistic model using the ordinal regression model. In their research, the objective was to determine the probability that a pipe falls into one of five condition grade categories considering several selected covariates (e.g., age and material). The other major branch of the deterioration model is the Markov chain model. Since the most commonly adopted sewer pipe assessment systems, e.g., PACP (NASSCO, 2015) and MSCC (WRc, 2013), usually classify the pipe into five classes from very poor to very good, the Markov chain process can mimic the states changing process from one class to another. For example, a newly installed pipe is assigned to the "very good" class; however, over a certain time period, it is possible the class changes to a worse level. The Markov chain model is used to model this process with a fixed time period (e.g., one year or two years) by constructing a transition matrix, which is the primary objective in such deterioration model development. Different methods can be used to calculate the transition matrix, such as Bayesian inference (Micevski et al., 2002), ordinal regression (Baik et al., 2006; Lubini & Fuamba, 2011), and the Markov Chain Monte Carlo (MCMC) (Lin et al., 2019; Micevski et al., 2002; Tran et al., 2008). Survival models that were commonly used in population prediction or medical research were also used in previous research to develop the sewer pipe deterioration model (e.g., El-Housni et al., 2017; Duchesne et al., 2013; Baur & Herz, 2002). In addition, decision-making tools such as analytic hierarchy process (AHP) (Ennaouri & Fuamba, 2013), decision tree (Syachrani et al., 2013), and random forests (Harvey & McBean 2014) were used to rate or classify sewer pipes considering a certain number of predictors.

Table 5.2: Summary of the mathematical models used in previous deterioration models.

Model type	Researcher
Markov chain model	Micevski et al. (2002); Baik et al. (2006); Tran et al. (2008); Dirksen & Clemens (2008); Lubini & Fuamba (2011); Lin et al. (2019)
Cohort survival model	Baur & Herz (2002)
NN	Najafi & Kulandaivel (2005); Tran et al. (2006)
Analytic hierarchy process (AHP)	Ennaouri & Fuamba (2013)
Ordinal regression	Tran et al. (2008); Younis & Knight (2010)
Multiple regression	Chughtai & Zayed (2008)
Logistic regression	Ana et al. (2009); Lubini & Fuamba (2011); Ana et al. (2009); Lubini & Fuamba (2011)
Dynamic deterioration model	Syachrani et al. (2011)
Decision tree	Syachrani et al. (2013)
Survival analysis model	Duchesne et al. (2013)
Random forests	Harvey & McBean (2014)
Cox model	El-Housni et al. (2017)
Bayesian geoadditive regression model	Balekelayi & Tesfamariam (2019)

The relationship between the deterioration pattern and the deterioration factors mentioned above is complex as well. It is difficult to determine whether the deterioration speed has a linear or nonlinear relationship with any single factor or a combination of factors (Balekelayi & Tesfamariam, 2019). The NN could be an effective tool for dealing with such a complex problem. Najafi & Kulandaivel (2005) developed a deterioration model with a NN using seven input variables to predict the rating score of the sewer pipe. Using a three-layer network with a structure of 15-10-1 (15 input neurons, 10 hidden neurons, and 1 output neuron), the authors concluded that it is feasible to use a NN to solve such a problem. However, they also noted that accuracy (70%) still needs to be improved by increasing the size of the dataset and collecting additional input variables. Tran et al. (2006) applied a probabilistic NN (PNN) model to classify concrete storm pipes into three categories (i.e., good, fair and poor). The accuracy of their experiment was 66.9%, which can be improved by adding more predictors; however, additional

predictors were not available in their research. Another approach to improving the accuracy is to feed the model a properly distributed dataset for these three categories since their dataset contained a large portion of pipes in poor condition, which may cause bias during prediction. Therefore, in this research, a three-layer NN model is proposed to predict the individual pipe condition considering a new combination of factors and with the goal of achieving a higher prediction accuracy.

5.2.2 Factors for deterioration model

The factors considered in previous sewer pipe deterioration problems are summarized in Table 5.3. Obviously, age is the most frequently used predictor since natural deterioration (without any external impact) is caused by age. After that, the diameter is the second most frequently considered factor, and several researchers have claimed that smaller pipes may deteriorate faster than larger pipes (e.g., Balekelayi & Tesfamariam, 2019; Lubini & Fuamba, 2011; Micevski et al., 2002). However, other researchers have argued that the diameter may not influence the deterioration rate (e.g., Baik et al., 2006; Tran et al., 2008). Similar contradicting views also exist for other factors such as depth of cover (Ana et al., 2009; Balekelayi & Tesfamariam, 2019), and length (Balekelayi & Tesfamariam, 2019; Lin et al., 2019). Research consistency does exist, however, for some factors. For example, one commonly held opinion is that different materials deteriorate at different rates (e.g., Balekelayi & Tesfamariam, 2019; Lubini & Fuamba, 2011). When investigating previous research, these contradicting conclusions are commonly found for other factors, and the contradictions may be caused by the fact that all the results are derived from statistical analysis of historical data, which may cause bias when different factors are considered or when a dataset of a different size is used. The aforementioned problem is one of the limitations of this type of statistical analysis-based deterioration model development. A more

convincing method to develop the deterioration problem is to find all the influencing factors and analyze them based on physical and chemical principles, and this may be undertaken in future research.

Table 5.3: Summaries of influencing factors used in the previous deterioration model.

#	Researcher / Factor	Micevski et al., 2002	Baur & Herz, 2002	Najafi & Kulandaivel, 2005	Baik et al., 2006	Tran et al., 2006	Tran et al., 2008	Chughtai & Zayed, 2008	Dirksen & Clemens, 2008	Ana et al., 2009	Younis & Knight, 2010	Lubini & Fuamba, 2011	Syachrani et al., 2011	Ennaouri & Fuamba, 2013	Syachrani et al., 2013	Duchesne et al., 2013	Harvey & McBean, 2014	El-Housni et al., 2017	Lin et al., 2019	Balekelayi & Tesfamariam, 2019	Frequency
1	Age	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	19
2	Diameter	✓	✓	✓	✓	✓	✓	✓		✓		✓	✓	✓	✓		✓	✓	✓	✓	16
3	Length	✓		✓	✓			✓		✓		✓	✓	✓	✓		✓	✓	✓	✓	13
4	Material	✓	✓	✓	✓			✓			✓	✓	✓	✓	✓	✓		✓		✓	13
5	Slope		✓	✓		✓	✓	✓				✓	✓		✓		✓	✓	✓	✓	12
6	Location		✓			✓		✓					✓	✓			✓		✓	✓	8
7	Type (storm/sanitary)	✓	✓	✓	✓			✓						✓				✓			7
8	Depth of cover			✓		✓	✓	✓						✓			✓			✓	7
9	Soil type	✓				✓	✓	✓						✓							5
10	Number of trees					✓	✓					✓			✓						4
11	Groundwater											✓	✓								2
12	Maint. history																		✓	✓	2
13	Shape		✓											✓							2
14	Water breaks																✓				1
15	Orientation change																✓				1
16	Slope change																✓				1
17	Elevation*																✓				1
18	Bedding condition							✓													1
19	Serviceability	✓																			1
20	Exposure	✓																			1
21	Capacity													✓							1
22	Category (trunk/branch)																✓				1

*up/downstream invert elevation

The location factor listed in Table 5.3 is a factor that encompasses several specific factors, for example, the traffic load above the pipe (Baur & Herz, 2002; Chughtai & Zayed, 2008; Tran et al., 2006), and the sewer area (commercial, residential, industrial, etc.) (Harvey & McBean, 2014;

Syachrani et al., 2011). Balekelayi & Tesfamariam (2019) considered the location (geographical) information using a Bayesian Geoadditive model, in which they include some unobserved factors (such as nature of the soil, groundwater fluctuations, activities in the vicinity, etc.) in the deterioration model. The frequency of factors considered in previous research is tabulated in Table 5.3, where a higher frequency indicates the factor is an important consideration in such a problem since it is commonly considered by experts in the field, while a lower frequency indicates that the factor is less important in comparison to those with higher frequency. The importance of these factors according to the frequency is not absolute truth; however, it can reflect the importance of that factor has on deterioration to some degree. Thus, the frequency in Table 5.3 will be used as a reference to rank the importance of the factors in this research.

The deterioration of sewer pipes is a very complicated process that may be caused by many variables for which measurements may not be available, for example, factors related to the surrounding environment such as soil condition, surrounding utilities, and activities, and groundwater condition (Balekelayi & Tesfamariam, 2019). It is assumed that pipes located within a relatively small geographical area, i.e., neighborhood, are likely to experience similar impacts from factors related to the surrounding environment (Balekelayi & Tesfamariam, 2019). As a result, the deterioration caused by such factors on similar pipes, i.e., PVC, clay, concrete, etc., will probably be of (approximately) the same magnitude. Thus, geographical information is included in this research for the purpose of predicting the average condition of the neighborhood. The geographical information can represent some unobserved information for pipes in the same neighborhood; therefore, experiments should be performed to determine whether the neighborhood-level pipe condition should be included as input in the individual-level pipe condition prediction.

5.3 Methodology

The methodology employed in the development of the proposed bi-level sewer pipe deterioration model to predict the sewer pipe condition at the neighborhood-level and individual-level is presented in Figure 5.1. The neighborhood-level prediction results serve as input for the individual prediction model since it represents some unobservable factors mentioned in the previous section. The model selection and influencing factor selection take into consideration previous research studies, which facilitate the design of the proposed method. Data collection is performed considering the results of the literature review and the availability of the data for our study. The influencing factors are retrieved from the sewer pipe inspection database. In addition, geographical data is collected to facilitate the neighborhood-level prediction and to feed the results to the individual prediction model, where LOF is the target for both prediction models. The methodology for the bi-level prediction model is described in detail in the following sections. Finally, with the developed model, a determination analysis for sewer pipes at the neighborhood-level and the individual-level is conducted.

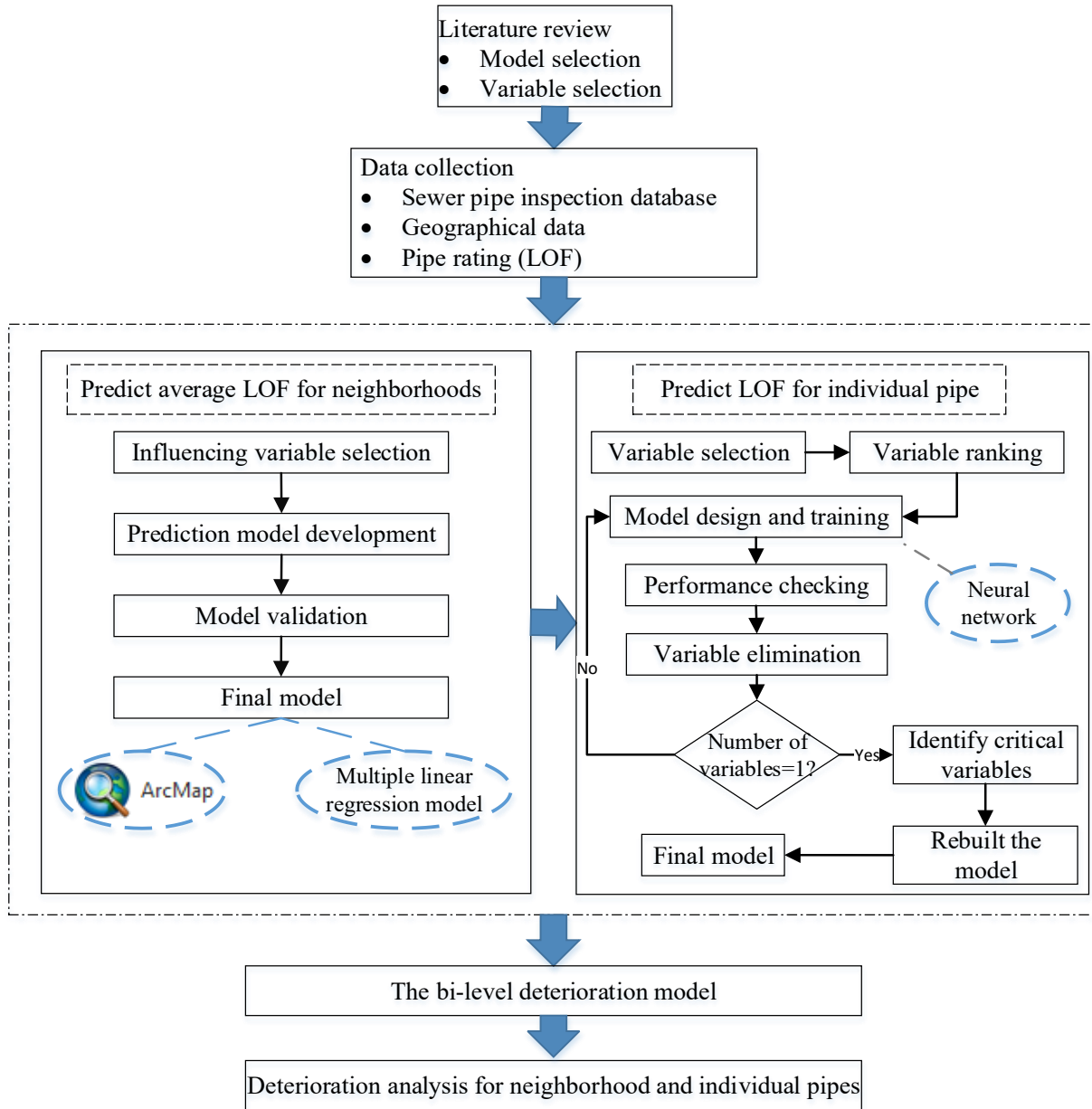Figure 5.1: Research methodology.

5.3.1 Predicting the LOF for neighborhoods

5.3.1.1 Variable selection

Since neighborhoods are considered as spatial units for predicting the neighborhood LOF of sewer pipes, the variables influencing pipe condition should also be selected in accordance with the spatial perspective. Based on the literature review and practical analysis of pipe maintenance

management, six quantitative variables (the second level) from four aspects (the first level) are established for the LOF prediction in Table 5.4 and are described in more detail following the table.

Table 5.4: Influencing variables for the LOF at the neighborhood-level.

First-level variable	Second-level variable	Definition
Pipe status (A_1)	Average age (A_{11})	Mean of ages of all pipes
Surroundings (A_2)	Tree density (A_{21})	The ratio of the number of trees to the total area surrounding the pipes
	Population density (A_{22})	The ratio of the population to the total residential area
Traffic condition (A_3)	Traffic coverage (A_{31})	The ratio of pipe length along the road to the total pipe length
	Traffic volume (A_{32})	Annual average weekday daily traffic volume
Land use (A_4)	Merged area ratio (A_{41})	The ratio of land used for commercial, residential, and industrial purposes to the total area

- Pipe status (A_1). Although there are many variables listed in Table 5.3 that could be used to describe the pipe itself, only one second-level variable, the average age, is selected to represent pipe status in each neighborhood. Based on previous studies, age tends to be the most widely and commonly used factor in pipe deterioration analyses. Furthermore, some variables (e.g., material, diameter, or slope) are related to detailed features of individual pipes and are not appropriate for a neighborhood-level analysis.

- Surroundings (A_2). Since tree roots can enter pipes through cracks or failed joints (or seals), trees have the potential to influence pipe performance. Previous research has considered the number of trees as one of the factors for predicting the individual pipe condition (e.g., Syachrani et al., 2011; Tran et al., 2006). With respect to the neighborhood, the tree density (A_{21}) along the pipe is taken into consideration. For a given neighborhood, only the trees surrounding underlaying pipes (or within a distance

threshold from pipes) would give rise to an adverse effect. In order to make a horizontal comparison among neighborhoods, a relative index is defined to measure the tree density (A_{21}) where a distance threshold from pipes need to be pre-set (e.g., 10 m) based on a specific set of conditions, such as tree types and growth rate. Then, the buffer function, the intersection function, and the spatial join function in geographic information system (GIS) can be utilized to collect data for the number of trees within the distance threshold and the total surrounding area of pipes in terms of each neighborhood. Furthermore, population density in the residential area is selected as another second-level variable to describe the surroundings, since, in theory, the more populated an area is, the more sewage will be transported from residential housing through pipelines.

- Traffic condition (A_3). This first-level variable is described in terms of two aspects: traffic coverage and traffic volume. As for the former, sewer pipelines are typically designed to run alongside roads (or streets) except in the case of a specific community or park, etc. As such, the ratio of pipe length along the road to the total pipe length in a neighborhood can be used to represent the percentage of pipes that are influenced by surrounding traffic. Similar to the index for tree density (A_{21}), the data for traffic coverage (A_{31}) can be gathered with the support of GIS. Furthermore, annual average weekday daily traffic volume is another index adopted to measure how busy the road is, which indirectly represents the degree to which neighboring traffic affects the pipes.

- Land use (A_4). When it comes to the influence of land use on the sewer pipe condition, the land use type is considered. Two types of land use are considered in the neighborhood-level deterioration analysis, which is type one (including residential, industrial, and commercial uses) and type two (other uses such as agriculture, park, etc.).

The majority of sewer pipelines provide service for activities in areas of type one land use and sewage produced from activities in type one land use areas also influence pipe aging. Rather than treat land use as categorical (or nominal) variables, this study collects data pertaining to the land area for the type one land-use area and further calculates the area ratio for each neighborhood. The variable, merged area ratio (A_{41}), can be used to present the portion of the whole neighborhood that is type one area land use.

5.3.1.2 Model development

The multiple linear regression model, which is a statistical approach used to determine the relationship between the dependent variable and multiple independent variables, has been widely used in business, social science, engineering, and other disciplines (Neter et al., 1989). This study will adopt this approach to predict the LOF for neighborhoods based on historical data, which is expressed as Equation (5-1).

$$LOF_i = \beta_0 + \sum_{k=1} \beta_k X_{ik} + \varepsilon_i \qquad (5\text{-}1)$$

where LOF_i is the LOF for the neighborhood i, which is equal to the average values of LOF for all observed pipes in the neighborhood; X_{ik} is the value of the independent variable k in the neighborhood i and variable $k(k \geq 1)$ is derived from the influencing variables established in Table 5.4; β_k is the coefficient of variable k; and ε_i is the independent error term in the neighborhood i satisfying a normal distribution $N(0, \delta^2)$.

Since $E(\varepsilon_i) = 0$ the response function for the above regression model can be transformed into Equation (5-2). In other words, the dependent variables LOF_i satisfy the normal distribution $N(E(LOF), \delta^2)$.

$$E(LOF) = \beta_0 + \sum_{k=1} \beta_k X_k \qquad\qquad (5\text{-}2)$$

In order to estimate the coefficients β_k in Equation (5-1), a classical approach called the least-squares method is generally used, but it only reflects the parameter fitting. As in the case for uncertain models, we can select the stepwise regression (Neter et al., 1989) to search the optimal linear regression model among pre-listed independent variables in consideration of other parameters such as the goodness of fit (R^2), t-stat, Durbin-Watson test, and P-P plot for regression residuals.

5.3.2 Predicting the LOF for individual pipes

5.3.2.1 NN architecture design

Essentially, the prediction of the sewer pipe condition is to classify the sewer pipe into several categories (five LOF grades in our case, see Table 5.1) using certain input variables. Back-propagation NN has been widely used in the classification problem in civil engineering and has proven to be efficient (Adeli, 2001; Al-Barqawi & Zayed, 2008; Chatterjee et al., 2017; Zayed & Halpin, 2005). The design of the NN mainly focuses on five aspects, which are the number of input neurons, number of output neurons, number of hidden layers, number of neurons in the hidden layers, and the activation function for each neuron (Moselhi & Shehab-Eldeen, 2000). The structure design and parameter selection will be in consideration of two aspects in this research: 1) the common practices in previous research and implementation, 2) the experiments conducted with the best results (trial and error procedure). The determination of the number of input neurons and output neurons are straightforward since they will be equal to the number of input variables and output variables, respectively.

The number of input variables is determined by the specific problem to be solved. Note that, one nominal variable should be converted to several binary variables (dummy variables) based on the

number of the category for the variable (Zaman et al., 2015). For example, the variable of sewer type can be converted to three dummy variables as sanitary (true:1, false:0), stormwater (true:1, false:0), and combined (true:1, false:0). The number of output neurons is the number of classes, in our case, the number of output neurons is equal to five representing the five LOF grades. A three-layer structure (one input layer, one hidden layer, and one output layer) is employed to solve the classification problem since it has been proven to be sufficient and efficient for such problem (Al-Barqawi & Zayed, 2008; Moselhi & Shehab-Eldeen, 2000; Zayed & Halpin, 2005).

The activation function used in the hidden layer is the sigmoid function, which is the most commonly used activation function for a similar problem in the previous research and has proven to be efficient (Sousa et al., 2014; Syachrani et al., 2013). The softmax function is used as the activation function in the output layer, where the function is capable of solving the multiple classes prediction problem and commonly used in the output layer for the NN (Cha et al., 2017; Li et al., 2019).

As for the number of hidden neurons, there is not a fixed number that can be used for any NN, and it can be decided according to the results of the experiment (Moselhi & Shehab-Eldeen, 2000). In this research, we use classification accuracy (which will be defined in Section 5.3.2.3) as the benchmark for the parameter (number of neurons in the hidden layer) selection. For each trial, we test the accuracy for the different number of neurons in the hidden layer. Note that, the NN training process is a nondeterministic process, which will generate the near-optimum solution (instead of the optimum solution) after each training process. Thus, several experiments (20 repeated experiments are used in this research) should be performed with the same parameters to calculate the average accuracy for later comparison purposes. Figure 5.2 is an example of the number of hidden neurons selection. As the number of hidden neurons increases,

the classification accuracy shows an increasing trend. In some circumstances, the larger number of hidden neurons give a lower accuracy than a smaller number of hidden neurons; however, the overall trend is increasing until reaching an upper limit (which is around 0.955 as shown in Figure 5.2). The curve shows no significant growth trend after 35 hidden neurons, therefore, 35 will be selected as the number of neurons in the hidden layer. In the later implementation stage, the same methods will be applied to select the number of neurons in the hidden layer. Overall, the architecture of the proposed NN is presented in Figure 5.3. The mechanism of NN has been discussed extensively, and the interested reader may consult Chong & Zak (2013) for a detailed description of the mechanism underlying NN.
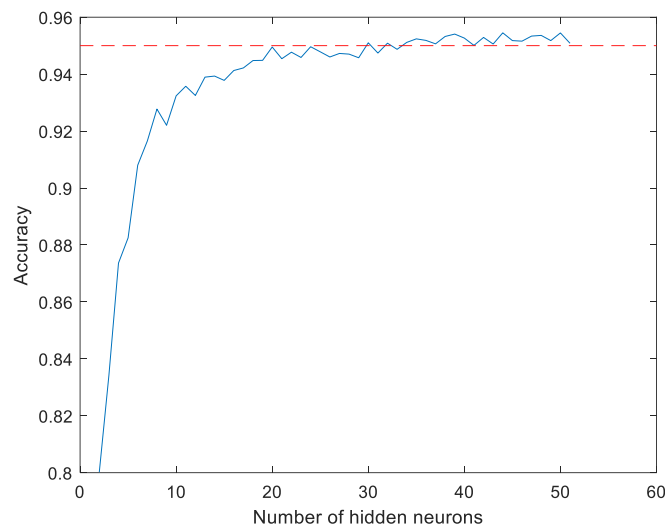
Figure 5.2: Classification accuracy for the different number of neurons in the hidden layer.
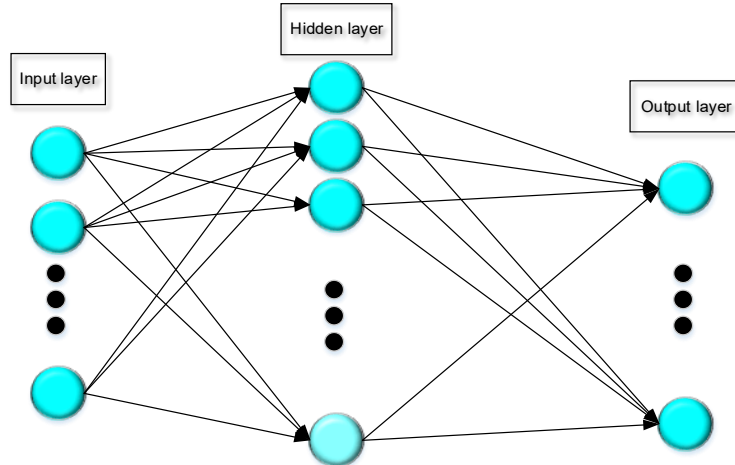
Figure 5.3: The architecture of the NN.

5.3.2.2 Backward elimination for variable selection

Variable selection is also known as feature selection, which aims at selecting an appropriate subset of variables for the model development, and the process could benefit the modeling process in three aspects: 1) improving the model performance, 2) providing faster and cost-effective input variables and 3) simplifying the models and making it easy to be understood (Guyon & A, 2003). The backward variable elimination is one of the commonly used variable selection methods and has been successfully used in previous prediction models developments, such as a prediction model for electricity demand (Vu, Muttaqi, & Agalgaonkar, 2015), a prediction model for the acute myocardial infarction mortality (Austin & Tu, 2004), etc. Thus, the backward variable elimination method is adapted in this research for selecting the proper input variables for the individual-level deterioration model. The variables used in the previous research have been reviewed in Section 5.2.2 and summarized in Table 5.3. As mentioned before, the frequency in Table 5.3 indicates the relative importance of the variables to some extent. The variables collected in our case will be sorted according to the importance degree as shown in the table. Then, a backward elimination is performed to determine which variables should be included in the model and which should not. One of the motivations for using the variable

150

elimination algorithm is the simplification of the model since only the variables that are statistically significant will be kept, which would allow achieving optimal accuracy. In fact, some variables are redundant and including them in the model will not contribute to increasing the classification accuracy. In practice, inputting more variables means more work in the data collection process. For example, if we determine that the capacity of the sewer pipe cannot influence the classification accuracy in our experiment, then we do not have to collect it in the process of any future deterioration model development. The criteria for determining whether a variable need to be eliminated is to check if it impacts the classification accuracy significantly. To some extent, it is a subjective judgment to decide if a variable should be included or not, since "significantly" is a subjective feeling by the model developer. In this research, we will try to use this procedure to conduct several experiments, and the results will be compared. During the comparison of the accuracy derived from each step, critical variables that can affect the model performance significantly will be identified. These critical variables will be fed into the NN model to compare the performance again with the experiments conducted before. Finally, the input variables will be determined after comparing all the experiment results.

5.3.2.3 Model validation and analysis

The metrics of precision, recall, and accuracy, shown in Equation (5-3), Equation (5-4), and Equation (5-5), respectively, are adopted in this research to measure the performance of the prediction model, where the definitions of TP, FN, FP, TN are summarized in Table 5.5. This measurement metric is widely accepted in measuring the performance of the classification problem (e.g., Chatterjee et al., 2017; Kumar et al., 2018; Li et al., 2019). The precision indicates the correct rate for a specific category, while the recall is the rate of completeness of the classified sample for a specific category (Martínez-Rojas et al., 2015). The accuracy gives the

percentage of the correctly classified samples by the model (Li et al., 2019). The confusion matrix will be adopted to interpreter the classification results (Stehman, 1997).

Table 5.5: Definition of TP, FN, FP, TN.

	True condition	
	Positive	Negative
Predicted Positive	True Positive (TP)	False Positive (FP)
Predicted Negative	False Negative (FN)	True Negative (TN)

$$\Pr ecision = \frac{TP}{TP+FP} \tag{5-3}$$

$$\mathrm{Re} call = \frac{TP}{TP+FN} \tag{5-4}$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{5-5}$$

5.4 Case study

5.4.1 Data collection

The city of Edmonton, Canada, is selected as the study area and its drainage systems are maintained by a municipal management company called EPCOR Drainage Services. The distribution of all sewer pipes in this city is presented in Figure 5.4. For the neighborhood prediction, two primary data sources are used in the present research. (1) The dataset used for this study contains pipe information including age and LOF from EPCOR Drainage Services for approximately 5,200 km (around 76,000 sewer pipes) of various kinds of sewer pipes, which accounts for 76% of the total pipe length in Edmonton, after data cleaning and filtering to exclude some data that contains missing information. (2) The dataset includes information regarding neighborhoods from the City of Edmonton's Open Data Portal (EODP), such as trees, population, traffic volume, and road network. The city is composed of 396 neighborhoods, but

only 382 neighborhoods are selected as the region units in the analysis due to data availability of sewer pipes. The data for influencing variables of the neighborhood LOF is then calculated with the support of GIS after collecting shapefiles for the above items at the neighborhood-level.

For the development of the individual pipe prediction, 9,892 pipes (685.266 km) are randomly selected from the original dataset to get a balanced dataset across the five levels of LOF. The reason for this random selection is that a larger data sample requires much more computing power, which is limited in our experiment. In addition, since the sample is selected randomly from the original dataset, it should contain most of the features in the original dataset: 75% of the data is used for training, while 15% and 15% of the data are used for validation and testing, respectively. These particular proportions for data separation are widely used in different model training processes (e.g., Cheng & Wang, 2018; Li et al., 2019). The samples used for training the model are evenly distributed among the five levels of LOF, which will prevent the model from overfitting for minority class (Li et al., 2019). For example, if most of the pipes in the dataset are in LOF 1 while the number of rest of the pipes (LOF 2-5) is much less than that of LOF 1, then the model tends to give a high accuracy for the pipe in LOF 1 while poor accuracy for pipes in other classes when we test the model with a generalized dataset.
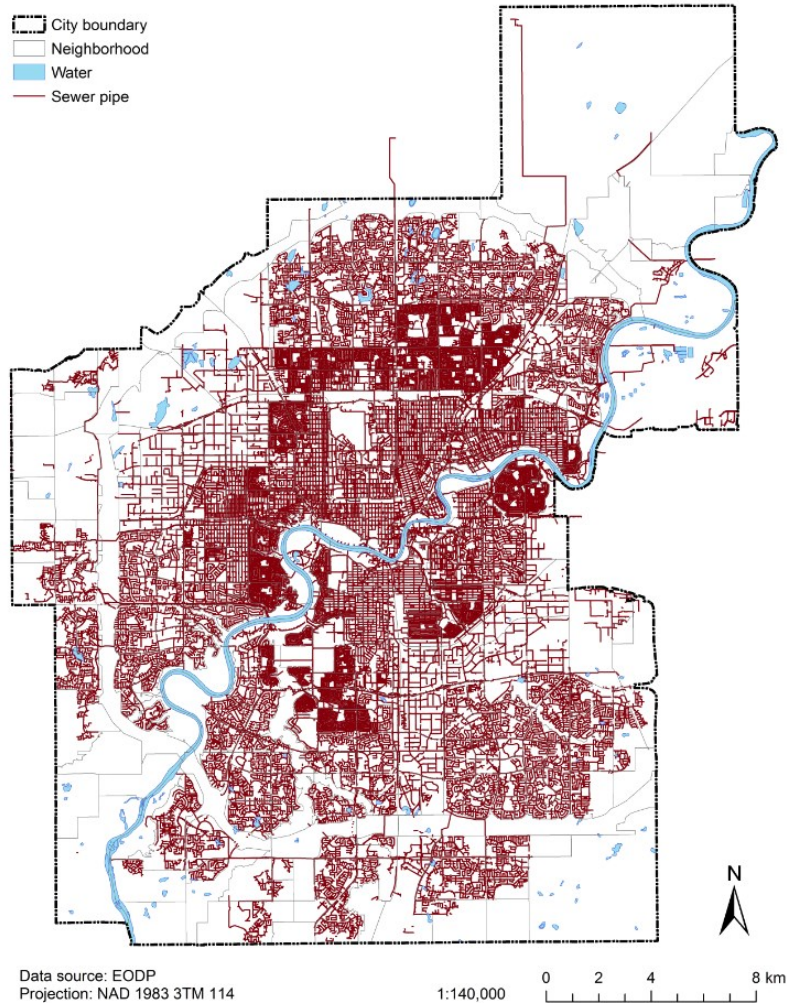
Figure 5.4: Distribution of sewer pipes in Edmonton, Canada.

5.4.2 Deterioration model for neighborhoods

5.4.2.1 Descriptive statistics

In order to clearly grasp the data underlying the neighborhood LOF and its influencing variables,

Table 5.6 provides the basic statistical features of all variables from all observed neighborhoods.

Of them, a significant imbalance exists in terms of traffic volume of all neighborhoods, while

regional differences in terms of population density, traffic coverage, and merged area ratio are

relatively minor. Specifically, the average age of pipes in the neighborhoods concentrating on the

city center is relatively high, while the pipe age tends to decrease as the distance from the center

154

of the city increases. If there is no residential area in a neighborhood, the population density is equal to 0, as such neighborhoods are primarily designed for industry. The neighborhoods with traffic coverage of less than 0.8 account for around 40% of the total observed neighborhoods. In addition, the neighborhood LOF is visualized in the 382 neighborhoods as shown in Figure 5.5. There is a clear pattern that as the city expanding outwards, the sewer pipes' condition gets better since they are newer pipes. It can be seen that the pipes in relatively bad condition (i.e., the value of neighborhood LOF exceeds 2) are primarily distributed in the central part of this city, and are surrounded by other neighborhoods with pipes in better condition.

Table 5.6: Descriptive statistics of the variables.

Variable	Unit	Min.	Max.	Mean	Standard Deviation
Average LOF	–	1	3.84	1.95	0.74
Average age	Year	3.07	68.74	32.84	15.87
Tree density	100No./km^2	0	63	14	12.40
Population density	person/m^2	0	0.21	0.01	0.01
Traffic coverage	–	0	0.97	0.66	0.23
Traffic volume	Vehicle/axle	0	65,925	9,738	9,637.31
Merged area ratio	–	0	0.92	0.43	0.24
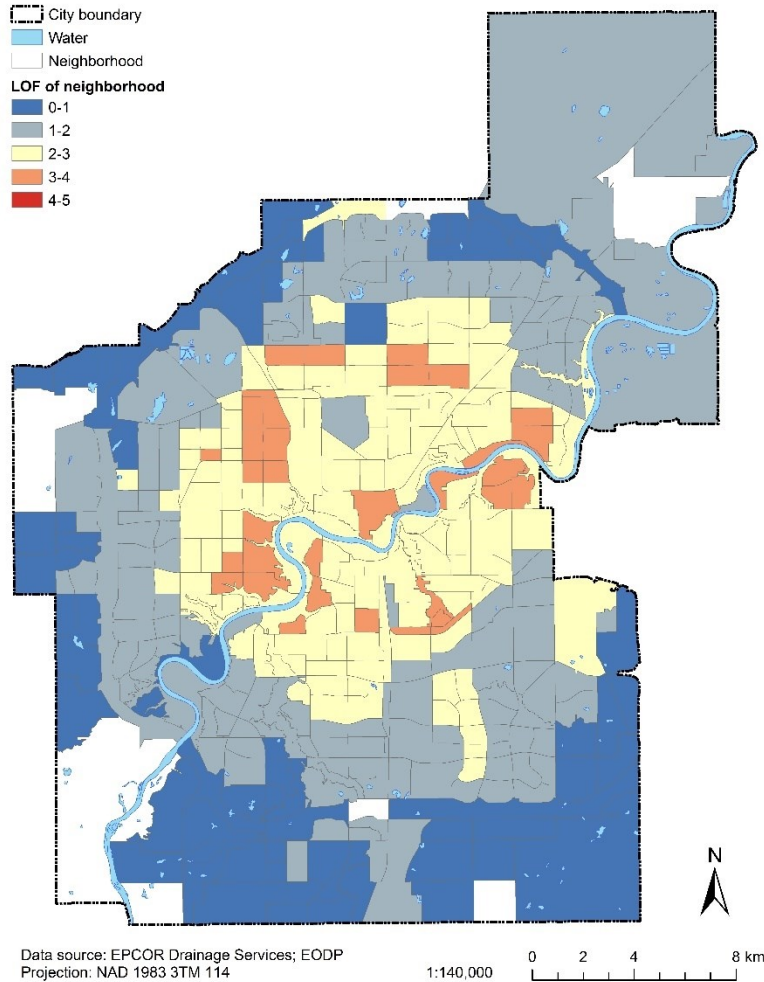
Figure 5.5: Neighborhood-level LOF of sewer pipes in Edmonton.

5.4.2.2 Model results

This study uses SPSS 24.0 software as the primary tool and builds the multiple linear regression model for the average LOF of neighborhoods (see Table 5.7). It can be seen that the fitness of the developed model is accepted since the R^2 or adjusted R^2 has a reasonable value. Meanwhile, *p*-values for all the variables selected for the models show that all coefficients pass the 0.01 significance level test. After using the stepwise regression, only three independent variables are entered into the model. One interesting result from the model is that the average LOF at the neighborhood-level has a negative correlation with traffic coverage. That implies, counterintuitively, that the higher the traffic coverage, the lower the average LOF. A possible

156

reason for this may be that the pipelines close to the road network are convenient in terms of inspection and maintenance, which means they attract more attention from facility managers, who may in turn, schedule the inspection activities more frequently. Furthermore, the results reveal that the average age of pipes and the tree density within the distance threshold can contribute to a higher average neighborhood-level LOF, where the age is the dominant factor in the performance of aging pipes.

Table 5.7: Linear regression for predicting the average LOF.

Variable	Coefficient	t-stat	Significance
Constant	0.806	11.352	0.000
Average age	0.041	36.249	0.000
Tree density	0.005	3.559	0.000
Traffic coverage	-0.391	-4.899	0.000

Note: model goodness-of-fit $R^2 = 0.788$ and adjusted $R^2 = 0.786$.

Compared with Table 5.4, we can see three variables (i.e., population density, traffic volume, and merged area ratio) are excluded in the regression model. In the case of population density, a higher variation would be observed if a city has a wider variety of housing types (e.g., apartments and houses) and a higher number of residential buildings at the neighborhood level. In the case study, the house (e.g., single-family or townhouse) is the most common residential type, which also explains why the standard deviation of population density in Table 5.6 is very small. As for traffic volume, the possible reason this variable is excluded from the regression model is data collection bias. The sites used to record this value can not provide coverage for all roads in this city, so the average traffic volume in a given neighborhood can not reflect the real value, although some pipelines are alongside the road network within a distance threshold. Merged area ratio has a positive correlation with the LOF, but the coefficient can not pass the significance test.

5.4.2.3 Model validation

In addition to the values of the R^2 and adjusted R^2 presented in the above section to evaluate the performance of the regression model, the model is further validated with various approaches. First, the Durbin-Watson value for the model is 1.961, which, being located between 1.5 and 2.5, indicates no autocorrelation exists among the residuals in the sample (Neter et al., 1989). Second, the model is validated by plotting the estimated neighborhood LOF against the observed values (see Figure 5.6(a)). The method is widely accepted to compare the deviation between the observed values and the estimated values (i.e., Zaman et al., 2015). In the best-case scenario, where the model can predict the real-condition values correctly, all the points should be distributed along the 45-degree reference line (i.e., diagonal line). From Figure 5.6(a), we can visually conclude that the points are moderately close to the diagonal line, which are considered as acceptable for the regression model in terms of prediction accuracy, but certainly has potential for improvement for future research. Third, the P-P plot is used to investigate the normality of the residuals (which is an assumption for the linear regression analysis) by plotting the observed cumulative probability of standardized residuals and their expected values. Similarly, for the case to be acceptable, the plotted points should scatter closely along the diagonal line, which means that the residuals satisfy a normal distribution. Therefore, the plot is shown in Figure 5.6(b), where points are closely located around the diagonal line, indicates the residuals in the regression model satisfy the normal distribution. Specifically, the mean is 5.59E-4, and the standard deviation is 0.995 based on the 382 observations. In summary, the validation process shows that there is no autocorrelation exists among the residuals, the predicted values and the observed values are close with each other, and the residuals satisfy the normal distribution assumption.

Thus, the developed regression model is considered as accepted in this research, and further improvement can be made in future research.
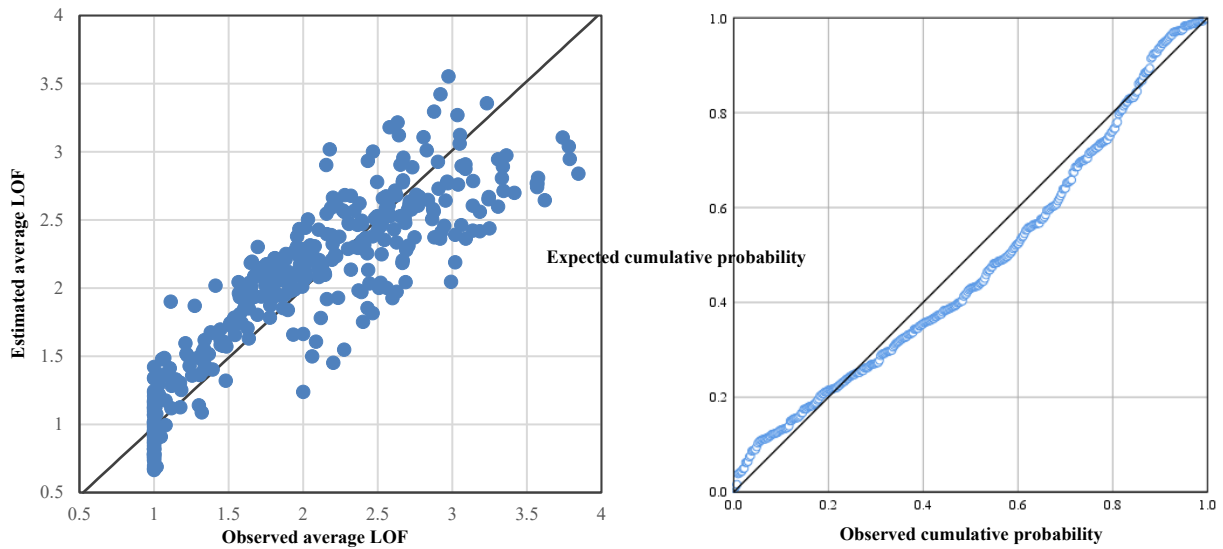
Figure 5.6: (a) Estimated vs observed average LOF; (b) P-P plot for regression residuals.

5.4.2.4 Neighborhood deterioration analysis

From the regression model, we know three variables are selected for predicting the LOF of neighborhoods, namely, average age, tree density, and traffic coverage. With the prediction model, the changing trend of the LOF (in the condition of no maintenance over time) can be depicted as shown in Figure 5.7. Where, the average age is set to increase by 10, 20, and 30 years from the year of 2018, and the figure representing the pipe condition in 2018 can refer to Figure 5.5. It can be seen that the percentages of neighborhoods in orange and red color will grow faster, most of which are concentrated in the city center. Meanwhile, the neighborhoods with the smaller LOF in 2018 will be gradually replaced with those with the slightly larger LOF. Specifically, the percentages of neighborhoods with medium and larger LOF (larger than 3) will be accounted for 70.4% of the total amount by the year 2028, which increases 52.8% compared with this value in 2018. When the pipes in these neighborhoods have been used for another 10–

20 years with no maintenance, there are much more neighborhoods are projected to arrive at a situation of LOF over 3. The figure proves the significance of the preventive maintenance for sewer pipes to keep the sewer system at an acceptable level of service (LOS). In addition, city managers could have a clear overview of the sewer pipe condition in the city-wide. The trend is clear that neighborhoods near the city center deteriorate into higher LOF (larger than 3) faster than others. More attention should be paid on the neighborhoods that transfer to orange or red in the next 10–30 years by assigning a higher priority on scheduling the maintenance operations (such as high-pressure flushing (HPF), CCTV inspection, etc.).

(a) The year of 2028

(b) The year of 2038

(c) The year of 2048

City boundary
Water
Neighborhood
LOF of neighborhood
0-1
1-2
2-3
3-4
4-5

N

0   2   4   8 km

1:138,000

Data source: EPCOR Drainage Services; EODP
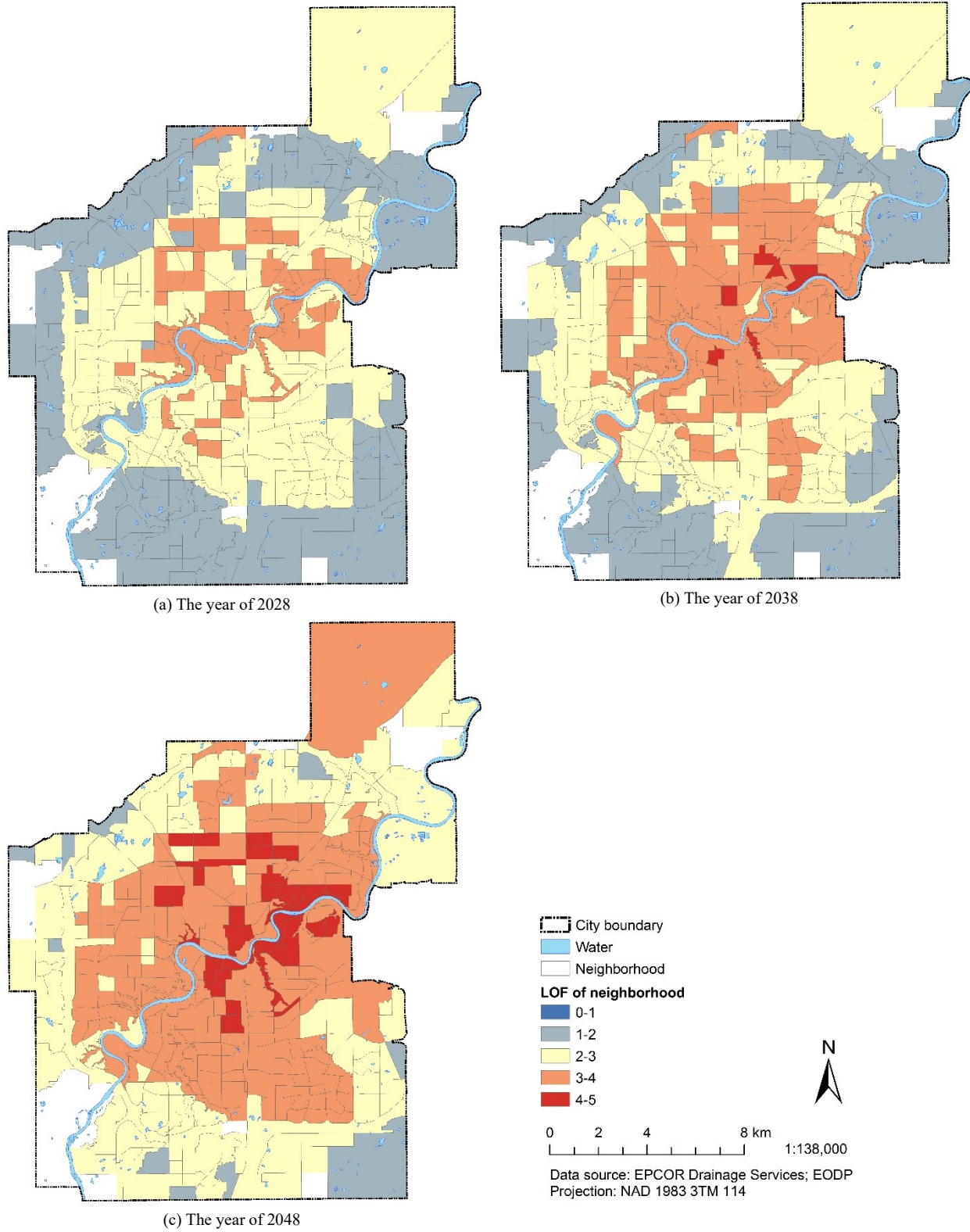Projection: NAD 1983 3TM 114

Figure 5.7: Neighborhood deterioration.

5.4.3 Deterioration model for individual pipes

5.4.3.1 Variable selection

The variables that are available for our case study are shown in Table 5.8. The description of the variables, the range of the values for each variable, variable type, and the relative ranking are also listed in Table 5.8. It is important to note that the age calculated in the case study is the number of years between the time the pipe was installed and 2018; however, if the pipe is recorded as having been renewed, then the age is calculated as the number of years between the most recent time it was renewed and 2018. If the pipe was renewed, then its LOF will change to 1 or 2; therefore, in the case study, we consider that pipe as a new pipe at the time of it was renewed. The rank of the variable is determined according to the frequency (importance) of the variable, as shown in Table 5.3. With this order, the backward elimination process will be conducted as presented in Table 5.9. Step 1 (iteration 1) will include all the available variables, followed by eliminating one variable with the lowest rank in Table 5.8 in each subsequent iteration. The included variables in each step will serve as the input for the NN with a structure as indicated in Figure 5.8. For those selected variables in each step, the number of neurons in the hidden layer still needs to be determined by calculating the average accuracy derived from the different number of hidden neurons, which has been described in Section 5.3.2.1. After comparing the accuracy curve for each step, several critical variables can be identified. These critical variables will be identified and fed into the NN model.

Table 5.8: Summary of the input variables.

Variable	Description	Range	Variable type	Rank
Age	The age of the pipe segment in 2018	1–111 year	Scale	1
Diameter	The diameter of the pipe segment	150–3150 mm	Scale	2

Length	The length of the pipe segment	1–313.21 m	Scale	3
Material	The material of the pipe segment	Concrete, clay/asbestos, PVC, others	Nominal	3
Slope	The slope of the pipe segment	-28.6%–55.05%	Scale	3
Average of LOF	Average LOF for the neighborhood	1–3.84	Scale	6
Waste type	The type of waste transported by the pipe segment	Sanitary, stormwater, combined	Nominal	7
Up/Down stream depth	Pipe buried depth	0–42 m	Scale	8
Repair history	The number of times that the pipe segment has been repaired	0–4	Scale	9
Up/Down stream invert elevation	The invert elevation the sewer pipe	0–732 cm	Scale	10
Capacity	The water flow rate inside the pipe segment	0–67.633L/s	Scale	11
Category	The category that the pipe segment belongs to	Large trunk, small truck, local sewer	Nominal	12

Table 5.9: Backward variable elimination process.

Iterations	Included variables
Step 1	Age, Diameter, Length, Material, Slope, Average of LOF, Waste type, Up/Down stream depth, Repair history, Up/Down stream invert elevation, Capacity, Category
Step 2	Age, Diameter, Length, Material, Slope, Average of LOF, Waste type, Up/Down stream depth, Repair history, Up/Down stream invert elevation, Capacity
Step 3	Age, Diameter, Length, Material, Slope, Average of LOF, Waste type, Up/Down stream depth, Repair history, Up/Down stream invert elevation
Step 4	Age, Diameter, Length, Material, Slope, Average of LOF, Waste type, Up/Down stream depth, Repair history
Step 5	Age, Diameter, Length, Material, Slope, Average of LOF, Waste type, Up/Down stream depth
Step 6	Age, Diameter, Length, Material, Slope, Average of LOF, Waste type
Step 7	Age, Diameter, Length, Material, Slope, Average of LOF
Step 8	Age, Diameter, Length, Material, Slope
Step 9	Age, Diameter, Length, Material
Step 10	Age, Diameter, Length
Step 11	Age, Diameter
Step 12	Age

5.4.3.2 Model results

Part (a) of Figure 5.8 shows the results of the 12-step variable elimination process. We can see very clearly that the results can be divided into four clusters, which are cluster 1 (step 1, the blue curve on the top of part (a) of Figure 5.8), cluster 2 (steps 2–6, part (b) of Figure 5.8), cluster 3 (steps 7–8, part (c) of Figure 5.8), and cluster 4 (steps 10–12, part (d) of Figure 5.8). For each step, the accuracy (see Equation 5-5) increases drastically as the number of neurons in the hidden layer increase at the initial stage (around 2 to 10 hidden neurons). Then, the increasing trend becomes more moderate and finally reaches a stable level. The highest accuracies achieved by the four clusters with different numbers of hidden neurons are 0.955 (step 1 with 44 hidden neurons), 0.932 (step 2 with 49 hidden neurons), 0.870 (step 9 with 49 hidden neurons), and 0.681 (step 11 with 46 hidden neurons).
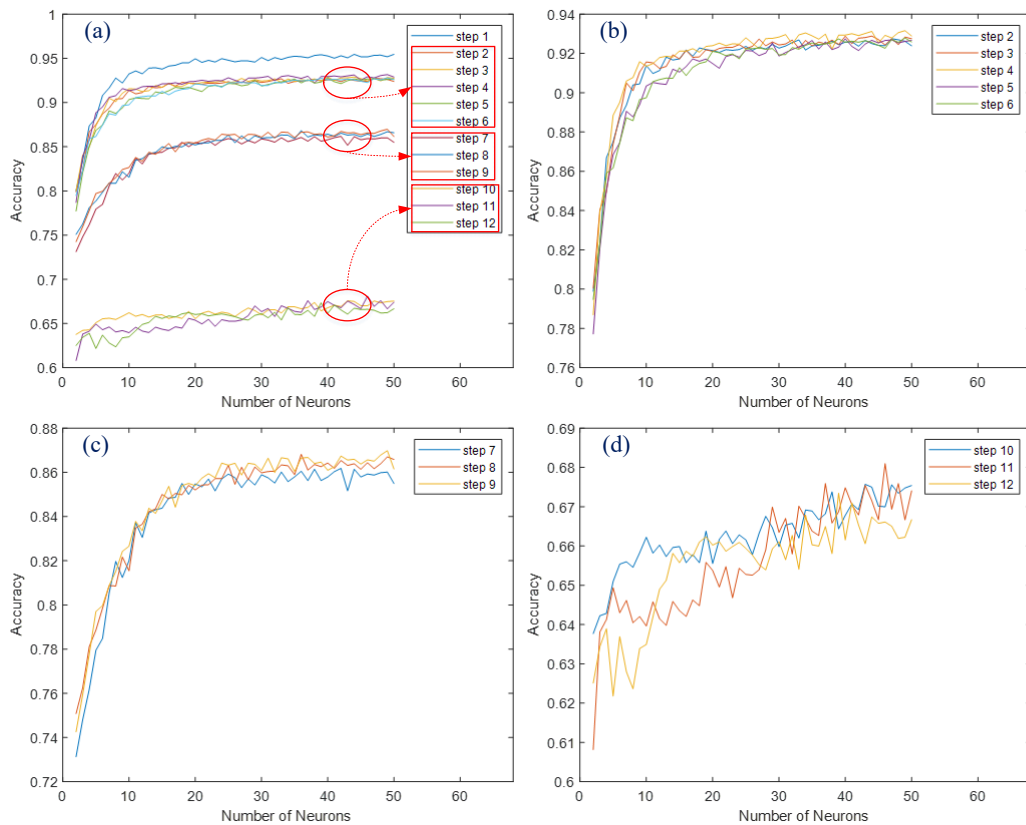

Figure 5.8: Variable elimination process.

There is a significant decrease in accuracy between steps 1 and 2, steps 6 and 7, and steps 9 and 10. With respect to the accuracy of the prediction model, clear gaps can be observed between the four clusters shown in part (a) of Figure 5.8, while there is no significant difference in accuracy within each cluster. For example, the accuracy of cluster 2 (which includes steps 2 to 6, part (b) of Figure 5.8) is approximately 0.92. The variables eliminated within cluster 2 are capacity, up/down stream invert elevation, repair history, and up/down stream depth, successively in that order, from step 2 to step 6. When the variable of waste type is eliminated (in the transition from step 6 to step 7), the accuracy decreases to approximately 0.85, where cluster 3 (which includes step 7 to step 9) can be observed. In other words, when the waste type variable is eliminated, the model accuracy drops significantly, while eliminating the variables of capacity, up/down stream invert elevation, repair history, and up/down stream depth produces no significant decrease in the accuracy of the model. Based on this, we can conclude that waste type is an important input variable. Similarly, the elimination of category and of material as part of the backward variable elimination process both cause significant decreases in accuracy (see the gap between clusters 1 and 2, and the gap between clusters 3 and 4 shown in Figure 5.8). Thus, the variables of category and material are identified as critical input variables based on the same logic. Furthermore, in the backward variable elimination process, the average LOF is eliminated between step 7 and step 8 as per Table 5.9, and it can be observed that there is no significant difference in the accuracy between step 7 and step 8 (as shown in parts (a) and (c) of Figure 5.8). Therefore, the average LOF variable is not considered as a critical variable since it does not contribute significantly to the prediction accuracy in this case study. It should be noted that the average LOF (or any other of the eliminated variables in this case study) could be identified as a critical variable in other case studies. Therefore, several critical variables that could significantly influence the accuracy

165

of the model can be identified as the category, waste type, and material; however, the category has the least influence among those three variables. The results of the 12-step variable elimination process are summarized in Table 5.10.

Table 5.10: Summary of the results of the 12-step variable elimination.

Cluster	Highest accuracy	The step with the highest accuracy	Number of hidden neurons
1	0.955	1	44
2	0.932	2	49
3	0.870	9	49
4	0.681	11	46

Among the four clusters, cluster 4 (part (d) of Figure 5.8) is the least acceptable model since it only has an accuracy of 0.681, while cluster 3 (part (c) of Figure 5.8) has the highest accuracy of 0.87, which is much better than cluster 4. Step 9, which has the highest accuracy in cluster 3, will be selected as the base model for further improvement. The reason to choose step 9 as the base model for further operation is that it has an acceptable classification accuracy, and which also makes sure that further operation on the base model will retrieve the least accuracy of around 0.87. As mentioned above, the significant decrease in accuracy among the four clusters is caused by the elimination of the variables of category, waste type, and material. Step 9 includes the material variable, thus, further improvement of the model in step 9 would involve adding the other two critical variables (category, waste type) into the base model step by step. These further operations on the base model are named step 13 and step 14.

The variables included in each of these two steps are summarized in Table 5.11. The results of step 13 and 14 are plotted in Figure 5.9, which also consists of the best curve in cluster 1 (step 1), 2 (step 2), and 3 (step 9). From the figure, we can see that the accuracy of step 13, which has an extra variable of waste type when compared to step 9, shows a significant improvement compared to step 9. Improvements also occur when adding another variable of the category into

166

step 13. The accuracy curve of step 14 is slightly higher than step 1, generally speaking, and the highest accuracy of step 14 is 0.957 (45 hidden neurons), which is slightly higher than step 1 (0.955) that includes all the collected variables. Overall, step 14 has the best performance in terms of classification accuracy; however, it contains only 6 variables (6 variables less than step 1). Therefore, step 14 will be selected as the final deterioration model considering that it has the best performance while needing fewer input variables. The final model can be concluded as a three-layer NN (6–45–5), with the input variables of age, diameter, length, material, waste type, category.

Table 5.11: Summary of variables in steps 13 and 14.

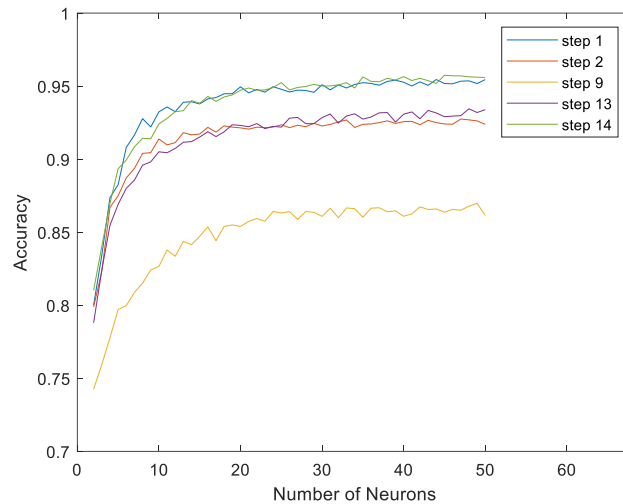| Iterations	Included variables
Step 13	Age, Diameter, Length, Material, Waste type
Step 14	Age, Diameter, Length, Material, Waste type, Category

Figure 5.9: Include identified critical variables.

For the final model, the confusion matrix of the classification results is presented in Table 5.12. The precision and recall are also calculated as shown in Equation (5-3) and Equation (5-4), respectively. The overall accuracy is 0.955, while the precision and recall are higher than 0.92.

167

Notably, the values of precision and recall are better for LOF 1, LOF 4, and LOF 5 than that of LOF 2 and LOF 3. That may be explained by there being a smaller magnitude in the differences between the conditions of pipes in LOF 2 and LOF 3 when compared to any other two adjacent classes. From the table, we can see that there are 14 pipes in LOF 3 that were predicted as being classified in the class of LOF 2, which is the largest number of misclassifications. It is important to note that the misclassified samples are mostly misclassified as an adjacent class. For example, there is no pipe in LOF 1 that is misclassified as LOF 5, and vice versa, which is a positive outcome for the proposed model.

Table 5.12: Confusion matrix for the prediction model.

		True condition							
		LOF 1	LOF 2	LOF 3	LOF 4	LOF 5	Precision	Recall	Accuracy
	LOF 1	**290**	4	2	0	0	0.980	0.970	
	LOF 2	7	**267**	14	1	0	0.924	0.960	
Predicted	LOF 3	2	7	**309**	8	3	0.939	0.925	0.955
	LOF 4	0	0	7	**281**	4	0.962	0.949	
	LOF 5	0	0	2	6	**270**	0.971	0.975	

5.4.3.3 Sewer pipe deterioration analysis

With the developed classification model, deterioration curves that present the LOF trend regarding pipe age can be plotted. As mentioned above, variables of material, waste type, and category influence the model significantly; therefore, the deterioration curves are plotted separately for pipes in these different classes (see Figure 5.10–5.12).

Figure 5.10 shows that the deterioration patterns are different for pipes of different materials, all else characteristics being equal. For example, the deterioration curves in Figure 5.10 are shown for pipes that transport sanitary waste and that are classified as local sewers. The reason that

clay/asbestos pipe and concrete pipe are classified as LOF 2 at an early age may be due to the age calculating method (which has been mentioned above), where if the pipe has been renewed, we calculate the age by using 2018 minus the year of the most recent pipe renew time. A pipe that has been renewed will have a LOF of 1 or 2 in the beginning years. In addition, pipes in LOF 1 and 2 classes are almost the same since both LOF 1 and 2 classes are pipes in very good condition that should not be considered for preventive maintenance (i.e., inspection) in the near future. The PVC pipes have a faster deterioration rate and are predicted to be classified as LOF 4 at the age of 47, while the corresponding age is 58 and 59 for clay/asbestos pipe and concrete pipe, respectively. However, all the pipes will be classified as LOF 5 at around the same age (64, 65, and 68 for clay/asbestos pipe, concrete pipe, and PVC pipe, respectively).
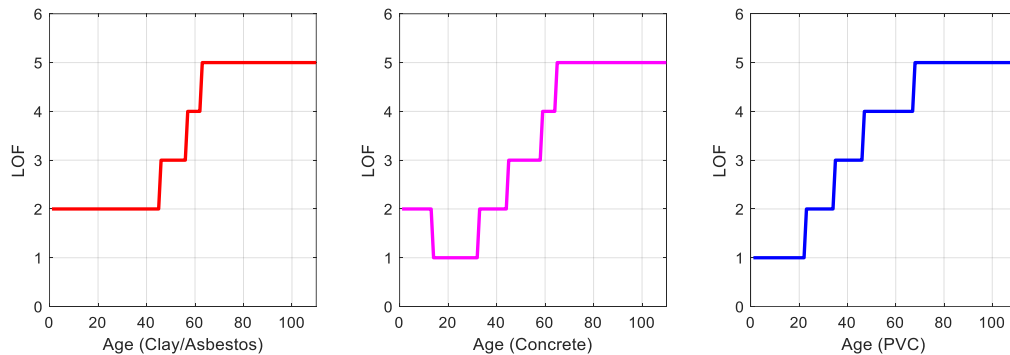
Figure 5.10: Material (sanitary, local sewer).

Figure 5.11 shows the deterioration curves for pipes transporting different kinds of waste, namely, sanitary, stormwater, and combined (transporting both sanitary and stormwater). The combined pipe is the last one among the three to reach the LOF of 4 (at the age of 97 for combined pipe, 57 for sanitary pipe, and at the age of 84, the stormwater pipe reaches LOF of 5). In fact, the combined pipe should be the oldest pipe in the city, since only combined pipes were used at an early age. However, many of these pipes were renewed several times to keep them in functional condition. Therefore, few combined pipes are in LOF of 5 which resulting in the

model classifies no pipes in LOF of 5. In addition, as mentioned above, the pipe age is recalculated if the pipe was renewed, thus, most of the combined pipes in the age around 40 have the best condition. The stormwater pipe deterioration rate is slower than the sanitary, which may be because the waste that is transported by sanitary pipes contains more complex compositions (e.g., grease or solid matter) that may cause corrosion or condensation. The curves suggest that sanitary pipes should be paid more attention to at around the age of 57 and at the age of 84 for stormwater pipes.
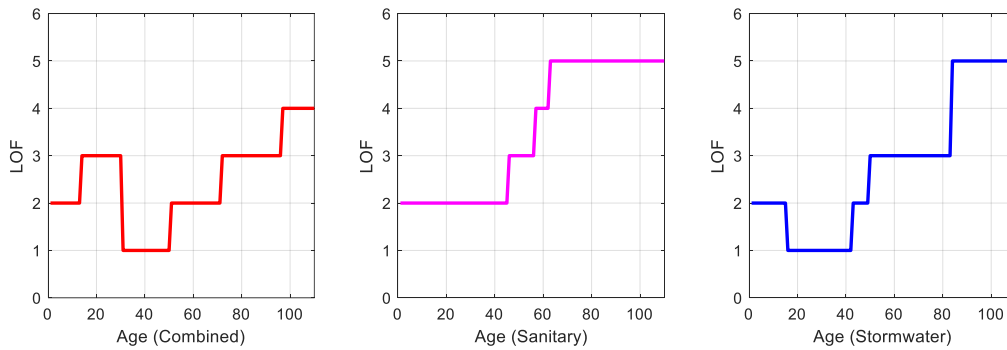
Figure 5.11: Sewer type (Clay/Asbestos, local sewer, diameter=300mm, length=100m)

Pipes classified in different categories (i.e., large trunk, local sewer, and small trunk) also show a different pattern of deterioration (see Figure 5.12). Overall, the large trunk deteriorates at the lowest rate compared with the local sewer and small trunk. The reason for that maybe that the large trunk with a larger diameter (usually greater than 1200mm) will garner more attention and be under stricter supervision during construction, which may lead to a better bedding or surrounding conditions. The local sewer (100mm–550mm) and small truck (600mm–1200mm) show a similar deterioration pattern. They all require a high priority maintenance schedule at around the age of 60.
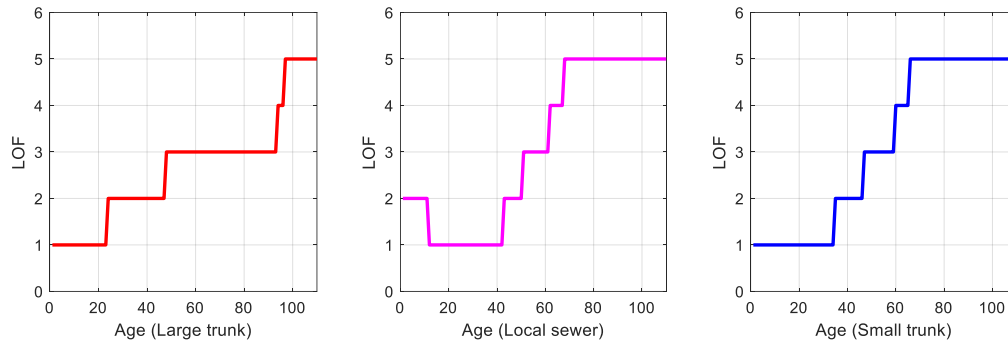
Figure 5.12: Category (Concrete, stormwater, Large trunk=1500mm, local sewer =300mm, small truck=900mm).

The diameter and length of the pipe also have some influence on the deterioration rate (see Figure 5.13). For the concrete stormwater pipe at the age of 70, the large the diameter of the pipe, the better the condition, which is consistent with the results in Figure 5.12. As for the influence of length for concrete sanitary pipe, at the age of 50, the longer the length, the worse the condition. The priority for maintenance should be given to pipes with small diameters and longer lengths.
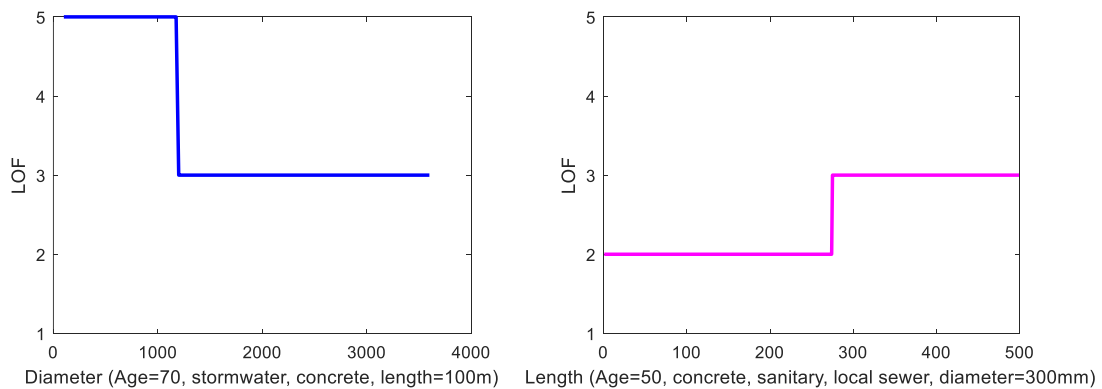
Figure 5.13: Influence of diameter and length.

5.5 Discussion

The proposed model to predict the condition of pipes at the neighborhood-level and individual-level are derived from the historical data, which means that the results can only be applied for the purpose of creating a maintenance schedule for the given sewer network understudy for which

171

the historical data was provided. The contradictions about the influence of some variables on the pipe deterioration exist between this research and some of the previous studies, which is a common phenomenon as discussed in the literature review. For example, the depth of cover may be identified as an important variable for the deterioration model; however, this variable has been eliminated in this study since it does not contribute to improving the performance of the deterioration model. The consistency of some variables in terms of the influence on pipe deterioration also exists. For example, it is widely agreed that different materials have different patterns of deterioration, and this has been confirmed by this research as well. Overall, although the results of the model may not be applied to all sewer pipe systems, the proposed methods can be applied in various kinds of deterioration model development for other civil infrastructures (e.g., bridges, tunnels, pavement, etc.).

The criterion used to determine whether a variable is excluded from or included in the backward variable elimination process conducted in this study is based on whether the variable contributes significantly to decreasing or increasing the model's accuracy. If the elimination of a variable will not cause an obvious or significant decrease in accuracy, then it is determined that the variable can be eliminated. On the other hand, if the elimination of the variable will cause a significant decrease in accuracy, then the variable will be identified as one of the critical variables that must be included in the final model. In our case, the variables of category, waste type, and material are identified as critical input variables since in each case the variable's elimination can be seen to cause a significant drop in the prediction accuracy of the model, and, for this reason, these variables are added back into the model in steps 13 and 14. The results in Figure 5.9 show that, when adding these critical variables back into the model, the prediction accuracy increases significantly, which further proves that these variables are critical to the

172

prediction model. In this research, the neighborhood-level LOF prediction is firstly proposed to predict the sewer pipe condition in city-wide. Then, the neighborhood prediction was tested to determine whether it has a significant influence on the performance of the individual-level prediction model. As a result of the backward variable elimination experiment, the neighborhood-level LOF was excluded from the individual-level predictors for the final model in our case study since including the neighborhood LOF does not contribute to increasing the accuracy of the individual-level prediction. However, the situation may be different in the case of other cities because the model is determined based on historical data for the specific sewer network, as described above. Since the proposed framework can be generalized for developing deterioration models for sewer pipe systems in other cities or even other types of city infrastructure, the neighborhood LOF could be identified as an essential input variable that will be included in the individual-level deterioration model after the variable selection process proposed in this paper for other cases. Therefore, it is necessary that the neighborhood LOF serves as the original input for the individual-level deterioration model.

The proposed bi-level deterioration model could be an effective tool for streamlining sewer pipe maintenance practice. For instance, the practice of CCTV inspection of sewer pipes, which is one of the most widely used inspection methods, includes both on-site CCTV collection and off-site CCTV video assessment. The development of the on-site CCTV video collection schedule is based on the sewer pipe deterioration model. The sewer pipes/neighborhoods that are predicted to be in the poorest condition in the future are assigned a higher priority on the schedule for on-site CCTV video collection. However, the indicator used to develop the deterioration model (e.g., LOF, in this case) is the output of the off-site CCTV video assessment. By viewing the CCTV footage, the technicians assess the health condition of the sewer pipes and assign a score using an

indicator according to certain specifications (e.g., PACP, in this case); the deterioration model is then developed based on these indicators. Therefore, the proposed deterioration model is a key component of the current sewer pipe maintenance workflow.

This research makes four main contributions to the body of knowledge:

1) This research proposes the first bi-level deterioration model for the sewer pipe system, which helps to improve the maintenance schedule on the neighborhood scale and individual level. The proposed method could serve as a framework for other municipal departments in different cities to develop their own bi-level deterioration model for sewer pipe systems.

2) The average LOF for the neighborhood is predicted, and the results are presented by GIS visually. It is the first prediction model for the future condition of sewer pipes in the neighborhood level compared to the previous models that focus on the individual level deterioration. The results of the neighborhood-level deterioration model show a trend of the sewer pipe deterioration in the neighborhood scale for city managers, which can help them to identify critical neighborhoods that require more attention than others.

3) The development procedure of an individual-level deterioration model provides an option for determining the satisfactory solution for the architecture of the NN model by combining the knowledge from literature and the proposed accuracy-test procedure.

4) The feature selection technique is integrated with the NN model to achieve the objective of using proper input variables to get a great prediction accuracy. The significant influencing variables are identified by the proposed backward variable elimination process, which is further investigated for developing the deterioration curves.

5.6 Conclusions and limitations

A bi-level deterioration model for sewer pipes was developed in this research, namely the neighborhood-level and individual-level deterioration model. The neighborhood-level deterioration model was constructed using the multiple regression analysis, resulting in a final model with an adjusted R^2 of 0.786. Variables of average age, tree density, and traffic coverage were identified as significant influencing variables for the neighborhood-level deterioration model. Furthermore, time series neighborhood LOF can be visualized city-wide in GIS, and with that, city managers could have a clear overview of the sewer pipe condition for each neighborhood, which in turn facilitates creating a city-wide maintenance schedule. For the individual-level deterioration model, an NN classification model and backward variable elimination method were adopted, resulting in a deterioration model with a classification accuracy of 0.955. Age, diameter, length, material, waste type, and category were selected as significant influencing variables for the individual-level deterioration model. Based on that, deterioration curves for different pipes can be developed, which can help to assign different levels of maintenance priorities for pipe in different conditions (e.g., a higher maintenance priority for pipe in a worse condition). Overall, the model could provide city managers with both macro and micro perspectives of sewer pipe conditions. In addition, the proposed model development procedures can be used for sewer pipe networks in other cities, and even used for deterioration model development for other types of infrastructure (e.g., bridges, tunnels, pavement, etc.).

The research has its limitations in terms of data availability and input variable selection. A more comprehensive dataset with more input variables is recommended in order to find the best combination of input variables in future research. The variable elimination process was

conducted in a backward direction, which cannot provide every possible combination of the variables listed in Table 5.8. In future research, more advanced variable selection methods such as variable selection algorithm based on the genetic algorithm or the particle swarm optimization algorithm could be used for finding the best combination of the input variables for the proposed model. In addition, a deterioration model based on physical and chemical principles could be developed in the future to analyze the influence of different variables on the deterioration rates.

Chapter 6: DATA-DRIVEN SEWER PIPE DATA RANDOM GENERATION AND VALIDATION[5]

Preface

Data is the foundation for machine learning-based and data-driven research. In this study, CCTV inspection data for sewer pipe systems serve as the basis for preventive maintenance in the context of sewer pipe condition ratings and maintenance schedule planning. In considering sewer pipe maintenance operations in practical terms, the following CCTV inspection data for sewer pipes are of particular interest to this research: length of the pipes, defect interval, and defect sequence for different types of defects (and taps). However, the data collection process using CCTV inspections is typically expensive and time-consuming from the perspective of the municipal department. In this context, an input modeling technique that aims to exploit the potential value of historical data is proposed by combining the Markov chain model with distribution fitting techniques and other relevant methods. The data generation research presented in this chapter is complementary research to my proposed methodology of the machine learning-based framework for preventive maintenance of sewer pipe systems. The generated data could be used to improve the simulation model proposed in Chapter 2 and the deterioration model presented in Chapter 5. For other researchers, who want to conduct similar studies of preventive maintenance of sewer pipes, encounter data deficiency problems could adopt this proposed framework to generate a larger dataset for their research.

6.1 Introduction and background

Sewer pipe systems are indispensable in terms of the infrastructure required for the development of modern cities due to their function of transporting sewer water and stormwater to treatment plant or water streams, which facilitates to prevent the crisis of water flooding, water-borne disease, etc. (Ana & Bauwens, 2010; Jensen et al., 2016). The importance of the sewer pipe system is also due to the significant amount of inventory and its significant replacement value (Wirahadikusumah et al., 2001). For instance, approximately 6,500 km of all kinds of sewer pipes are installed underground in Edmonton, Canada, for which the replacement value is around 18 billion Canadian dollars (Kurach et al., 2019).

One of the most challenging issues is that the sewer pipe system is deteriorating, the rate of which may be influenced by several factors (e.g., pipe material, soil type) (Balekelayi & Tesfamariam, 2019). In order to maintain the sewer pipe system at a functional level of service (LOS), preventive maintenance is required on a regular basis (Fenner, 2000). In practice, CCTV inspection results serve as the basis for the development of deterioration models, maintenance schedules, etc. (Yin et al., 2019).

In previous studies, simulation models and optimization algorithms are shown to improve the productivity and lower the cost of preventive maintenance, thereby increasing its effectiveness (Agbulos et al., 2006; Navab-Kashani et al., 2013; Zaman et al., 2017). For the development of those models, data collection plays an important role in the process (AL-Alawi et al., 2018). Trybula (1994) finds that data collection usually takes 10–40% of the total time required to build a simulation model. Moreover, some data may be unavailable or cost-prohibitive to obtain in some circumstances (Robertson & Perera, 2002). The interpretation and the understanding of the original data structure and its inherent characteristics are valuable processes because they offer

the potential to maximize the value of the existing data and provide the basis for generating new data that could represent the original data in the case where there is insufficient data. In addition, sewer pipe data (for convenience of expression, the term "sewer pipe data" will be used to refer to CCTV inspection data for sewer pipes) are valuable for sewer pipe preventive maintenance, e.g., sewer pipe condition ratings, maintenance schedule planning, etc. In this context, this research aims at randomly generating sewer pipe data by exploring the inherent randomness of the original data. Then, the randomly generated data will go through a rigorous validation process to prove that the similarity between the original dataset and the generated dataset is acceptable.

The data generation process, which is also known as input modeling, is usually conducted by first selecting a probability distribution to mimic the real-life behavior of the input variables (Abourizk et al., 2016). An accurate input modeling process serves as the foundation for an accurate simulation model (Wu et al., 2020). Input modeling techniques have been widely investigated in previous studies; for example, Abourizk et al. (1994) investigate fitting construction operations as a beta distribution based on sample data; Zouaoui & Wilson (2004) propose a Bayesian Model Averaging (BMA)-based approach to deal with the input model uncertainties; and Wu et al. (2020) develop a method for real-time updating of the input by Bayesian inference combined with Markov chain Monte Carlo (MCMC). However, when the input data structure becomes more complicated, further operations need to be undertaken to construct a reasonable input model. For example, for the generation of industrial pipeline data, AL-Alawi et al. (2018) propose a Markov chain model to generate the industrial pipeline structure with different components randomly (e.g., tube, elbow, tee, etc.) each having different locations within the pipe. After data generation, a rigorous validation process is conducted to

validate the quality of the generated pipeline data. The industrial pipeline data generation research provides inspiration for the present research, and the framework proposed by Al-Alawi et al. (2018) is adopted for use in this research to solve the specific problem of sewer pipe system data generation. However, the structure of sewer pipe data is significantly different than industrial pipeline data since the industrial pipeline data contains thirteen types of standard elements without any distance-related variables, while the sewer pipe data include variables of seven types of defects, defect interval, and total pipe length (which will be described in Section 6.2). The differences in the data structure between these two types of targeted datasets make it necessary to employ a different data generation method and validation process. To be more specific, the distribution fitting and Markov chain model are integrated into the data generation method, and X-means clustering and Hausdorff distance are adopted for the purpose of validation, which will be described in detail in Sections 6.4 and 6.5.

To summarize, several conclusions can be made based on the above analysis: 1) sewer pipe data are valuable for the practice of sewer pipe preventative maintenance; 2) input modeling techniques are widely used in the construction industry, especially in the context of building simulation models; 3) among the input modeling techniques, none of them were found to target the generation of sewer pipe data, which is considered graphical data (or tree type data) that is more complicated than a purely numerical type of data. Based on the above reasoning, the present research is motivated to present an input modeling framework that can randomly generate sewer pipe data that maintains the major characteristics of the original sewer pipe data collected in the field. The major contribution of this research is to propose a novel input modeling framework to randomly generate sewer pipe data. Through the process, the characteristics of the sewer data are well investigated and understood. A rigorous validation

framework to compare the similarity of original data and the generated data is described and adopted, which could be employed in other research for the purpose of dataset similarity comparison.

6.2 Sewer pipe dataset

The sewer pipe dataset used in this research is collected by EPCOR Water Services Inc., the industry partner located in Edmonton, Canada. Figure 6.1 shows the overall data collection process.
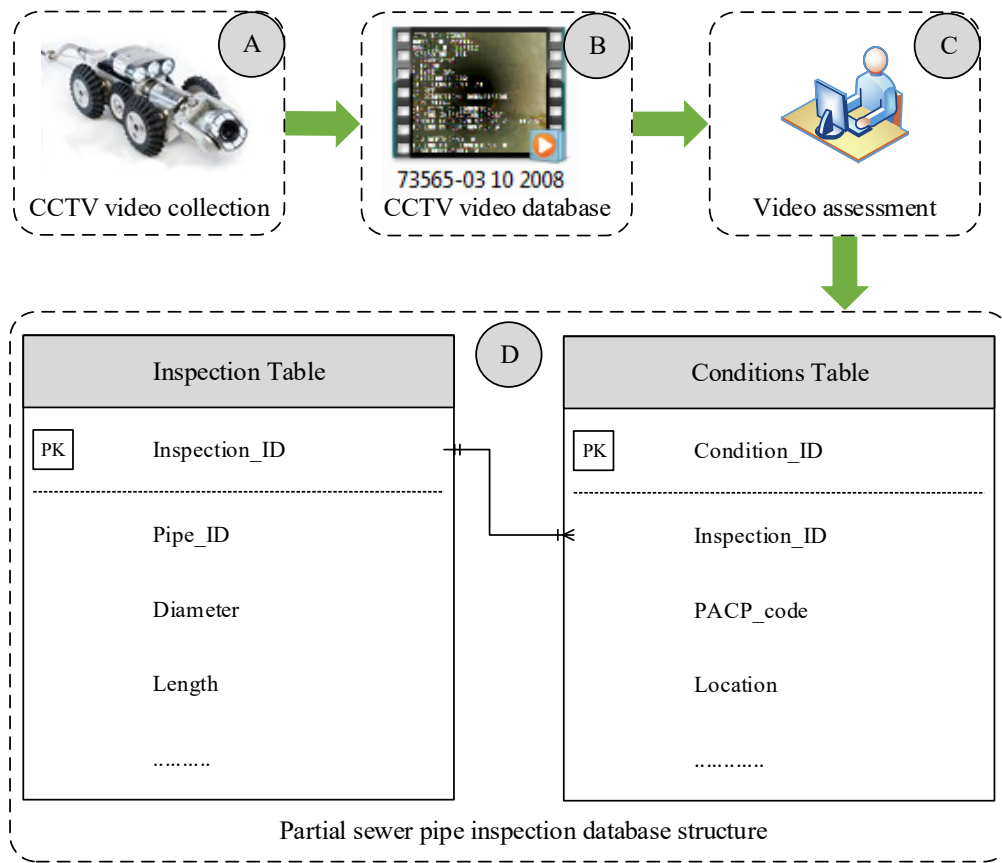

Figure 6.1: Data collection process.

The first step is to collect CCTV videos of sewer pipes recorded on-site by a crawler equipped with a camera and lights (part A of Figure 6.1). The output of this step is the CCTV footage shown in part B of Figure 6.1. Then the video assessment is conducted as described in section

1.1.2. Six types of defects (i.e., broken, hole, deposits, crack, fracture, root) and one type of construction feature (i.e., tap) are recorded as shown in Figure 6.2 (Yin et al., 2019). These defects and construction features are of interest in this research considering the real-life maintenance operations conducted by the industry partner. It should be noted that, since only one construction feature (tap) is included in this research, we will consider a tap as representative of any construction feature for the convenience of expression in the present study.

(a). Broken   (b). Hole   (c). Deposits

(d). Crack   (e). Fracture   (f). Root   (g). Tap
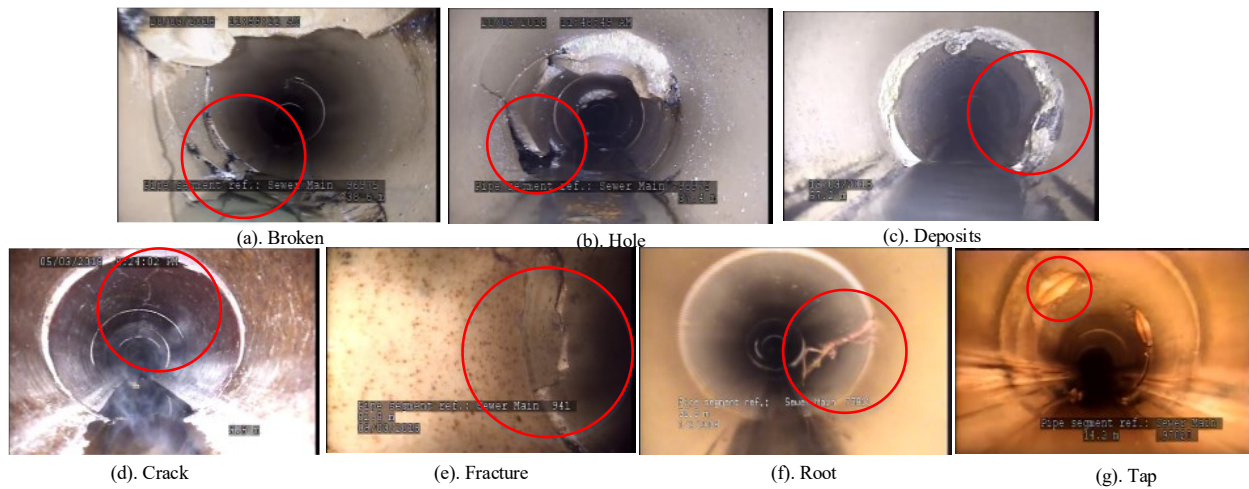
Figure 6.2: Examples of defects and tap.

In this research study, the PACP coding method is adopted to be consistent with the dataset we collected. In the PACP coding manual, the crack defect includes several subclasses: CL (longitudinal), CC (circumferential), CM (multiple), CS (spiral), CH (hinge). In this research study, we will merge these subclasses into one defect type, since the subclasses would be too complicated to include and will be considered in future research. For instance, in the example mentioned above, we consider CL, CC, CM, CS, and CH as crack. After the video assessment process, the output is the sewer pipe inspection database where the inspection information of the sewer pipe is recorded.

The database structure is partially presented in part D of Figure 6.1. The dataset is retrieved from the *Inspection Table* and *Conditions Table* in the database. The *Inspection Table* records the

properties of the sewer pipe, which include pipe ID, diameter, length, material, etc. The *Conditions Table* records the CCTV video assessment results, which include the PACP code for the defects and taps, location for these defects and taps, etc. The *Inspection Table* and *Conditions Table* have primary keys of Inpsection_ID and Condition_ID, respectively, and the two tables are connected with Inpsection_ID, which is the foreign key of *Conditions Table.*

A database query manipulation is conducted to translate the database into the type of real-life information presented in Figure 6.3, which shows an example of a sewer pipe with inspection results. The sewer pipe is shown in Figure 6.3 starts and ends at manholes, with a total pipe length of 12 m. Along with the inspection direction, different types of defects and taps are marked with names and the corresponding distance to the starting point (manhole). For example, at a distance of 2.9 m, there is a tap. The defect interval is the distance between two adjacent defects; for instance, the tap and crack have a defect interval of 2.3 m. Figure 6.3 shows an example sewer pipe from the randomly generated dataset with locations of defects and taps, as indicated in the figure. To summarize, three elements are needed to represent one sewer pipe data: total pipe length, defect interval, and defect sequence. As mentioned above, the generated data will serve as valuable input for simulation models that may be developed for various purposes. The following sections will describe the data generation process in further detail.
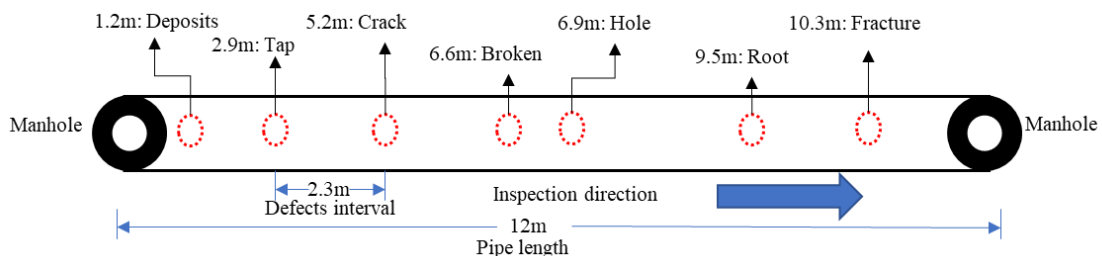

Figure 6.3: An example of a sewer pipe.

The original dataset contains 1,793 sewer pipes whose total length is around 101 km. For those pipes, there are 17,863 records that include taps and six types of defects. The distribution of the

original dataset for defects and taps is presented in Figure 6.4, and shows that the type of deposits represents 45% of the total records in the dataset, while taps (26.83% of the total records) represent the second-largest number the original dataset. In the meantime, the broken and hole types of defects are the least frequent, with 121 (0.68%) and 52 (0.29%) occurrences, respectively. Broken and hole are the two types of defects that are considered major defects, which means that any pipe with one of these two types of defects needs immediate maintenance operation.
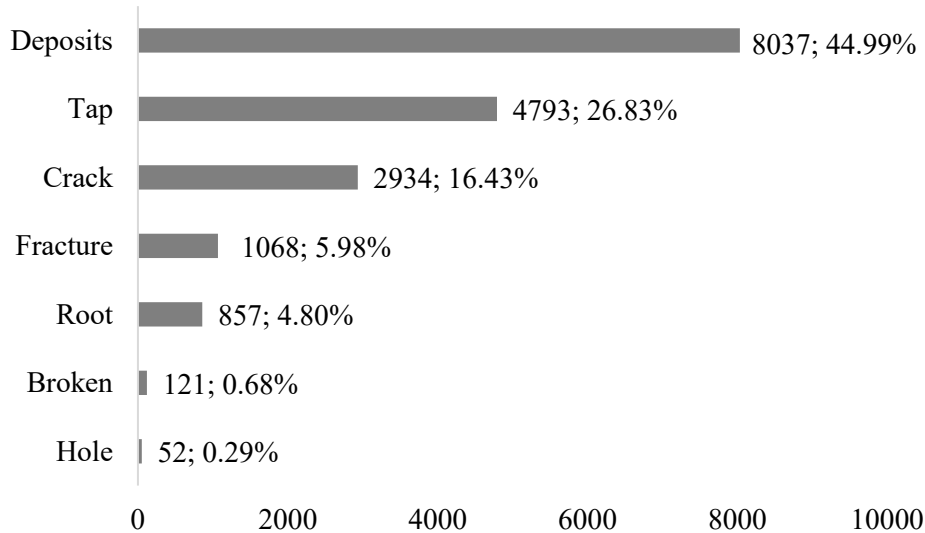
Figure 6.4: The number of each type of defect in the original dataset.

Further analysis is conducted to reveal the correlation among different types of defects by calculating the correlation matrix and converting it to a correlation heatmap (see Figure 6.5). The correlation analysis was conducted on the whole original dataset (1,793 sewer pipes) to reveal the correlation between the number of defects (and taps) in one sewer pipe. The significance level is 0.01, which means that the p-value for the correlation coefficient in Figure 6.5 is less than 0.01. The objectives of the correction analysis are twofold: 1) to investigate the characteristics of the original dataset, which helps to understand the structure of the dataset; and

2) to serve as a benchmark for comparing the generated dataset to the original dataset. All the correlation coefficients are positive and the highest correlation is between fracture and broken. The second-highest correlation is between fracture and crack. The correlation heatmap of the original dataset is going to be compared with the correlation heatmap for the generated dataset in Section 6.5.1 to examine the similarity between the original dataset and the generated dataset.
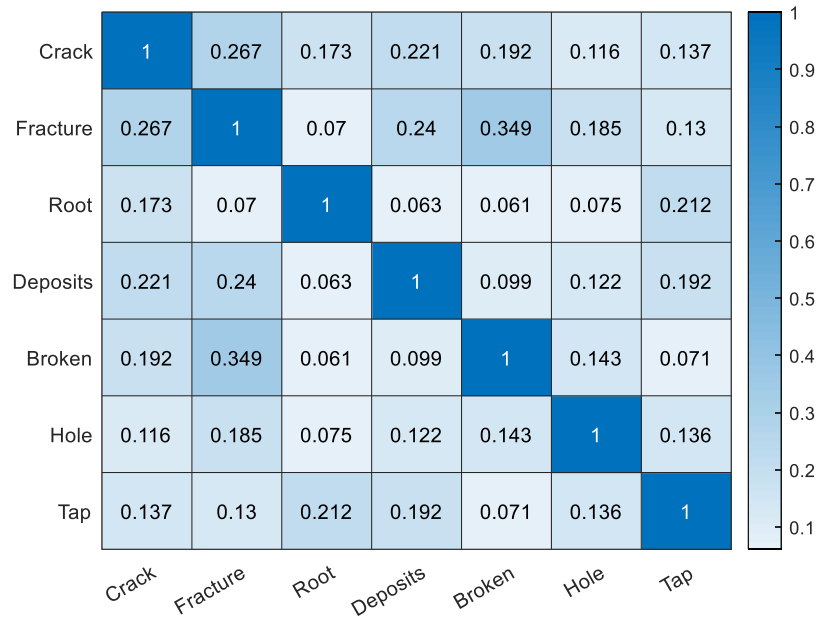
Figure 6.5: Correlation heatmap of the original dataset.

6.3 Methodology

The overall methodology, as presented in Figure 6.6, starts with data collection and statistical analysis for the original dataset. These two parts of the process have been described in Section 6.2. As described further in Section 6.4, the subsequent step is the proposed data generation model development, based on which the sewer pipe data can be generated. Then, statistical analysis will be conducted on the generated dataset to give the total number of taps and the total number of each type of defect, and the correlation coefficient between them, etc. The statistics of the generated dataset will be compared with the original dataset to determine the similarity

between these two datasets. While the statistical analysis and comparison can give an overview of the similarity of the defects and taps from the perspective of quantity, their location can not be measured by this method. In order to measure the similarity between the original dataset and the generated dataset from the perspective of the location of defects and taps and their distribution within the sewer pipe (an example can be seen in Figure 6.3), the feature vector comparison method is adopted in this research. The feature vector comparison method is proposed by Liu et al. (2006) to compare the similarities of DNA sequences, and the method is also adopted by AL-Alawi et al. (2018) to compare the generated industrial pipeline data with the real industrial pipeline data. The feature vector comparison method is enhanced and adjusted to compare the original sewer pipe dataset and generated sewer pipe dataset with the X-means clustering method and Hausdorff distance calculation. The validation process is presented in detail in Section 6.5. Finally, a conclusion section will summarize the whole research and propose some potential future works.
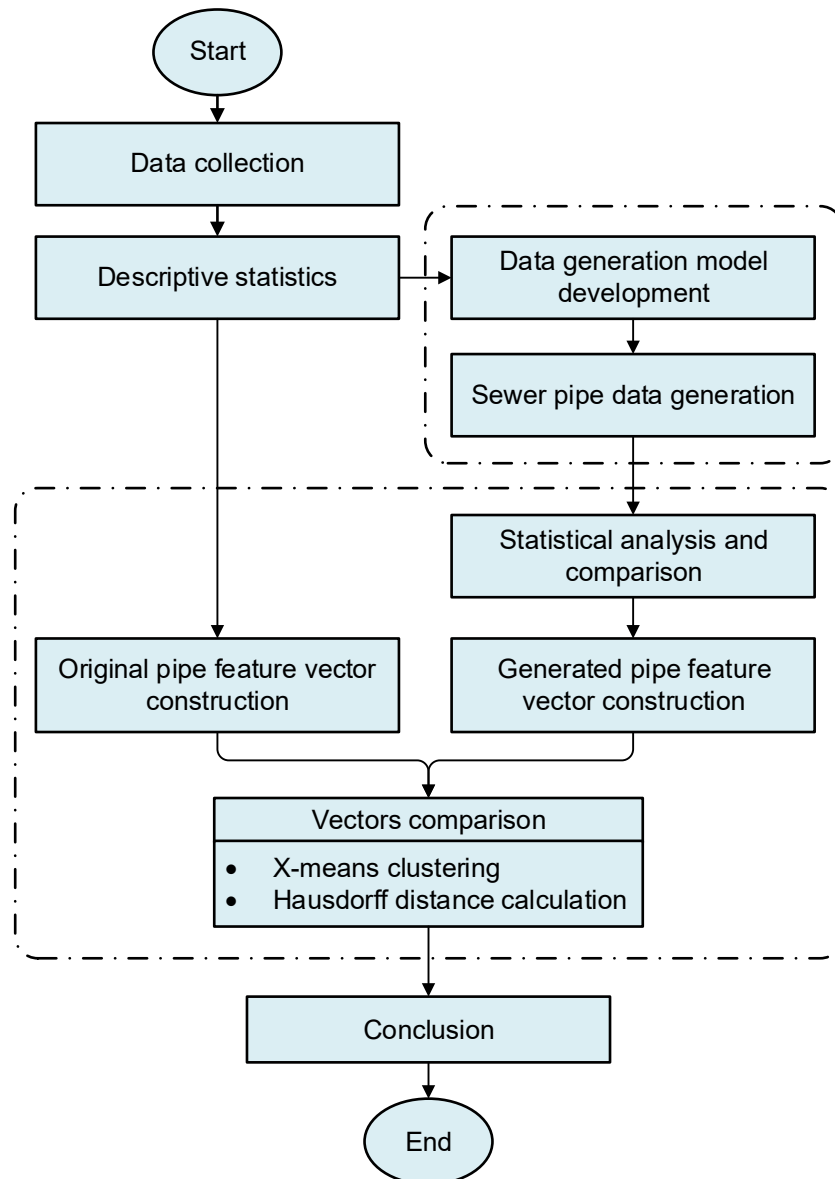
Figure 6.6: The methodology of the research.

6.4 Sewer pipe data generation

6.4.1 Overview of sewer pipe data generation process

The sewer pipe data generation process is presented in Figure 6.7. As mentioned in Section 6.2, three elements are needed to generate one sewer pipe dataset, including the total pipe length, defect interval, and defect sequence. Therefore, the process starts with developing models to generate these three elements (part A of Figure 6.7). The variables of pipe length and defect

interval are generated with conventional distribution fitting techniques since these variables are simple univariant, and this technique is widely used in previous research (S. M. Abourizk et al., 1994; Lu, 2002). The defect sequence is generated with the Markov chain model, which will be described in the following sections. The samples of pipe length, defect interval, and defect sequence can be generated by the model developed in the first stage (part A). Then, the sewer pipe data integration process may begin, which is indicated in part C of the flow chart shown in Figure 6.7. The integration process (part C) is undertaken to integrate the data samples (part B) generated from the first stage (part A) together to generate the sewer pipes. The integration process starts with a sample pipe length, L, and three initial counting variables (i.e. M, N and i) set to 0. Then, from the generated samples of defect interval, a defect interval is sampled as m_i, and a defect is sampled from the generated defect sequence D_i. Subsequently, the distance between the starting point and the defect D_i is accumulated in the variable M. If $M < L$ it means that the defect D_i is within the pipe, and another iteration that samples another defect can begin. Otherwise, if $M > L$, the defect D_i is located in the next sewer pipe. The defect interval m_i will be used as the first defect interval (the distance from starting manhole to the defect D_i) in the next sewer pipe. Finally, if a certain number of sewer pipes are generated (e.g., 1,793 in this study), the process can be terminated. In this study, the whole process will output 1,793 sewer pipes with human-assigned defects with their specific locations (as illustrated in Figure 6.3).
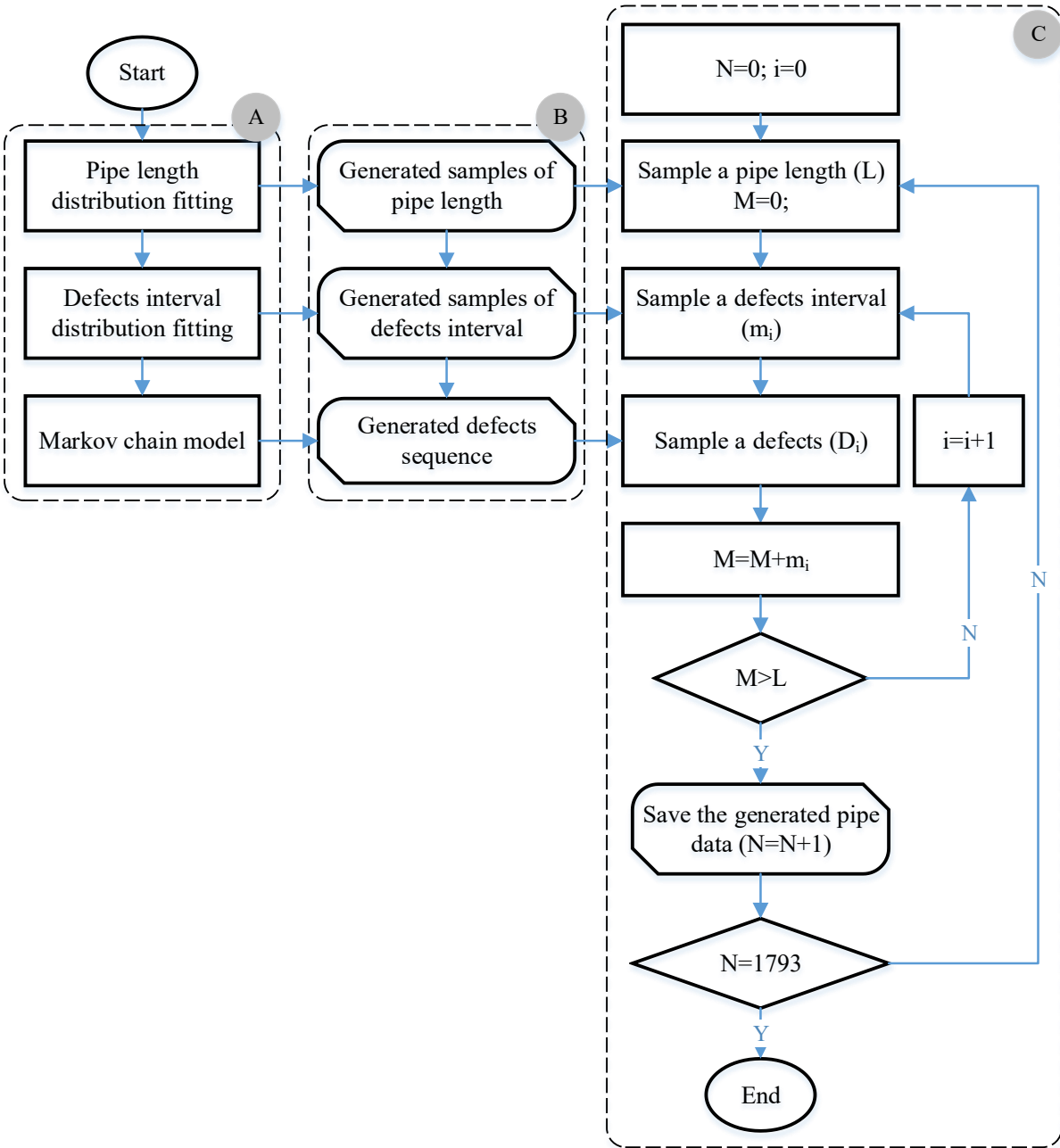
Figure 6.7: The process of sewer pipe data generation.

6.4.2 Markov chain model

The Markov chain model is typically used to describe a stochastic process wherein the probability of jumping into a specific state at a future step only depends on the state of the current step (Micevski et al., 2002). The mathematical format can be presented, as shown in Equation (6-1) (Baik et al., 2006).

$$P(X_{t+1} = i_{t+1} \mid X_t = i_t, X_{t-1} = i_{t-1}, ..., X_1 = i_1) = P(X_{t+1} = i_{t+1} \mid X_t = i_t) \tag{6-1}$$

where the probability of variable X in state i_{t+1} at step $t+1$ only conditionally depends on its state i_t in the step t. For a Markov chain model, the assumption is that the conditional probability remains the same over time. For any variable X in state i at step t, the probability of transfer to state j at step $t+1$ can be expressed as shown in Equation (6-2).

$$P(X_{t+1} = i \mid X_t = j) = p_{ij} \tag{6-2}$$

where p_{ij} represent the transfer probability. If multiple states (n states, where $n > 1$) exist in the target Markov chain process, a transfer matrix is usually required as shown in Equation (6-3).

$$P = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{pmatrix} \tag{6-3}$$

If we define the initial state as a vector $Q^{(0)}$, after n steps, the state vector will be $Q^{(n)} = Q^{(0)} P^n = Q^{(n-1)} P$.

The Markov chain model has been widely used in various engineering domains (Lecchini-Visintini et al., 2010; Yang et al., 2005). As for the field of sewer pipe maintenance, the Markov chain model is widely used to develop the deterioration model, where the model can mimic the deterioration of the condition of the sewer pipe over time (Baik et al., 2006; Micevski et al., 2002; Tran et al., 2008). In previous studies, conclusions show that the Markov chain model can simulate real-life stochastic processes in an efficient manner. For example, studies using the Markov chain model to predict daily rainfall (Haan et al., 1976), and to reposition ambulances (Alanis et al., 2013) show good performance. Essentially, the generation of sewer pipe defects data is a discrete-time stochastic process, which is a problem that could be solved efficiently by

the Markov chain model. The present research employs the Markov chain model to generate the sewer pipe defect data taking the following considerations into account: 1) the characteristics of the sewer pipe defect generation problem itself, which is a discrete-time stochastic process that could be mimicked by using the Markov chain model, and 2) the performance of the Markov chain model is then validated in the present study, and the results show that the generated dataset is highly similar to the original dataset, which means that the Markov chain model solved this problem properly. For the problem described in this research, we propose that the sequence of the defects can be generated by the Markov chain model. As the CCTV crawler advances through the sewer pipe in a certain direction, the $n+1$ defect (or tap) can be predicted by the n defects (or taps) using the Markov chain model. For example, if a crack is found in the CCTV video of the sewer pipe as the first defect, then along the inspection direction, with a certain interval, the next defect (or tap) could be any one of the seven components (i.e., crack, fracture, root, deposits, broken, hole, tap) according to the specific probability of each. In addition, besides the sequence of the defects (and taps) for the sewer pipe, two other variables must be generated in the sewer pipe data: the pipe length and defect interval distance. The pipe length and defect (and tap) interval distance can be generated using conventional input modeling techniques, which is to fit the original data with proper distributions, as mentioned in Section 6.4.1. The defect (and tap) sequence generation modeled by the Markov chain, the generation of pipe length data, and the generation of defect (tap) interval distance are described in detail in the next section.

6.4.3 Model results

The variables of pipe length and defect interval are retrieved from the original dataset, which provides the basis for the distribution fitting process. Figure 6.8 shows the histogram of the pipe

length variable in the original dataset and the fitted beta distribution. The original dataset has a pipe length range of 0–916 m, which is assumed to be the pipe length range for the generated pipe length since the latter is sampled from the distribution of the former. Note that all the distribution fittings are validated by chi-square goodness of fit under the significance level of 0.05. As for the defect interval variable, the original dataset was separated into two parts to fit two distributions since, in a prior experiment, it was found that to fit the data with one distribution results in unsatisfactory performance in terms of representing the original data. In the original dataset, 91% of the defect intervals are less than 30 m, which inspires a separation of the defect intervals into two parts: shorter than 30 m, and longer than 30 m. The distribution fittings for these two parts of the dataset, which are presented in Figure 6.9, are two beta distributions with the distribution parameters shown in the figure. When sampling the defect interval data in the data generation process, 91% of the defect intervals will be randomly generated following the first distribution shown in Figure 6.9, while the rest of the defect intervals will be randomly generated following the second distribution shown in Figure 6.9.
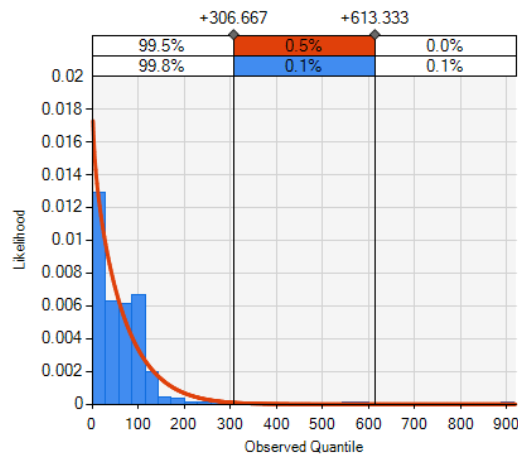
Figure 6.8: Distribution fitting of pipe length (Beta: $\alpha=0.929$, $\beta=12.831$, min=0, max=916).
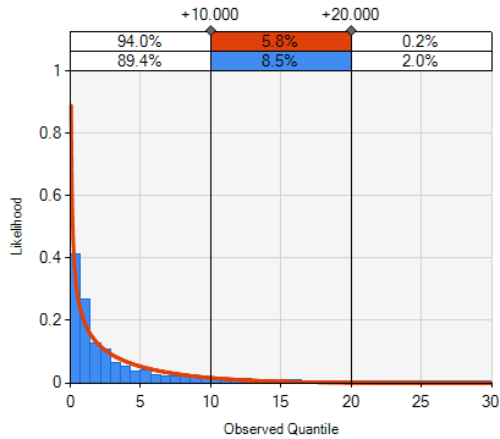
Defect interval (Shorter than 30m, Beta: α=0.543, β=4.832, min=0, max=30)
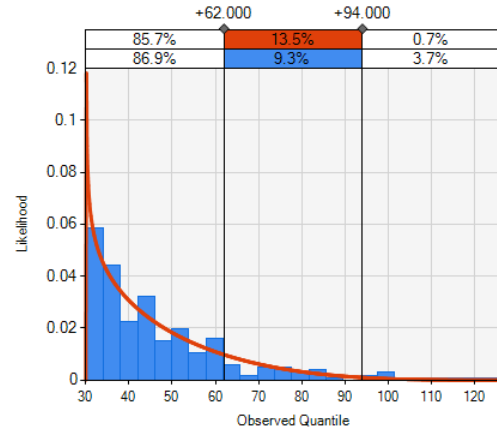
Defect interval (Longer than 30m, Beta: α=0.809, β=4.082, min=30.1, max=125.2)

Figure 6.9: Distribution of defect interval.

The generation of defect sequence requires an input of the initial state, which can be derived from the original dataset (see Figure 6.4) as the initial vector:

$$Q^{(0)} = (16.43\%(crack), 5.98\%(fracture), 4.80\%(root), 44.99\%(deposits), 0.68\%(broken), 0.29\%(hole), 26.83\%(tap))$$

The result of the transfer matrix is presented below, which is the transfer matrix P shown in Equation (6-3). The rows of the matrix are labeled accordingly and the columns are arranged in the same sequence. For example, the elements of (1,1) (1,2) are 46.49% and 5.76%, respectively, which means that if the current defect is crack, there is a probability of 46.49% that the next defect would be a crack, and a probability of 5.76% that the next defect would be a fracture, and so on. Each of the numbers in the matrix is calculated based on the historical dataset (original dataset). For example, to calculate the first row in the transfer matrix P, all the cracks need to be picked out in the original dataset and then the frequency of the next defect subsequent to the picked crack is determined separately for each of the different types of defects. It was determined

that 46.49% of the defects subsequent to a crack are still cracks, and 5.76% of the defects

subsequent to a crack are fractures, and so on.

$$P = \begin{array}{r} crack \\ fracture \\ root \\ deposits \\ broken \\ hole \\ tap \end{array} \begin{pmatrix} 46.49\% & 5.76\% & 3.51\% & 27.88\% & 0.58\% & 0.17\% & 15.61\% \\ 16.76\% & 29.68\% & 3.09\% & 27.25\% & 2.15\% & 0.56\% & 20.51\% \\ 11.09\% & 3.15\% & 42.82\% & 17.74\% & 0.82\% & 0.70\% & 23.69\% \\ 10.03\% & 3.70\% & 1.53\% & 66.58\% & 0.50\% & 0.19\% & 17.48\% \\ 21.49\% & 12.40\% & 4.96\% & 28.10\% & 5.79\% & 3.31\% & 23.97\% \\ 13.46\% & 15.38\% & 11.54\% & 26.92\% & 3.85\% & 7.69\% & 21.15\% \\ 9.53\% & 4.90\% & 4.57\% & 28.75\% & 0.52\% & 0.25\% & 51.47\% \end{pmatrix}$$

Therefore, after n steps, the defect states can be calculated as: $Q^{(n)} = Q^{(0)} P^n = Q^{(n-1)} P$. Following

this transfer matrix, the sequence of defects can be generated. With the three generated elements

(i.e., pipe length, defect interval, and defect sequence), the sewer pipe data can be generated

following the process shown in part C of Figure 6.7. For the purpose of validating the method,

the comparison of the generated dataset and the original dataset is described in Section 6.5.

6.5 Model validation

6.5.1 Statistical analysis of the generated data

Figure 6.10 shows comparison of the number of each type of defect in the generated dataset and

the original dataset. It can be seen from the figures (Figure 6.10) that the structure of the

generated dataset is quite similar to that of the original. The total number of defects (including

taps) in the original dataset is 17,862, while the number is 18,014 in the generated dataset. The

most frequent defect is deposits, while the least frequent defect is hole in both the original

dataset and the generated dataset. The ranking of the other defects in terms of the total number is

also the same between the two datasets. As for the percentages of each type of defect, the two

datasets are quite similar in this regard. For instance, the root defect represents 4.80% of the total

defects (including taps) in the original dataset, while the same defect represents 4.33% of the total defects (including taps) in the generated dataset.
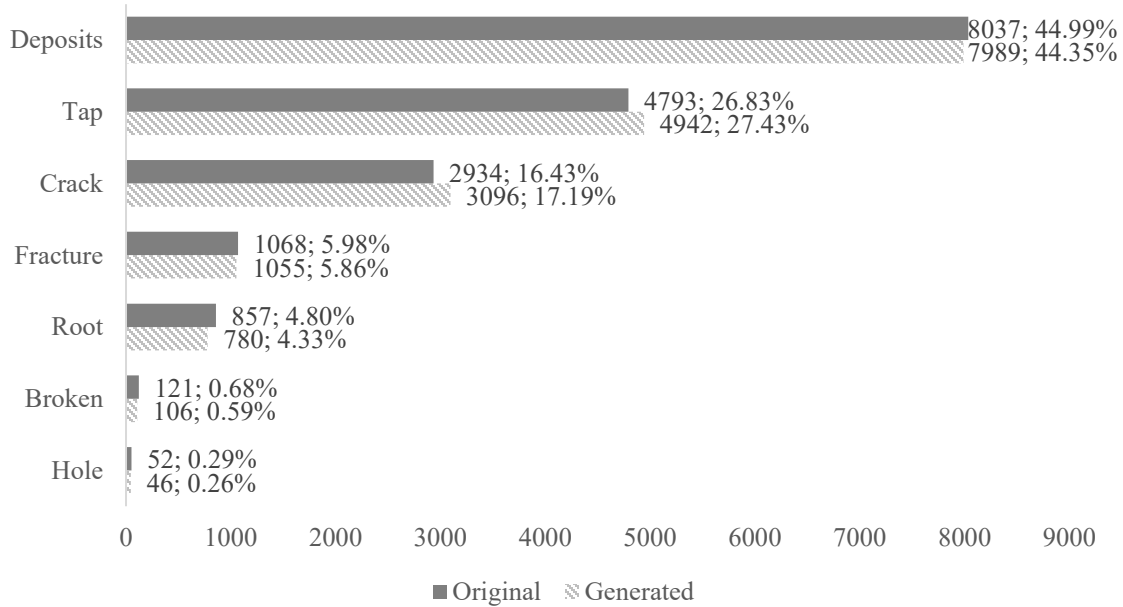
Figure 6.10: The number of each type of defect in the generated dataset.

In order to determine the pattern of the data structure in terms of the number of each type of defect in the original dataset and in the generated dataset, several statistics are calculated and summarized in Table 6.1. In the original dataset, the number of cracks within one sewer pipe ranges between 0–56, while this range for the generated dataset is 0–25, which indicates some variance between the two datasets. The standard deviation for the original dataset is 4.49, while the standard deviation for the generated dataset is 2.49. However, the average number of cracks for each sewer pipe is 1.64 for the original dataset and 1.73 for the generated dataset, which is quite similar statistically. The situation is similar in the case of fracture, root, and deposits: the range for each type of defect varies when comparing one dataset to the other, but the average number of defects for each type of defect for each sewer pipe is relatively similar for both

datasets. As for the taps and the broken and hole defects, all the statistics (i.e., min, max, mean, and standard deviation) in Table 6.1 show a high degree of similarity.

Table 6.1: Statistical analysis of the original dataset and the generated dataset for each type of defect in one sewer pipe.

	Original						
	Crack	Fracture	Root	Deposits	Broken	Hole	Tap
Min	0	0	0	0	0	0	0
Max	56	48	39	62	5	3	33
Mean	1.64	0.60	0.48	4.48	0.07	0.03	2.67
Standard deviation	4.49	2.22	2.35	7.17	0.32	0.20	4.40
	Generated						
Min	0	0	0	0	0	0	0
Max	25	10	8	37	3	2	31
Mean	1.73	0.59	0.44	4.46	0.06	0.03	2.76
Standard deviation	2.49	1.15	1.02	5.18	0.27	0.17	3.53

To compare the number of each type of defect (including taps) for one sewer pipe, seven histograms are plotted in terms of the taps and the six types of defects for the two datasets (see Figure 6.11). The greater the overlapping area between the two histograms, the higher the similarity between the two datasets. As can be seen in Figure 6.11, the overlapping area represents the majority of the area of the histogram for every pair, especially for the broken and hole defects. The reason for the large overlapping area for the broken and hole defects is due to the narrow range of frequency for these two types of defects. The other five pairs of histograms show a similar pattern: for the bin that is closest to zero, the frequency of the original dataset is slightly larger than that of the generated dataset. And for the rest of the bins, the situation is the opposite. The phenomenon illustrates that the numbers of cracks, fractures, roots, deposits, and taps for the generated dataset are slightly more evenly distributed than the original dataset. However, the difference is small compared with the overlapping area of the histograms.
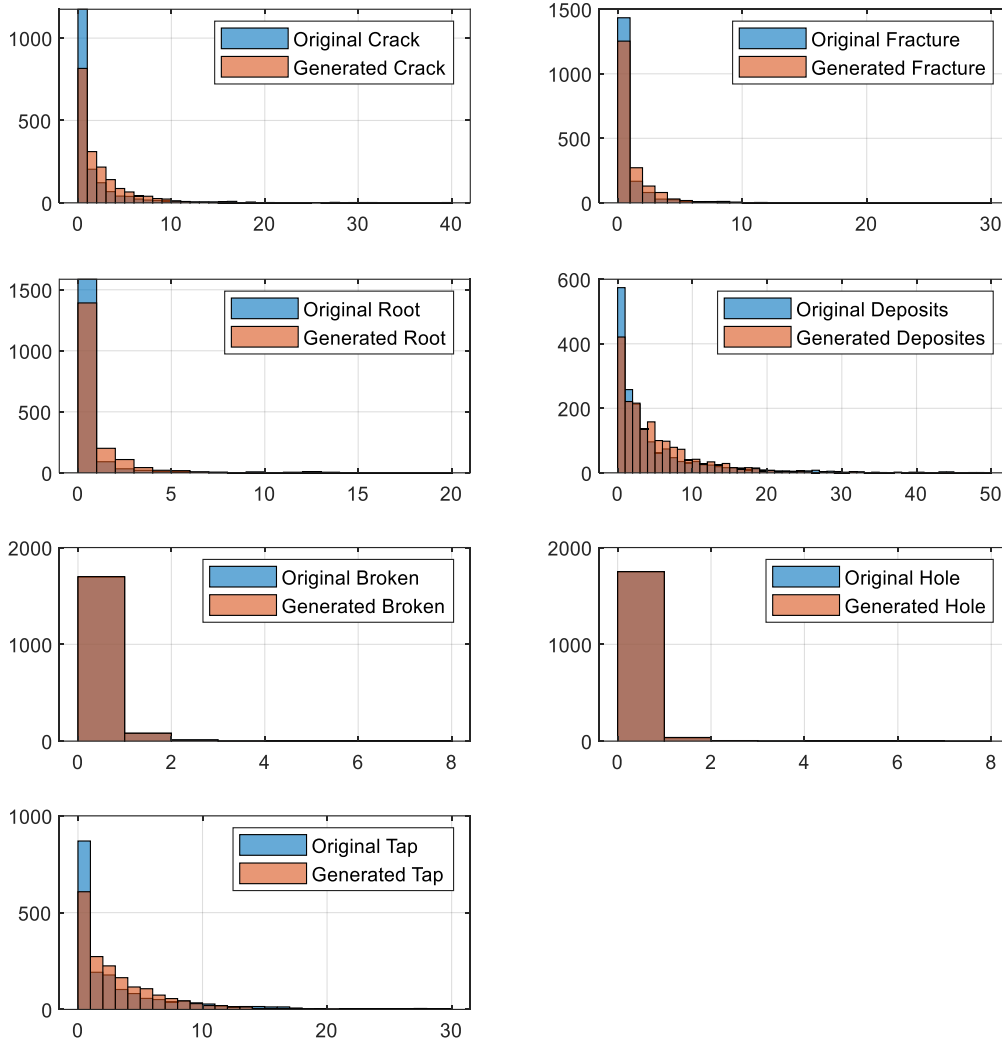
Figure 6.11: Histogram overlap comparison between the original dataset and generated the dataset for each type of defect.

Finally, the correlation coefficients are calculated for the generated dataset, and a correlation heatmap (see Figure 6.12) is plotted to compare with Figure 6.5. Overall, the correlation coefficients are higher than those found for the original dataset. The highest correlation coefficient is 0.616 between deposits and tap, while the correlation coefficient is 0.349 between fracture and broken. The lowest correlation coefficient is 0.096 between root and hole, while the correlation coefficient is 0.061 between root and broken. However, the trend is similar between the two figures; for example, both crack and fracture, and root and tap have a medium to a higher

degree of correlation for the two datasets. To summarize, in terms of statistical analysis, the generated dataset is reasonable and has been determined to have a high degree of similarity with the original dataset.
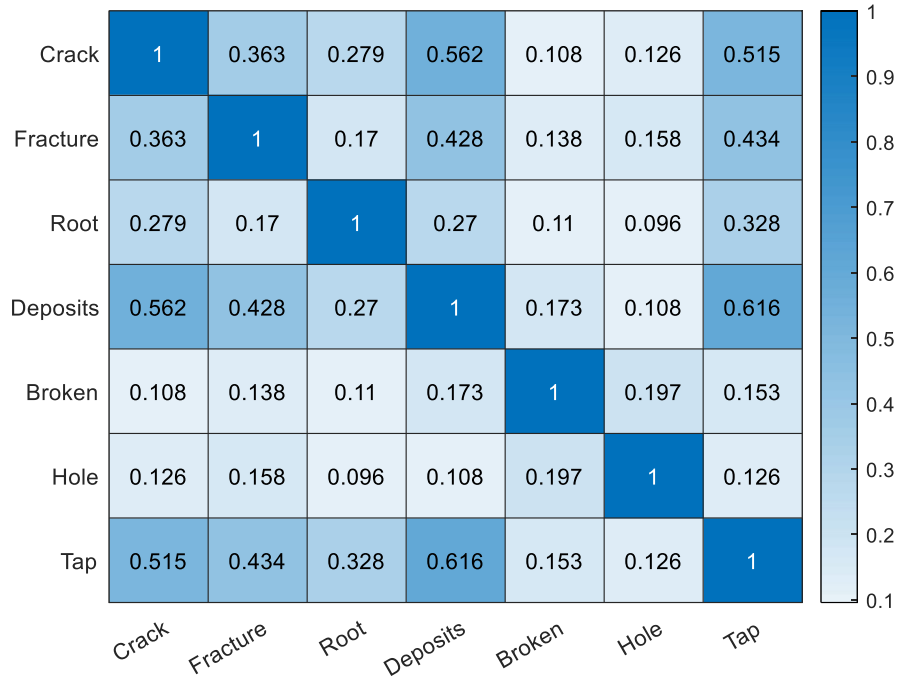
Figure 6.12: Correlation heatmap of the generated dataset.

6.5.2 Sewer pipe feature vectors

The similarity between the original dataset and the generated dataset regarding the number of each type of defects can be investigated in terms of the statistical analysis conducted in Section 6.5.1. However, the statistical analysis can not measure the similarity of the datasets in terms of the location of the defect. For example, two pipes with the same length that both have two cracks and one fracture are presented in Figure 6.13. From a statistical point of view, the two pipes are quite similar, since the total length, defect interval, and the number of each type of defect are all equal. But the two pipes are definitely different pipes visually speaking since the relative location of the defects is different. Therefore, a sewer pipe feature vector is proposed to represent

the sewer pipe structure. The vector method is originally proposed by Liu et al. (2006) to measure the DNA sequence similarity as mentioned above.
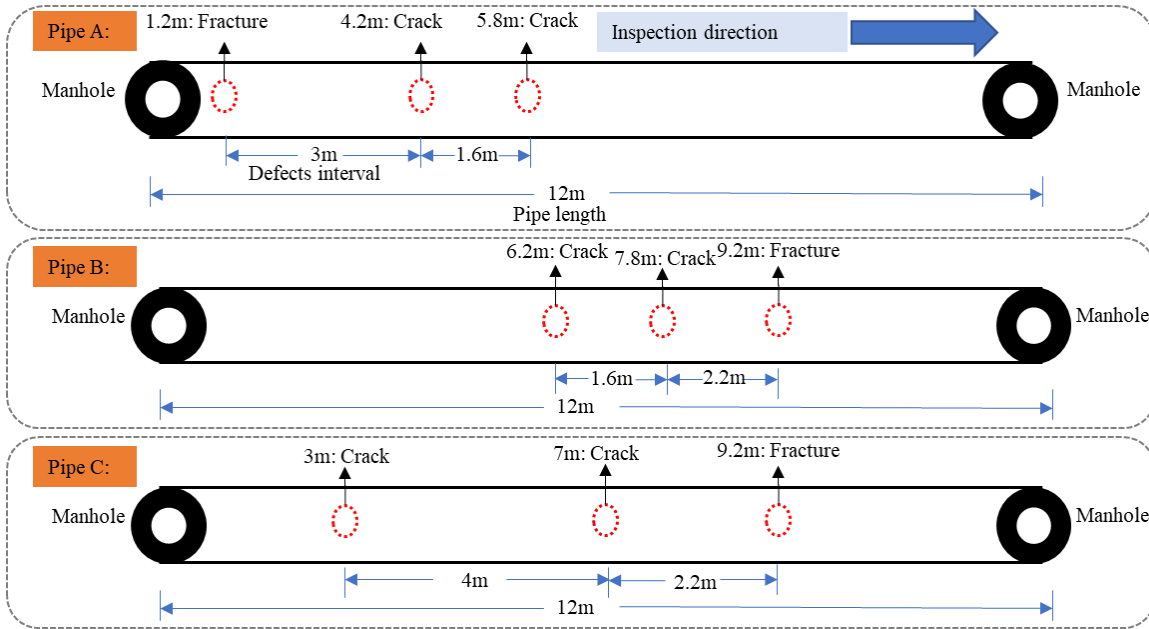
Figure 6.13: Example of two pipes with the same defects.

Before constructing the sewer pipe feature vector, several variables are introduced, as shown in Equations (6-4)-(6-6).

$$T_i = \sum_{j=1}^{N_i} t_j \qquad (6\text{-}4)$$

$$\mu_i = \frac{T_i}{N_i} \qquad (6\text{-}5)$$

$$D_i = \sum_{j=1}^{N_i} \frac{(t_j - \mu_i)^2}{N_i} \qquad (6\text{-}6)$$

where

- i represents different types of defects (tap) (i.e., crack, fracture, root, deposits, broken, hole, tap)

- N_i is the total number of defects (taps) i for the sewer pipe

- t_j is the distance to the starting point (manhole) of j th defect (tap) i

- T_i is the total distance for each type of defect (tap)

199

- μ_i is the average distance of each type of defect (tap)

- D_i is created to represent the distribution of each type of defect

Table 6.2: N_{crack}, T_{crack}, D_{crack} of pipe A, B and C.

Pipe	N_{crack}	T_{crack}	D_{crack}
A	2	10	0.64
B	2	14	0.64
C	2	10	4

To illustrate the three equations above, the variables of N_{crack}, T_{crack}, and D_{crack} are calculated and summarized in Table 6.2. For pipe A, N_{crack} is equal to 2, T_{crack} is equal to 10 (4.2 + 5.8), μ_{crack} is equal to 5 (10 ÷ 2), and D_{crack} is equal to 0.64 ($\frac{(4.2-5)^2+(5.8-5)^2}{2}$). N_{crack}, T_{crack}, and D_{crack} for pipes B and C can be calculated similarly as well. From Figure 6.13, the similarities and differences among three pipes can be clearly observed; however, in order to show these similarities and differences, the combination of these three indices needs to be considered simultaneously. All the pipes have 2 cracks, so $N_{crack} = 2$, which means it is not possible to distinguish these three pipes using this single index alone. However, $N_{fracture} = 1$ holds for pipes A, B, and C, which helps to show the difference between crack and fracture. The relative location (defect interval) for the two cracks in both pipes A and B is the same (1.6 m), which leads to the distribution index of $D_{crack} = 0.64$, which holds for both pipes. Therefore, we also cannot distinguish pipes A and B using the crack distribution index. The total distances for cracks (T_{crack}) in pipes A and C are the same (10), thus, we cannot distinguish these two pipes with this single index. Therefore, the combination of these three indices could show the major characteristics of a pipe.

Considering the variables in Equations (4–6) and other properties of the sewer pipe, a 22-dimensional feature vector structure is proposed to represent the sewer pipe (see Figure 6.14). The first element L represents the length of the sewer pipe. Block 1 represents the three variables (N_{crack}, T_{crack}, D_{crack}) for crack, block 2 represents the three variables ($N_{fracture}$, $T_{fracture}$, $D_{fracture}$) for fracture, and so on. Therefore, 21 elements constitute one sewer pipe feature vector to represent the sewer pipe in terms of length of the pipe, number of each type of defect (tap), and location of each type of defect (taps). Both the original dataset and the generated dataset are converted to sewer pipe feature vector format for comparison purposes. A comparison between the two datasets is conducted by comparing the feature vectors of the two datasets. The feature vector represents the major characteristics of a sewer pipe, and the similarities and differences between the feature vectors of the pipes can represent the similarities and differences between sewer pipes. The comparison process is described in the following two sections.
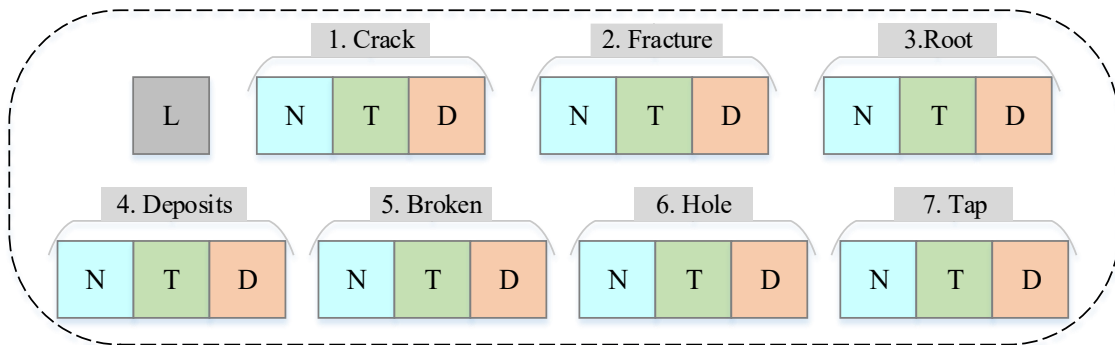
Figure 6.14: Sewer pipe feature vector structure.

6.5.2.1 Cluster analysis and comparison

Cluster analysis, which is an unsupervised machine learning technique, is one of the most useful tools for understanding and interpreting the data structure and other information inherent to the dataset (Shen et al., 2019; Wagstaff et al., 2001). Cluster analysis has been widely used in the construction management field, for instance, to classify construction contractors (Chinyio et al.,

1998; Holt, 1996). A distance-based clustering algorithm, X-means clustering (Pelleg & Moore, 2015), is adopted in this research for its many benefits compared to the conventional *k*-means clustering method. The advantages of X-means clustering are three-fold: 1) the speed is faster than repeatedly using k-means for different values of *k*; 2) the number of clusters *k* is not required; and 3) it provides better clustering results on both synthetic and real-life datasets (Pelleg & Moore, 2015).

The sewer pipe feature vectors for the original dataset and the generated dataset are fed into the X-means clustering algorithm, respectively, and the clustering results are summarized in Table 6.3. In the table, L_S is the abbreviation for the length of the sewer pipe, NC is N_{crack}, TC is T_{crack}, DC is the D_{crack}, and so on. Both the original dataset and the generated dataset can be grouped into two clusters. For the original dataset, 1,792 feature vectors are assigned to cluster 0, while only 1 feature vector is assigned to cluster 1 by the algorithm automatically. The 1 sewer pipe in cluster 1 is characterized by very long length, with only 5 deposits and 1 tap. The other variables are all equal to 0. For the generated dataset, 1,710 feature vectors are assigned to cluster 0, while 83 feature vectors are assigned to cluster 1. Cluster 1 (83 sewer pipes) of the generated dataset is also characterized by pipes of relatively long length; however, other variables are relatively larger when compared to cluster 1 of the original dataset. The comparison of the two clusters in terms of the cluster centroid can be observed in Figure 6.15 and Figure 6.16. As shown in Figure 6.15, the centroid for all the variables in cluster 0 is quite close with each other, which means that these two clusters in the original dataset and the generated dataset have a high degree of similarity. From Figure 6.16, it can be concluded that the degree of similarity between cluster 1 of these two datasets is not very high, but they both represent a small percentage of the total dataset (0.06% and 4.63% for the original dataset and the generated dataset, respectively).

202

Overall, the majority of the generated dataset is quite similar to the majority of the original dataset, except for the 4.63% of the generated dataset that shows some variation from the original dataset. It is fair to say that the generated dataset is reasonable and has captured the major characteristics of the original dataset.

Table 6.3: The centroids for the X-means cluster analysis.

	Original dataset		Generated dataset	
	Cluster 0	Cluster 1	Cluster 0	Cluster 1
Original	1792 (99.94%)	1 (0.06%)	1710 (95.37%)	83 (4.63%)
L_S	56.00	580.30	53.13	236.26
NC	1.64	0.00	1.49	6.58
TC	61.20	0.00	62.82	778.86
DC	86.01	0.00	110.42	3501.87
NF	0.60	0.00	0.51	2.14
TF	22.30	0.00	21.72	260.53
DF	30.48	0.00	43.21	1504.17
NR	0.48	0.00	0.39	1.45
TR	18.02	0.00	15.59	161.85
DR	12.89	0.00	26.59	618.45
ND	4.48	5.00	3.84	17.19
TD	171.56	2238.70	151.80	2100.86
DD	250.34	31401.75	234.94	4726.51
NB	0.07	0.00	0.05	0.20
TB	2.07	0.00	1.97	29.60
DB	4.01	0.00	0.19	88.11
NH	0.03	0.00	0.02	0.10
TH	0.03	0.00	0.71	13.29
DH	1.49	0.00	1.74	0.06
NT	2.67	1.00	2.37	10.76
TT	109.43	400.30	96.05	1336.56
DT	172.04	0.00	182.58	3914.19
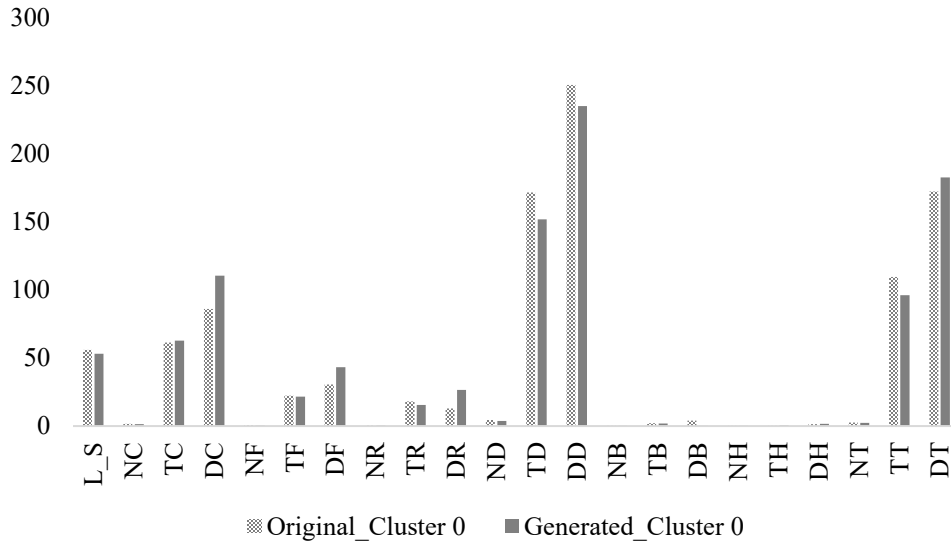
Figure 6.15: The cluster 0 centroid comparison of the original data and the generated data.
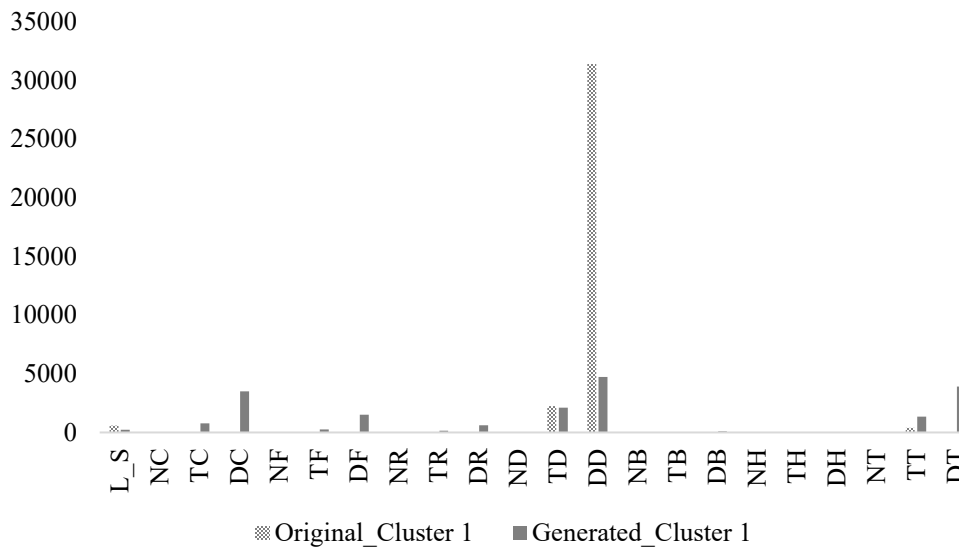
Figure 6.16: The cluster 1 centroid comparison of the original data and the generated data.

6.5.2.2 Distance-based similarity comparison

Hausdorff distance is commonly used to measure the similarity between two pointsets by calculating the maximum of the minimum distance between them (Nutanong et al., 2011). For instance, the Hausdorff distance has been widely used for measuring the similarity between 3D surfaces (Aspert et al., 2002), for objective matching (Sim et al., 1999), and face detection

(Jesorsky et al., 2001), etc. For two pointsets $A = \{a_1, a_2, ..., a_{N_a}\}$, $B = \{b_1, b_2, ..., b_{N_b}\}$, the Hausdorff

distance between A and B is defined by Equations (6-7) - (6-9).

$$h(A,B) = \max_{a \in A} \min_{b \in B} \| a - b \| \tag{6-7}$$

$$h(B,A) = \max_{b \in B} \min_{a \in A} \| a - b \| \tag{6-8}$$

$$H(A,B) = \max[h(A,B), h(B,A)] \tag{6-9}$$

where $\| a - b \|$ is Euclidean distance between the point a and b, $h(A,B)$ is the maximum of the

minimum distance from any point in the dataset A to dataset B, $h(B,A)$ is the maximum of the

minimum distance from any point in the dataset B to dataset A, $H(A,B)$ is Hausdorff distance,

which is the maximum between $h(A,B)$ and $h(B,A)$. If the two pointsets are exactly the same,

then the Hausdorff distance between them is equal to 0. However, the Hausdorff distance can be

highly influenced by the outliers of the pointset (Sim et al., 1999). In order to enhance the

Hausdorff distance measurement, Dubuisson & Jain (1994) propose a modified Hausdorff

distance (MHD) to measure the similarity between two objects. The MHD for the two datasets

($A = \{a_1, a_2, ..., a_{N_a}\}$, $B = \{b_1, b_2, ..., b_{N_b}\}$) is calculated by Equations (6-10) - (6-12).

$$h_{MHD}(A,B) = \frac{1}{N_a} \sum_{a \in A} \min_{b \in B} \| a - b \| \tag{6-10}$$

$$h_{MHD}(B,A) = \frac{1}{N_b} \sum_{b \in B} \min_{a \in A} \| a - b \| \tag{6-11}$$

$$H_{MHD}(A,B) = \max[h_{MHD}(A,B), h_{MHD}(B,A)] \tag{6-12}$$

Given the format of Equations (6-10) - (6-12) and the experiment conducted by Dubuisson &

Jain (2002), it can be concluded that the advantages of MHD can be expressed in terms of two

aspects: 1) *"its value increases monotonically as the amount of difference between the two sets of*

edge points increases", and 2) "*it is robust to outlier points that might result from segmentation errors*" (Dubuisson & Jain, 1994). In order to present the calculation process, $Dis(A,B)$ is defined here as a set of variables that is a collection of minimum distances from points in the dataset A to any of the points in the dataset B, according to Equation (6-13); $Dis(B,A)$ is also defined as a set of variables that is a collection of minimum distances from points in the dataset B to any points in the dataset A, according to Equation (6-14).

$$Dis(A,B) = \{\min_{b \in B} \| a-b \|, a \in A\} \tag{6-13}$$

$$Dis(B,A) = \{\min_{a \in A} \| a-b \|, b \in B\} \tag{6-14}$$

Therefore, $h(A,B) = \max(Dis(A,B))$ and $h(A,B) = \max(Dis(B,A))$.

The original Haufdorff distance and the MHD are calculated to measure the similarity between the original dataset and the generated dataset. The results of $Dis(Original, Generated)$ and $Dis(Generated, Original)$ are plotted as histograms presented in Figure 6.17. The detailed information of these two sets, $Dis(Original, Generated)$ and, $Dis(Generated, Original)$ is also presented in the figure. The two histograms are both skewed to the left, which means that most of the points can find a point in the opposite dataset that has a distance that is close to 0. $H(Original, Generated)$ is equal to 23,612, which is a relatively large value considering the definition of the Hausdorff distance. However, the value is influenced by the outliers of the original dataset and the generated dataset, which can be inferred by the smaller mean value and larger standard deviation presented in Figure 6.17. The percentiles of the two histograms are also calculated and presented in Figure 6.17 to prove that most of the values of $Dis(Original, Generated)$ and $Dis(Generated, Original)$ are small values, which means that most of the sewer pipe feature vectors can find similar sewer pipe feature vectors in the opposite dataset.

In addition, the MHD is also calculated and the value is 406.34, which is a relatively small value compared to the Hausdorff distance between the two datasets. Overall, the distance-based method conducted in this research can prove that the generated dataset has a high degree of similarity to the original dataset, that the generated dataset is reasonable, and that the generated dataset presents the main features of the original dataset.
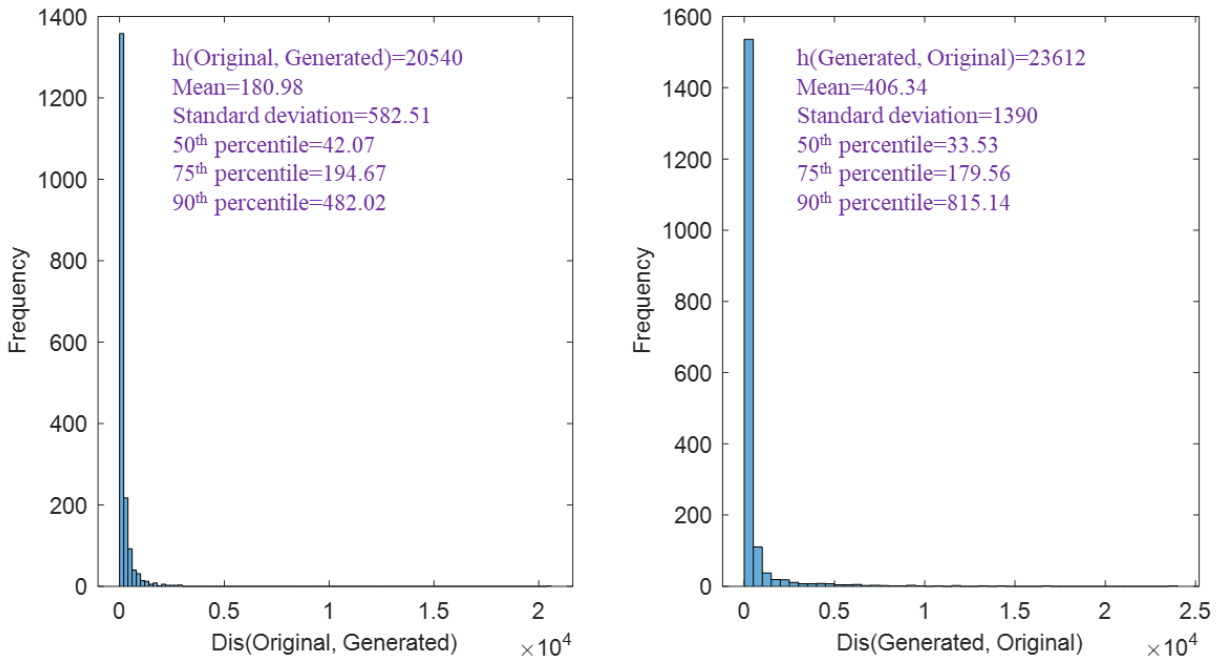
Figure 6.17: Histogram of Hausdorff distance results.

6.6 Discussion

Essentially, to develop a simulation model is to abstract the real-world with mathematical methods, which requires defining an appropriately narrow scope for the development of the simulation model, otherwise the scope could be too broad to be simulated. As for the rate at which defects occur in a sewer pipe, the location of a sewer pipe could be an important factor that influences the deterioration rate of sewer pipes (Balekelayi & Tesfamariam, 2019; Yin et al., 2020b). In this research, we exclude the location information for the following reasons: (1) the availability of location information in the dataset used in this research is limited, (2) to consider

more variables is to increase the scope of the research and the complexity of this issue deserves much more attention and another comprehensive study in addition to the current study should be conducted. Therefore, in our future research, other factors, such as location information, sewer type, pipe material, pipe age, etc., could be incorporated in the data generation process to mimic the real world in a more detailed and comprehensive manner.

In this research, the main objective is to develop a framework to 1) develop a method to generate sewer pipe data from a sample of sewer pipe data, and 2) develop a process to validate the generated dataset. Therefore, in practice, instead of using the data generated in the present study for other sewer pipe systems, a municipal department should adopt the proposed framework to generate their data and validate the generated data following the proposed validation method. If the generated data serves as any kind of input in the sewer pipe maintenance model (e.g., working process scheduling model, deterioration model, etc.), then the generated data should be updated at a certain frequency that would be determined by the real-world application of the data. Essentially, updating the generated dataset requires updating the original dataset (which is the real-life data collected from the targeted sewer network), and the generated dataset would then have similar characteristics as the new version of the original dataset. The generated data could be used as input for the maintenance scheduling model, and the frequency with which it is updated should be determined by the application of the generated data. For example, in order to schedule the next annual maintenance plan for a specific city, a certain amount of sewer pipe data should be collected according to standard sampling techniques to ensure the sample can represent the whole population to an appropriate extent. With the collected dataset, a new dataset that contains a larger number of data points could be generated following the proposed

framework. In this annual maintenance plan example, the original dataset should be updated at least once annually to reflect the current condition of the sewer pipe network.

The current research proposes an input modeling method, which will generate the human-made dataset that contains the major characteristics of the original dataset. Deviations exist between the original dataset and the generated dataset, which also means that the generated dataset cannot represent the real-life situation for one hundred percent. Therefore, there is a certain level of risk of using such a generated dataset. For example, using the generated dataset to predict the next annual cost of the sewer pipe maintenance, there are risks in this process and may lead to inaccurate predictions since the generated dataset can only represent the major characteristics of the sewer pipe system. The level of risk is influenced by 1) the performance of the proposed data generation framework, 2) the quality of the original dataset, and 3) the performance of the simulation model fed with the generated dataset. To measure the level of risk, a comprehensive study is needed that takes into consideration all the possible influencing factors that could contribute to this type of risk.

In this research, the dataset similarity comparison depends primarily on the subjective evaluation of the selected indices, such as the general statistical comparison and the feature vector similarity comparison. In terms of quantitatively measuring the similarity between both datasets, a limitation of the present research is the lack of quantitative indices to provide this evidence in a conclusive manner. In the future, a more determined quantitative comparison method should be investigated and used for the purpose of dataset similarity comparison.

As mentioned in Section 6.2, sewer pipe defects can be categorized in a more detailed manner according to PACP; for example, the crack can be classified as CL, CC, CM, CS, and CH. However, this research only considers the major categories, which means crack is used to

represent all its subclasses. In future research, the types of defects can be further detailed to randomly generate sewer pipe data using the subclasses of sewer pipe defects. Another enhancement can be achieved in the future by using a larger size of the original dataset to include more comprehensive features of the sewer pipe data.

6.7 Conclusion

An input modeling process for the random generation of sewer pipe data is proposed in this research. The model is driven by historical CCTV inspection data for sewer pipes, which include the sewer pipe lengths, types of features and defects (which include crack, fracture, root, deposits, broken, hole, and tap), and location of the defects (and taps). The dataset is valuable for sewer pipe preventive maintenance in various ways. By interpreting the characteristics of the original dataset, a model that combines the Markov chain model and distribution fitting techniques is used to randomly generate sewer pipe data. The transfer matrix is the core part of the Markov chain model, and together with the initial state of the sewer pipe data, the defect sequence is generated. The sewer pipe length and defect interval are fitted with a beta distribution. However, the defect interval is separated into two parts to fit with two beta distributions, respectively. Then, following the proposed sewer pipe data generation algorithm, 1,793 sewer pipes are randomly generated to compare with the same quantity from the original data. The generated data is validated using a rigorous validation process, which includes statistical analysis and comparison, X-means clustering analysis, and a Hausdorff distance-based similarity comparison. The validation process proves that the generated dataset is reasonable since it has a high degree of similarity with the original dataset and also has inherited the major characteristics of the original dataset.

The contributions of the research are four-folds: 1) The characteristics of the CCTV inspection data are well investigated and understood. 2) The sewer pipe data is randomly generated by integrating the conventional input modeling technique and the proposed Markov chain process. 3)The similarity of the generated data and the original data is validated by a rigorous validation process, which can serve as a framework for other purposes of statistical similarity comparison in the perspective of methodology. 4) Overall, the research proposes an input modeling process that could generate human-made sewer pipe inspection data that inherent the major characteristic of the real-life data; The generated data could be used for various purposes, such as input data for the simulation model to simulate the sewer pipe maintenance process in the context of data deficiency; city/country-wide sewer pipe health assessment, where retrieving the complete data is impossible or cost-prohibitive.

Chapter 7: CONCLUSIONS

7.1 Research summary

Regular preventive maintenance is necessary to keep a sewer pipe system at a functional level of service. The main tool used in this process is CCTV inspection, which includes the on-site CCTV video collection, and off-site video assessment. After the video assessment process, assessment reports that record the detailed information (e.g., defect type and defect location) of the sewer pipes are generated by the pipe technologists. The overall health condition of pipes can be calculated based on the assessment report. Indicators such as LOF could represent the health of the pipe to some extent. Given the large amount of data that records the detailed information of sewer pipes (including health condition), a deterioration model can be developed to predict the future condition of sewer pipes, which is the basis for decision making in future maintenance planning processes. This research is focusing on the above-mentioned process with a goal to improve productivity and consistency, and increase the level of automation in this process. In order to realize the goal of this research, the following objectives guided the research activities that are summarized below:

1) *Simulation model development for CCTV recording process*. The CCTV recording process accounts for a large portion (29%) of the total time required to complete the entire CCTV collection process (which includes flushing, recording, traveling between manholes, locating manholes, and camera installation). Since other major processes have been investigated in previous studies, the CCTV recording process is selected as the focus in this research. A machine learning algorithm is used to extract the benchmark for the CCTV recording process, which provides a standard for future recording processes. The simulation model is developed by integrating the benchmark and the results from the

regression analysis. The simulation model can be used for predicting the CCTV video duration in the context of a large-scale video collection process, which is valuable information for planning and scheduling CCTV video collection as well as the offsite assessment process. Building the simulation model also helps to understand the CCTV recording process. In the case study, one of the major findings is that the CCTV crawler is moving too fast compared to the regulated speed in PACP, which is one of the aspects that need to be improved in future daily operations. The proposed novel method of data separation using the RANSAC algorithm could serve as a framework that is used in other studies that require a data separation operation.

2) *Defect detection system development.* To assess the CCTV videos is a tedious and time-consuming job for the technologists in current practice. In addition, the inconsistency of the assessment results is another problem in the manual assessment process since human error is always inevitable. Therefore, this research proposes a deep learning-based defect detection system using YOLOv3 as the core object detection algorithm. The system outperforms other similar system in terms of processing speed. The processing accuracy is also better in comparison with other systems. A framework is proposed in this research to formulate the information and data flow required to build such a defect detection system for sewer pipes.

3) *Video interpretation system development.* To assess the CCTV video requires extracting the information from video format to text format. The information primarily includes the type and location of a defect or construction feature within a sewer pipe. The defects and construction features can be detected and labeled by the defect detector developed to accomplish the second objective; however, the information is still embedded in the video,

which means the technologists still need to watch through every video and record the information manually even though the videos are already labeled. Therefore, the author proposed a sewer pipe video assessment (SPVA) system that features a novel video interpretation algorithm, VIASP, which could translate the information from video format to text format automatically. With the proposed SPVA system, the productivity and the level of automation of the overall CCTV video assessment process is improved significantly. The proposed framework could be used as a benchmark for developing automated video assessment systems for other types of civil infrastructures (such as bridge video assessment and pavement video assessment).

4) *Deterioration model development*. A thorough literature review is conducted focusing on the previously developed deterioration models for sewer pipes. The results show that there is little research published on the topic of developing deterioration models at the neighborhood level, instead, most of the previous works are focused on deterioration at the individual pipe level. In addition, previous deterioration model development follows the idea of trying to involve as many relevant influencing factors as possible, which is adjusted in this research as to include only the critical influencing factors. The neighborhood-level deterioration model is developed using stepwise regression analysis and the model results are presented visually by GIS. The individual-level deterioration model is built by the NN model integrated with the backward variable elimination process. The deterioration curves for different types of pipes are developed as a result. The bi-level model provides city managers with both the macro (neighborhood level) and micro (individual level) views of the sewer pipe condition, which is critically important information for making future maintenance plans.

214

5) *Input model development for CCTV inspection data*. Data is a valuable resource in the context of data-driven decision making, which is true for data-driven preventive maintenance for sewer pipes as well. This research investigates the characteristics of the CCTV inspection dataset for sewer pipes and abstracts the dataset such that it can be represented by a mathematical method, which is the input model in this case. The input model integrates the Markov chain model with distribution fitting techniques to generate a human-made CCTV inspection dataset that retains similar characteristics to the real-life dataset. The generated dataset has been through a rigorous validation process including statistical comparison, cluster analysis and comparison, and distance-based similarity comparison. This study generates a large human-made dataset that can be used as input for various models (e.g., simulation model, and deterioration model), and proposes a method that serves as a framework for generating and validating human-made datasets in other engineering areas with data deficiency problems.

7.2 Research contributions

The research proposed a framework using machine learning techniques for improving the productivity, consistency, and level of automation of the major processes (e.g., CCTV recording, CCTV video assessment, and deterioration model development) involved in the preventive maintenance of sewer pipes. The contributions of the present research in terms of industrial application and in terms of the body of knowledge are listed in the sections that follow.

7.2.1 Industrial contributions

1) A simulation model is proposed for the purpose of simulating the CCTV recording process of sewer pipes. The model has the ability to predict the recording duration (i.e., video duration) of large-scale CCTV video collection, and this prediction of the duration

serves as the basis for scheduling future CCTV collection and video assessment processes effectively at the city scale.

2) The SPVA system has been developed for processing the CCTV video to extract and translate the information (e.g., type and location of defects) embedded within the CCTV video from video format to text format automatically. This tool can realize the automation of the CCTV assessment process, increase the productivity of this previously time-consuming job, and decrease the required number of technologists for this job as well.

3) A bi-level deterioration model has been developed to predict the health condition of sewer pipes both at the neighborhood level and individual pipe level. The neighborhood-level deterioration model helps to identify critical neighborhoods that require more attention than others, while the individual-level deterioration model helps to identify sewer pipes that need immediate maintenance. Therefore, this tool can help city managers effectively project required resources and make future maintenance plans.

7.2.2 Academic contributions

1) A machine learning algorithm (i.e., RANSAC) is proposed for benchmark extraction and dataset separation (i.e., the CCTV recording dataset in this research). In particular, the algorithm is more suitable for 2D or 3D space separation considering visibility.

2) VIASP has been proposed to interpret the information contained in CCTV videos of sewer pipes. The goal is to avoid duplication and noise in the process of translating the video information to text information. VIASP employs a novel concept, namely the defect frame cluster (DFC), which is aimed at grouping video frames that record the same defect together, at the same time, filtering out the noisy frames within the grouped video

frames. The algorithm can be extensively used for other civil infrastructures video assessment processes.

3) The concept of feature selection has been adopted in the development of the deterioration model. Unlike the previous deterioration model development processes that attempt to include as many related influencing factors as possible, this research follows the idea that it is possible to use only the critical influencing factors in the deterioration model, considering several reasons, such as the number of required collected variables normally positively correlated to the cost of data collection.

4) An input model to generate a human-made CCTV video inspection dataset has been developed and validated. The proposed method integrates the Markov chain model with distribution fitting techniques to extract the characteristic of the real-life dataset and to generate a new dataset that retains all the major characteristics. The proposed validation process, which includes statistical comparison, cluster analysis and comparison, and distance-based similarity comparison, serves as a framework that can be used for other dataset similarity comparisons.

7.3 Limitations and future research

In order to improve the performance of the proposed framework, the following research directions can be pursued in future work:

1) Increasing the data quantity and quality could always be a direction for improvement in future research. More advanced data collection methods could be adopted for data collection. For example, instead of using the current camera truck that is equipped with an adjustable facing-forward camera, a new version of a camera truck could be outfitted with a panoramic camera for use in future video collection. A panoramic camera can

217

capture a much wider view than the normal camera, which gives the camera truck the ability to take detailed images of defects without stopping or tilting.

2) The defect detector and the SPVA system can be improved for detecting and outputting information in addition to the current information that includes the defect type, confidence level, frame number, and location in meters. The additional information could be the sub-classes of the defect type according to PACP, the position of the defect on a clock face, the condition grade of each detected defect, and the overall calculated condition (e.g., pipe score or LOF) of sewer pipes.

3) One of the more valuable research directions is to investigate how to effectively train the deep learning-based defect detector. Currently, to train such a defect detector requires a significant amount of manual operation (e.g., to label the training images), computing resources, and computing time, which potentially could be improved by redesigning the training strategy.

4) More advanced feature selection methods could be adopted in the development of the deterioration model. This research uses a stepwise method to select the input variables for the regression analysis and a backward variable elimination process for the NN model. In future research, more advanced methods such as the genetic algorithm or particle swarm optimization algorithm could be used for finding the best combination of the input variables for the deterioration model.

5) In addition to the data-driven method used in this research, future studies based on physical and chemical principals is also a valuable research direction. For example, the relationship between the sewer pipe deterioration rates and the influencing factors could be explained by conducting physical or chemical experiments in future studies.

REFERENCES

Abourizk, S. (2010). Role of simulation in construction engineering and management. *Journal of Construction Engineering and Management*, *136*(10), 1140–1153. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000220

Abourizk, S. M., Hague, S. A., & Ekyalimpa, R. (2016). *Construction Simulation: An Introduction Using SIMPHONY*.

Abourizk, S. M., Halpin, D. W., & Wilson, J. R. (1994). Fitting beta distributions based on sample data. *Journal of Construction Engineering and Management*, *120*(2), 288–305. https://doi.org/10.1061/(ASCE)0733-9364(1994)120:2(288)

AbouRizk, S., & Mohamed, Y. (2000). Simphony-an integrated environment for construction simulation. *2000 Winter Simulation Conference Proceedings (Cat. No.00CH37165)*, *2*, 1907–1914 vol.2. https://doi.org/10.1109/WSC.2000.899185

Adeli, H. (2001). Neural networks in civil engineering: 1989-2000. *Computer-Aided Civil and Infrastructure Engineering*, *16*(2), 126–142. https://doi.org/10.1111/0885-9507.00219

Agbulos, A., Mohamed, Y., Al-Hussein, M., Abourizk, S., & Roesch, J. (2006). Application of lean concepts and simulation analysis to improve efficiency of drainage operations maintenance crews. *Journal of Construction Engineering and Management*, *132*(3), 291–299. https://doi.org/10.1061/(ASCE)0733-9364(2006)132:3(291)

AL-Alawi, M., Bouferguene, A., & Mohamed, Y. (2018). Random generation of industrial pipelines' data using Markov chain model. *Advanced Engineering Informatics*, *38*, 725–745. https://doi.org/10.1016/j.aei.2018.10.003

Al-antari, M. A., Al-masni, M. A., Choi, M. T., Han, S. M., & Kim, T. S. (2018). A fully integrated computer-aided diagnosis system for digital X-ray mammograms via deep learning detection, segmentation, and classification. *International Journal of Medical Informatics*, *117*(June), 44–54. https://doi.org/10.1016/j.ijmedinf.2018.06.003

Al-Barqawi, H., & Zayed, T. (2008). Infrastructure Management: Integrated AHP/ANN Model to Evaluate Municipal Water Mains' Performance. *Journal of Infrastructure Systems*, *14*(4), 305–318. https://doi.org/10.1061/(asce)1076-0342(2008)14:4(305)

Al-masni, M. A., Al-antari, M. A., Park, J. M., Gi, G., Kim, T. Y., Rivera, P., … Kim, T. S.

(2018). Simultaneous detection and classification of breast masses in digital mammograms via a deep learning YOLO-based CAD system. *Computer Methods and Programs in Biomedicine*, *157*, 85–94. https://doi.org/10.1016/j.cmpb.2018.01.017

Alanis, R., Ingolfsson, A., & Kolfal, B. (2013). A markov chain model for an EMS system with repositioning. *Production and Operations Management*. https://doi.org/10.1111/j.1937-5956.2012.01362.x

Altaf, M. S., Bouferguene, A., Liu, H., Al-Hussein, M., & Yu, H. (2018). Integrated production planning and control system for a panelized home prefabrication facility using simulation and RFID. *Automation in Construction*, *85*(November 2017), 369–383. https://doi.org/10.1016/j.autcon.2017.09.009

Ana, E., Bauwens, W., Pessemier, M., Thoeye, C., Smolders, S., Boonen, I., & de Gueldre, G. (2009). An investigation of the factors influencing sewer structural deterioration. *Urban Water Journal*, *6*(4), 303–312. https://doi.org/10.1080/15730620902810902

Ana, E. V, & Bauwens, W. (2010). Modeling the structural deterioration of urban drainage pipes: The state-of-the-art in statistical methods. *Urban Water Journal*, *7*(1), 47–59. https://doi.org/10.1080/15730620903447597

Anbari, M. J., Tabesh, M., & Roozbahani, A. (2017). Risk assessment model to prioritize sewer pipes inspection in wastewater collection networks. *Journal of Environmental Management*. https://doi.org/10.1016/j.jenvman.2016.12.052

Ariaratnam, S. T., & MacLeod, C. W. (2002). Financial Outlay Modeling for a Local Sewer Rehabilitation Strategy. *Journal of Construction Engineering and Management*. https://doi.org/10.1061/(asce)0733-9364(2002)128:6(486)

Arthur, S., Crow, H., Pedezert, L., & Karikas, N. (2009). The holistic prioritisation of proactive sewer maintenance. *Water Science and Technology*, *59*(7), 1385–1396. https://doi.org/10.2166/wst.2009.134

Aspert, N., Santa-Cruz, D., & Ebrahimi, T. (2002). MESH: Measuring errors between surfaces using the Hausdorff distance. *Proceedings - 2002 IEEE International Conference on Multimedia and Expo, ICME 2002*, *1*, 705–708. https://doi.org/10.1109/ICME.2002.1035879

Austin, P. C., & Tu, J. V. (2004). Automated variable selection methods for logistic regression produced unstable models for predicting acute myocardial infarction mortality. *Journal of*

Clinical Epidemiology, *57*(11), 1138–1146. https://doi.org/10.1016/J.JCLINEPI.2004.04.003

Baah, K., Dubey, B., Harvey, R., & McBean, E. (2015). A risk-based approach to sanitary sewer pipe asset management. *Science of the Total Environment*, *505*, 1011–1017. https://doi.org/10.1016/j.scitotenv.2014.10.040

Baik, H., Seok, H., Jeong, D., & Abraham, D. M. (2006). Estimating Transition Probabilities in Markov Chain-Based Deterioration Models for Management of Wastewater Systems. *Journal of Water Resources Planning and Management*, *132*(February), 15–24. https://doi.org/10.1061/(ASCE)0733-9496(2006)132:1(15)

Balekelayi, N., & Tesfamariam, S. (2019). Statistical Inference of Sewer Pipe Deterioration Using Bayesian Geoadditive Regression Model. *Journal of Infrastructure Systems*, *25*(3), 04019021. https://doi.org/10.1061/(asce)is.1943-555x.0000500

Barath, D., & Matas, J. (2018). Graph-Cut RANSAC. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 6733–6741. https://doi.org/10.1109/CVPR.2018.00704

Baur, R., & Herz, R. (2002). Selective inspection planning with ageing forecast for sewer types. *Water Science and Technology*, *46*(6–7), 389–396. https://doi.org/10.2166/wst.2002.0704

Cha, Y. J., Choi, W., & Büyüköztürk, O. (2017). Deep Learning-Based Crack Damage Detection Using Convolutional Neural Networks. *Computer-Aided Civil and Infrastructure Engineering*, *32*(5), 361–378. https://doi.org/10.1111/mice.12263

Chatterjee, S., Sarkar, S., Hore, S., Dey, N., Ashour, A. S., & Balas, V. E. (2017). Particle swarm optimization trained neural network for structural failure prediction of multistoried RC buildings. *Neural Computing and Applications*, *28*(8), 2005–2016. https://doi.org/10.1007/s00521-016-2190-2

Cheng, J. C. P., & Wang, M. (2018). Automated detection of sewer pipe defects in closed-circuit television images using deep learning techniques. *Automation in Construction*, *95*, 155–171. https://doi.org/10.1016/j.autcon.2018.08.006

Chinyio, E. A., Olomalaiye, P. O., Kometa, S. T., & Harris, F. C. (1998). A needs-based methodology construction clients and selecting contractors. *Construction Management and Economics*, *16*, 91–98. https://doi.org/10.1080/014461998372628

Chong, E. K. P., & Zak, S. H. (2013). An Introduction to Optimization: Fourth Edition. In *Wiley*

Series in Discrete Mathematics and Optimization, *76*. https://doi.org/10.1002/9781118033340

Chughtai, F., & Zayed, T. (2007). Sewer Pipeline Operational Condition Prediction using Multiple Regression. *Pipelines 2007: Advances and Experiences with Trenchless Pipeline Projects*, 1–11. Retrieved from https://doi.org/10.1061/40934(252)18

Chughtai, F., & Zayed, T. (2008). Infrastructure Condition Prediction Models for Sustainable Sewer Pipelines. *Journal of Performance of Constructed Facilities*, *22*(5), 333–341. https://doi.org/10.1061/(ASCE)0887-3828(2008)22:5(333)

CIRC. (2016). *2016 Canadian Infrastructure Report Card: Informing the Future*. Retrieved from https://www.pppcouncil.ca/web/pdf/infra_report_card_2016.pdf

CIRC. (2019). *Canada Infrastructure Report Card 2019*. 1–56. Retrieved from http://canadianinfrastructure.ca/en/about.html

Ćorović, A., Ilić, V., Đurić, S., Marijan, M., & Pavković, B. (2018). The Real-Time Detection of Traffic Participants Using YOLO Algorithm. *26th Telecommunications Forum TELFOR*, 1–4. https://doi.org/10.1109/telfor.2018.8611986

Dang, L. M., Hassan, S. I., Im, S., Mehmood, I., & Moon, H. (2018). Utilizing text recognition for the defects extraction in sewers CCTV inspection videos. *Computers in Industry*, *99*, 96–109. https://doi.org/10.1016/j.compind.2018.03.020

Dirksen, J., & Clemens, F. H. L. R. (2008). Probabilistic modeling of sewer deterioration using inspection data. *Water Science and Technology*, *57*(10), 1635–1641. https://doi.org/10.2166/wst.2008.308

Dong-Gyu Sim, Oh-Kyu Kwon, & Rae-Hong Park. (1999). Object matching algorithms using robust Hausdorff distance measures. *IEEE Transactions on Image Processing*, *8*(3), 425–429. https://doi.org/10.1109/83.748897

Dubuisson, M.-P., & Jain, A. K. (1994). A modified Hausdorff distance for object matching. *Proceedings of 12th International Conference on Pattern Recognition*, *1*(1), 566–568. https://doi.org/10.1109/ICPR.1994.576361

Duchesne, S., Beardsell, G., Villeneuve, J. P., Toumbou, B., & Bouchard, K. (2013). A Survival Analysis Model for Sewer Pipe Structural Deterioration. *Computer-Aided Civil and Infrastructure Engineering*, *28*(2), 146–160. https://doi.org/10.1111/j.1467-8667.2012.00773.x

Dung, C. V., & Anh, L. D. (2019). Autonomous concrete crack detection using deep fully convolutional neural network. *Automation in Construction*, *99*(December 2018), 52–58. https://doi.org/10.1016/j.autcon.2018.11.028

Duran, O., Althoefer, K., & Seneviratne, L. D. (2007). Automated pipe defect detection and categorization using caniera/laser-based profiler and artificial neural network. *IEEE Transactions on Automation Science and Engineering*, *4*(1), 118–126. https://doi.org/10.1109/TASE.2006.873225

Duran, Olga, Althoefer, K., & Seneviratne, L. D. (2002a). State of the art in sensor technologies for sewer inspection. *IEEE Sensors Journal*, *2*(2), 73–81. https://doi.org/10.1109/JSEN.2002.1000245

Duran, Olga, Althoefer, K., & Seneviratne, L. D. (2002b). State of the art in sensor technologies for sewer inspection. *IEEE Sensors Journal*, *2*(2), 73–81. https://doi.org/10.1109/JSEN.2002.1000245

Ehret, E. (2011). Sewer Condition Assessment & Rehabilitation. Retrieved from http://www.ohiowea.org/docs/03_Sewer_Conditions_Rehab.pdf

El-Housni, H., Ouellet, M., & Duchesne, S. (2017). Identification of most significant factors for modeling deterioration of sewer pipes. *Canadian Journal of Civil Engineering*, *45*(3), 215–226. https://doi.org/10.1139/cjce-2015-0293

Elmasry, M., Zayed, T., & Hawari, A. (2018). Defect-based ArcGIS tool for prioritizing inspection of sewer pipelines. *Journal of Pipeline Systems Engineering and Practice*, *9*(4), 1–13. https://doi.org/10.1061/(ASCE)PS.1949-1204.0000342

Elmasry, M., Zayed, T., & Hawari, A. (2019). Multi-Objective Optimization Model for Inspection Scheduling of Sewer Pipelines. *Journal of Construction Engineering and Management*, *145*(2), 1–11. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001599

Ennaouri, I., & Fuamba, M. (2013). New integrated condition-assessment model for combined storm-sewer systems. *Journal of Water Resources Planning and Management*, *139*(1), 53–64. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000217

Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., & Zisserman, A. (2009). The pascal visual object classes (VOC) challenge. *International Journal of Computer Vision*, *88*(2), 303–338. https://doi.org/10.1007/s11263-009-0275-4

Fenner, R. A. (2000). Approaches to sewer maintenance: A review. *Urban Water*, *2*(4), 343–356.

https://doi.org/10.1016/s1462-0758(00)00065-0

Fischler, M. a, & Bolles, R. C. (1981). Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, *24*(6), 381–395. https://doi.org/10.1145/358669.358692

Fontecha, J. E., Akhavan-Tabatabaei, R., Duque, D., Medaglia, A. L., Torres, M. N., & Rodríguez, J. P. (2016). On the preventive management of sediment-related sewer blockages: A combined maintenance and routing optimization approach. *Water Science and Technology*, *74*(2), 302–308. https://doi.org/10.2166/wst.2016.160

Frigge, I., Hoaglin, D. C., & Iglewicz, B. (1989). Some Implementations of the Boxplot. *The American Statistician*, *43*(1), 50–54. https://doi.org/10.1080/00031305.1989.10475612

Gandhi, R. (2018). R-CNN, Fast R-CNN, Faster R-CNN, YOLO — Object Detection Algorithms. Retrieved March 22, 2019, from https://towardsdatascience.com/r-cnn-fast-r-cnn-faster-r-cnn-yolo-object-detection-algorithms-36d53571365e

Gao, J., Lei, L., & Yu, S. (2015). Big data sensing and service: A tutorial. *Proceedings - 2015 IEEE 1st International Conference on Big Data Computing Service and Applications, BigDataService 2015*, (March), 79–88. https://doi.org/10.1109/BigDataService.2015.45

Girshick, R. (2015). Fast R-CNN. *Proceedings of the IEEE International Conference on Computer Vision*, *2015 Inter*, 1440–1448. https://doi.org/10.1109/ICCV.2015.169

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 580–587. https://doi.org/10.1109/CVPR.2014.81

Glorot, X., Bordes, A., & Bengio, Y. (2011). Deep Sparse Rectifier Neural Networks. *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, *15*, 315–323. https://doi.org/10.1.1.208.6449

Gong, J., & Caldas, C. H. (2011). An object recognition, tracking, and contextual reasoning-based video interpretation method for rapid productivity analysis of construction operations. *Automation in Construction*, *20*(8), 1211–1226. https://doi.org/10.1016/j.autcon.2011.05.005

Guerzhoy, M. (2015). *Lecture Notes: CSC320 : Introduction to Visual Computing Last time : Non-Maximum Suppression*. Retrieved from

https://www.cs.toronto.edu/~guerzhoy/320/lec/upsampling.pdf

Guo, W., Soibelman, L., & Garrett, J. H. (2009). Automated defect detection for sewer pipeline inspection and condition assessment. *Automation in Construction*, *18*(5), 587–596. https://doi.org/10.1016/j.autcon.2008.12.003

Guyon, I., & A, E. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research*. https://doi.org/10.1162/153244303322753616

Haan, C. T., Allen, D. M., & Street, J. O. (1976). A Markov Chain Model of daily rainfall. *Water Resources Research*. https://doi.org/10.1029/WR012i003p00443

Hackl, J., Adey, B. T., & Lethanh, N. (2018). Determination of Near-Optimal Restoration Programs for Transportation Networks Following Natural Hazard Events Using Simulated Annealing. *Computer-Aided Civil and Infrastructure Engineering*, *33*(8), 618–637. https://doi.org/10.1111/mice.12346

Hahn, M. A., Palmer, R. N., Merrill, M. S., & Lukas, A. B. (2002). Expert system for prioritizing the inspection of sewers: Knowledge base formulation and evaluation. *Journal of Water Resources Planning and Management*, *128*(2), 121–129. https://doi.org/10.1061/(ASCE)0733-9496(2002)128:2(121)

Hajek, B. (1988). Cooling Schedules for Optimal Annealing. *Mathematics of Operations Research*, *13*(2), 311–329. https://doi.org/10.1287/moor.13.2.311

Harvey, R. R., & McBean, E. A. (2014). Predicting the structural condition of individual sanitary sewer pipes with random forests. *Canadian Journal of Civil Engineering*, *41*(4), 294–303. https://doi.org/10.1139/cjce-2013-0431

Hassan, S. I., Dang, L. M., Mehmood, I., Im, S., Choi, C., Kang, J., … Moon, H. (2019a). Underground sewer pipe condition assessment based on convolutional neural networks. *Automation in Construction*, *106*, 102849. https://doi.org/10.1016/j.autcon.2019.102849

Hassan, S. I., Dang, L. M., Mehmood, I., Im, S., Choi, C., Kang, J., … Moon, H. (2019b). Underground sewer pipe condition assessment based on convolutional neural networks. *Automation in Construction*, *106*, 102849. https://doi.org/10.1016/j.autcon.2019.102849

Haurum, J. B., & Moeslund, T. B. (2020). A Survey on Image-Based Automation of CCTV and SSET Sewer Inspections. *Automation in Construction*, *111*, 103061. https://doi.org/10.1016/J.AUTCON.2019.103061

Hawari, A., Alamin, M., Alkadour, F., Elmasry, M., & Zayed, T. (2018). Automated defect

detection tool for closed circuit television (cctv) inspected sewer pipelines. *Automation in Construction*, *89*(December 2017), 99–109. https://doi.org/10.1016/j.autcon.2018.01.004

Hawari, A., Alkadour, F., Elmasry, M., & Zayed, T. (2017). Simulation-Based Condition Assessment Model for Sewer Pipelines. *Journal of Performance of Constructed Facilities*, *31*(1). https://doi.org/10.1061/(ASCE)CF.1943-5509.0000914

Hawari, A., Alkadour, F., Elmasry, M., & Zayed, T. (2018). Condition assessment model for sewer pipelines using fuzzy-based evidential reasoning. *Australian Journal of Civil Engineering*, *16*(1), 23–37. https://doi.org/10.1080/14488353.2018.1444333

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2020). Mask R-CNN. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. https://doi.org/10.1109/TPAMI.2018.2844175

He, K., Zhang, X., Ren, S., & Sun, J. (2015a). Deep Residual Learning for Image Recognition. *International Conference on Machine Learning*, 1–17. https://doi.org/10.1109/CVPR.2016.90

He, K., Zhang, X., Ren, S., & Sun, J. (2015b). Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *37*(9), 1904–1916. https://doi.org/10.1109/TPAMI.2015.2389824

Hoang, N.-D., Nguyen, Q. L., & Tran, V. D. (2018). Automatic recognition of asphalt pavement cracks using metaheuristic optimized edge detection algorithms and convolution neural network. *Automation in Construction*, *94*(July), 203–213. https://doi.org/10.1016/j.autcon.2018.07.008

Hocking, R. . R. . (1976). The Analysis and Selection of Variables in Linear Regression. *BIOMETRICS*, *32*(1), 1–49. Retrieved from http://www.jstor.org/stable/2529336

Holt, G. D. (1996). Applying cluster analysis to construction contractor classification. *Building and Environment*, *31*(6), 557–568. https://doi.org/10.1016/0360-1323(96)00028-5

Infrastructure Canada. (2018). *Investing in Canada : Canada's long-term infrastructure plan.* Retrieved from https://www.infrastructure.gc.ca/site/alt-format/pdf/plan/icp-pic/IC-InvestingInCanadaPlan-ENG.pdf

Iyer, S., & Sinha, S. K. (2005). A robust approach for automatic detection and segmentation of cracks in underground pipeline images. *Image and Vision Computing*, *23*(10), 921–933. https://doi.org/10.1016/j.imavis.2005.05.017

Iyer, S., & Sinha, S. K. (2006). Segmentation of pipe images for crack detection in buried sewers.

Computer-Aided Civil and Infrastructure Engineering, *21*(6), 395–410. https://doi.org/10.1111/j.1467-8667.2006.00445.x

Jensen, H., Biggs, C. A., & Karunakaran, E. (2016). The importance of sewer biofilms. *Wiley Interdisciplinary Reviews: Water*, *3*(4), 487–494. https://doi.org/10.1002/wat2.1144

Jesorsky, O., Kirchberg, K. J., & Frischholz, R. W. (2001). Robust face detection using the Hausdorff distance. In J. Bigun & F. Smeraldi (Eds.), *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 2091, pp. 90–95). https://doi.org/10.1007/3-540-45344-x_14

Johnson, D. S., Aragon, C. R., McGeoch, L. A., & Schevon, C. (1991). Optimization by Simulated Annealing: An Experimental Evaluation; Part II, Graph Coloring and Number Partitioning. *Operations Research*, *39*(3), 378–406. https://doi.org/10.1287/opre.39.3.378

Jung, J., Jayakrishnan, R., & Park, J. Y. (2016). Dynamic Shared-Taxi Dispatch Algorithm with Hybrid-Simulated Annealing. *Computer-Aided Civil and Infrastructure Engineering*, *31*(4), 275–291. https://doi.org/10.1111/mice.12157

Kanan, R., Elhassan, O., & Bensalem, R. (2018). An IoT-based autonomous system for workers' safety in construction sites with real-time alarming, monitoring, and positioning strategies. *Automation in Construction*, *88*(November 2016), 73–86. https://doi.org/10.1016/j.autcon.2017.12.033

Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., & Li, F. F. (2014). Large-scale video classification with convolutional neural networks. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1725–1732. https://doi.org/10.1109/CVPR.2014.223

Kathuria, A. (2018a). How to implement a YOLO (v3) object detector from scratch in PyTorch: Part 1. Retrieved March 28, 2019, from https://blog.paperspace.com/how-to-implement-a-yolo-object-detector-in-pytorch/

Kathuria, A. (2018b). What's new in YOLO v3? Retrieved April 1, 2019, from https://towardsdatascience.com/yolo-v3-object-detection-53fb7d3bfe6b

Kirkpatrick, S., Gelatt, C. D., & Vecchi, M. P. (1983). Optimization by simulated annealing. *Science*. https://doi.org/10.1126/science.220.4598.671

Kirstein, S., Müller, K., Walecki-Mingers, M., & Deserno, T. M. (2012). Robust adaptive flow line detection in sewer pipes. *Automation in Construction*, *21*(1), 24–31.

https://doi.org/10.1016/j.autcon.2011.05.009

Koirala, A., Walsh, K. B., Wang, Z., & McCarthy, C. (2019). Deep learning for real-time fruit detection and orchard fruit load estimation: benchmarking of 'MangoYOLO.' *Precision Agriculture*, (0123456789). https://doi.org/10.1007/s11119-019-09642-0

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. In F. Pereira, C. J. C. Burges, L. Bottou, & K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems 25* (pp. 1097–1105). Retrieved from http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf

Kumar, S. S., Abraham, D. M., Jahanshahi, M. R., Iseley, T., & Starr, J. (2018). Automated defect classification in sewer closed circuit television inspections using deep convolutional neural networks. *Automation in Construction*, *91*(October 2017), 273–283. https://doi.org/10.1016/j.autcon.2018.03.028

Kurach, L., Zaman, H., & Chang, J. (2019). *Condition Assessment and Rehabilitation Programs at the City of Edmonton*. Edmonton, Alta.

Law, A. M. (2008). How to build valid and credible simulation models. *Proceedings - Winter Simulation Conference*, 39–47. https://doi.org/10.1109/WSC.2008.4736054

Lecchini-Visintini, A., Lygeros, J., & MacIejowski, J. M. (2010). Stochastic optimization on continuous domains with finite-time guarantees by markov chain monte carlo methods. *IEEE Transactions on Automatic Control*, *55*(12), 2858–2863. https://doi.org/10.1109/TAC.2010.2078170

Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Lee, J., Yi, J. S., & Son, J. (2019). Development of Automatic-Extraction Model of Poisonous Clauses in International Construction Contracts Using Rule-Based NLP. *Journal of Computing in Civil Engineering*. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000807

Leung, Y., Mei, C. L., & Zhang, W. X. (2000). Testing for spatial autocorrelation among the residuals of the geographically weighted regression. *Environment and Planning A*. https://doi.org/10.1068/a32117

Li, D., Cong, A., & Guo, S. (2019). Sewer damage detection from imbalanced CCTV inspection data using deep convolutional neural networks with hierarchical classification. *Automation

in Construction, *101*(January), 199–208. https://doi.org/10.1016/j.autcon.2019.01.017

Liao, X., Li, Q., Yang, X., Zhang, W., & Li, W. (2008). Multiobjective optimization for crash safety design of vehicles using stepwise regression model. *Structural and Multidisciplinary Optimization*, *35*(6), 561–569. https://doi.org/10.1007/s00158-007-0163-x

Lin, P., Yuan, X. X., & Tovilla, E. (2019). Integrative modeling of performance deterioration and maintenance effectiveness for infrastructure assets with missing condition data. *Computer-Aided Civil and Infrastructure Engineering*, 677–695. https://doi.org/10.1111/mice.12452

Liu, H., Al-Hussein, M., & Lu, M. (2015). BIM-based integrated approach for detailed construction scheduling under resource constraints. *Automation in Construction*, *53*, 29–43. https://doi.org/10.1016/j.autcon.2015.03.008

Liu, L., Ho, Y. kin, & Yau, S. (2006). Clustering DNA sequences by feature vectors. *Molecular Phylogenetics and Evolution*, *41*(1), 64–69. https://doi.org/10.1016/j.ympev.2006.05.019

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016). SSD: Single shot multibox detector. *European Conference on Computer Vision*, *9905 LNCS*, 21–37. https://doi.org/10.1007/978-3-319-46448-0_2

Liu, Z., & Kleiner, Y. (2013). State of the art review of inspection technologies for condition assessment of water pipes. *Measurement: Journal of the International Measurement Confederation*, *46*(1), 1–15. https://doi.org/10.1016/j.measurement.2012.05.032

Lu, M. (2002). Enhancing Project Evaluation and Review Technique Simulation through Artificial Neural Network-based Input Modeling. *Journal of Construction Engineering and Management*, *128*(5), 438–445. https://doi.org/10.1061/(ASCE)0733-9364(2002)128:5(438)

Lubini, A. T., & Fuamba, M. (2011). Modeling of the deterioration timeline of sewer systems. *Canadian Journal of Civil Engineering*, *38*(12), 1381–1390. https://doi.org/10.1139/L11-103

Martínez-Rojas, M., Marín, N., & Vila, M. A. (2015). An Approach for the Automatic Classification of Work Descriptions in Construction Projects. *Computer-Aided Civil and Infrastructure Engineering*, *30*(12), 919–934. https://doi.org/10.1111/mice.12179

Martínez-Rojas, M., Soto-Hidalgo, J. M., Marín, N., & Vila, M. A. (2018). Using Classification Techniques for Assigning Work Descriptions to Task Groups on the Basis of Construction Vocabulary. *Computer-Aided Civil and Infrastructure Engineering*, *0*, 1–16.

https://doi.org/10.1111/mice.12382

Martinez, P., Al-Hussein, M., & Ahmad, R. (2019). A scientometric analysis and critical review of computer vision applications for construction. *Automation in Construction*, *107*(July). https://doi.org/10.1016/j.autcon.2019.102947

Mashford, J., Rahilly, M., & Davis, P. (2008). An approach using mathematical morphology and support vector machines to detect features in pipe images. *Proceedings - Digital Image Computing: Techniques and Applications, DICTA 2008*, 84–89. https://doi.org/10.1109/DICTA.2008.25

Mashford, J., Rahilly, M., Davis, P., & Burn, S. (2010). A morphological approach to pipe image interpretation based on segmentation by support vector machine. *Automation in Construction*, *19*(7), 875–883. https://doi.org/10.1016/j.autcon.2010.06.001

McDonald, S. E., & Zhao, J. Q. (2001). Condition assessment and rehabilitation of large sewers. *International Conference on Underground Infrastructure Research*, 361–369. Retrieved from www.nrc.ca/irc/ircpubs

Meeker, E. (1971). The improving health of the United States, 1850-1915. *Explorations in Economic History*. https://doi.org/10.1016/0014-4983(71)90066-0

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*. https://doi.org/10.1063/1.1699114

Micevski, T., Kuczera, G., & Coombes, P. (2002). Markov Model for Storm Water Pipe Deterioration. *Journal of Infrastructure Systems*, *8*(2), 49–56. https://doi.org/10.1061/(ASCE)1076-0342(2002)8:2(49)

Mirza, S. (2007). *Danger Ahead: The coming collapse of Canada's municipal infrastructure, A Report for the Federation of Canadian Municipalities*. Retrieved from https://data.fcm.ca/documents/reports/Danger_Ahead_The_coming_collapse_of_Canadas_municipal_infrastructure_EN.pdf

Moselhi, O., & Shehab-Eldeen, T. (1999). Automated detection of surface defects in water and sewer pipes. *Automation in Construction*, *8*(5), 581–588. https://doi.org/10.1016/S0926-5805(99)00007-2

Moselhi, O., & Shehab-Eldeen, T. (2000). Classification of defects in sewer pipes using neural networks. *Journal of Infrastructure Systems*, *6*(3), 97–104.

https://doi.org/10.1061/(ASCE)1076-0342(2000)6:3(97)

Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. *27th International Conference on Machine Learning (ICML-10)*, 807–814. Retrieved from https://www.cs.toronto.edu/~hinton/absps/reluICML.pdf

Najafi, M., & Kulandaivel, G. (2005). Pipeline condition prediction using neural network models. *Pipeline Division Specialty Conference 2005*, 767–781. Retrieved from https://doi.org/10.1061/40800(180)61

Nakazawa, J., Kishimoto, Y., Iida, Y., Chen, Y., Oya, Y., & Mikami, K. (2019). DeepCounter: Using Deep Learning to Count Garbage Bags. *2018 IEEE 24th International Conference on Embedded and Real-Time Computing Systems and Applications (RTCSA)*, 1–10. https://doi.org/10.1109/rtcsa.2018.00010

NASSCO. (2014). *Pipe Condition Assessment Using CCTV: Performance Specification Guideline*. Retrieved from https://www.nassco.org/sites/default/files/SPECIFICATION GUIDELINE - CCTV 15Dec2014-2018logo.pdf

NASSCO. (2015). *Pipeline assessment and certification program (PACP)* (version 7). Retrieved from https://www.nassco.org/manual-pacp

Navab-Kashani, R., Gay, L. F., & Bayat, A. (2015). Productivity Improvement of Sewer CCTV Inspection through Time Study and Route Optimization. *Journal of Construction Engineering and Management*, *141*(6), 04015009. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000976

Navab-Kashani, R., Gay, L. F., & Bayat, A. (2019). Experimental and Numerical Study on Production Rate in Sewer Mainline Video Inspection. *Journal of Pipeline Systems Engineering and Practice*, *10*(2), 04019011. https://doi.org/10.1061/(asce)ps.1949-1204.0000373

Neter, J., Wasserman, W., & Kutner, M. H. (1989). *Applied Linear Regression Models.* (Second Edi). Retrieved from https://www.jstor.org/stable/2531657?origin=crossref

Neubeck, A., & Gool, L. Van. (2006). Efficient Non-Maximum Suppression. *18th International Conference on Pattern Recognition (ICPR'06)*, 0–5. https://doi.org/10.1109/ICPR.2006.479

Nguyen, V., Gächter, S., Martinelli, A., Tomatis, N., & Siegwart, R. (2007). A comparison of line extraction algorithms using 2D range data for indoor mobile robotics. *Autonomous Robots*, *23*(2), 97–111. https://doi.org/10.1007/s10514-007-9034-y

Nutanong, S., Jacox, E. H., & Samet, H. (2011). An incremental Hausdorff distance calculation algorithm. *Proceedings of the VLDB Endowment*, *4*(8), 506–517. https://doi.org/10.14778/2002974.2002978

Paya, I., Yepes, V., González-Vidosa, F., & Hospitaler, A. (2008). Multiobjective optimization of concrete frames by simulated annealing. *Computer-Aided Civil and Infrastructure Engineering*. https://doi.org/10.1111/j.1467-8667.2008.00561.x

Pelleg, D., & Moore, A. (2015). X-means: Extending K-means with Efficient Estimation of the Number of Clusters. *CEUR Workshop Proceedings*, *1542*, 33–36. https://doi.org/10.1017/CBO9781107415324.004

Raguram, R., Frahm, J.-M., & Pollefeys, M. (2008). A Comparative Analysis of RANSAC Techniques Leading to Adaptive Real-Time Random Sample Consensus. In *European Conference on Computer Vision* (pp. 500–513). https://doi.org/10.1007/978-3-540-88688-4_37

Redmon, J., Divvala, S., Girshick, R., & Ali, F. (2016). You Only Look Once: Unified, Real-Time Object Detection. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 779–788. https://doi.org/10.1109/CVPR.2016.91

Redmon, J., & Farhadi, A. (2017). YOLO9000: Better, Faster, Stronger. *CVPR*, 7626–7271. https://doi.org/10.1109/CVPR.2017.690

Redmon, J., & Farhadi, A. (2018). YOLOv3: An Incremental Improvement. *ArXiv Preprint ArXiv:1804.02767*. https://doi.org/10.1109/CVPR.2017.690

Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(6), 1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031

Roberts, D., & Golparvar-Fard, M. (2019). End-to-end vision-based detection, tracking and activity analysis of earthmoving equipment filmed at ground level. *Automation in Construction*, *105*, 102811. https://doi.org/10.1016/j.autcon.2019.04.006

Robertson, N., & Perera, T. (2002). Automated data collection for simulation? *Simulation Practice and Theory*, *9*(6–8), 349–364. https://doi.org/10.1016/S0928-4869(01)00055-6

Robinson, S. (2013). Conceptual modeling for simulation. *Proceedings of the 2013 Winter Simulation Conference - Simulation: Making Decisions in a Complex World, WSC 2013*, 377–388. https://doi.org/10.1109/WSC.2013.6721435

Rocha, R., Campilho, A., Silva, J., Azevedo, E., & Santos, R. (2011). Segmentation of ultrasound images of the carotid using RANSAC and cubic splines. *Computer Methods and Programs in Biomedicine*, *101*(1), 94–106. https://doi.org/10.1016/j.cmpb.2010.04.015

Rothe, R., Guillaumin, M., & Gool, L. Van. (2015). Non-maximum Suppression for Object Detection by Passing Messages Between Windows. In *ACCV 2014* (pp. 290–306). https://doi.org/10.1007/978-3-319-16865-4_19

Safari, S., & Aliyari Shoorehdeli, M. (2018). Detection and Isolation of Interior Defects Based on Image Processing and Neural Networks: HDPE Pipeline Case Study. *Journal of Pipeline Systems Engineering and Practice*, *9*(2), 05018001. https://doi.org/10.1061/(asce)ps.1949-1204.0000311

Sasaki, Y. (2007). The truth of the F-measure. *Teach Tutor Mater*.

Shehab, T., & Moselhi, O. (2005). Automated detection and classification of infiltration in sewer pipes. *Journal of Infrastructure Systems*, *11*(3), 165–171. https://doi.org/10.1061/(ASCE)1076-0342(2005)11:3(165)

Shen, Y., Pedrycz, W., Chen, Y., Wang, X., Gacek, A., & Member, S. (2019). Hyperplane Division in Fuzzy C-Means : Clustering Big Data. *IEEE Transactions on Fuzzy Systems*, *PP*(c), 1. https://doi.org/10.1109/TFUZZ.2019.2947231

Siddique, N., & Adeli, H. (2016). Simulated Annealing, Its Variants and Engineering Applications. *International Journal on Artificial Intelligence Tools*. https://doi.org/10.1142/S0218213016300015

Simonyan, K., & Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. *ICLR 2015*, 1–14. Retrieved from http://arxiv.org/abs/1409.1556

Sinha, S. K., & Fieguth, P. W. (2006a). Automated detection of cracks in buried concrete pipe images. *Automation in Construction*, *15*(1), 58–72. https://doi.org/10.1016/j.autcon.2005.02.006

Sinha, S. K., & Fieguth, P. W. (2006b). Neuro-fuzzy network for the classification of buried pipe defects. *Automation in Construction*, *15*(1), 73–83. https://doi.org/10.1016/j.autcon.2005.02.005

Sinha, S. K., Fieguth, P. W., & Polak, M. A. (2003). Computer vision techniques for automatic structural assessment of underground pipes. *Computer-Aided Civil and Infrastructure Engineering*, *18*(2), 95–112. https://doi.org/10.1111/1467-8667.00302

Sinha, S. K., & Karray, F. (2002). Classification of underground pipe scanned images using feature extraction and neuro-fuzzy algorithm. *IEEE Transactions on Neural Networks*, *13*(2), 393–401. https://doi.org/10.1109/72.991425

Sinha, S. K., & Knight, M. A. (2004). Intelligent system for condition monitoring of underground pipelines. *Computer-Aided Civil and Infrastructure Engineering*, *19*(1), 42–53. https://doi.org/10.1111/j.1467-8667.2004.00336.x

Son, H., Seong, H., Choi, H., & Kim, C. (2019). Real-Time Vision-Based Warning System for Prevention of Collisions between Workers and Heavy Equipment. *Journal of Computing in Civil Engineering*, *33*(5), 1–14. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000845

Sousa, V., Matos, J. P., & Matias, N. (2014). Evaluation of artificial intelligence tool performance and uncertainty for predicting sewer structural condition. *Automation in Construction*, *44*, 84–91. https://doi.org/10.1016/j.autcon.2014.04.004

Stehman, S. V. (1997). Selecting and interpreting measures of thematic classification accuracy. *Remote Sensing of Environment*, *62*(1), 77–89. https://doi.org/10.1016/S0034-4257(97)00083-7

Syachrani, S., Jeong, H. S. (David), & Chung, C. S. (2011). Dynamic Deterioration Models for Sewer Pipe Network. *Journal of Pipeline Systems Engineering and Practice*, *2*(4), 123–131. https://doi.org/10.1061/(asce)ps.1949-1204.0000085

Syachrani, S., Jeong, H. S., & Chung, C. S. (2013). Decision tree-based deterioration model for buried wastewater pipelines. *Journal of Performance of Constructed Facilities*, *27*(5), 633–645. https://doi.org/10.1061/(ASCE)CF.1943-5509.0000349

Taghaddos, H., Hermann, U., Abourizk, S., & Mohamed, Y. (2014). Simulation-based multiagent approach for scheduling modular construction. *Journal of Computing in Civil Engineering*, *28*(2), 263–274. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000262

Tian, Y., Yang, G., Wang, Z., Wang, H., Li, E., & Liang, Z. (2019). Apple detection during different growth stages in orchards using the improved YOLO-V3 model. *Computers and Electronics in Agriculture*, *157*(January), 417–426. https://doi.org/10.1016/j.compag.2019.01.012

Tran, D. H., Ng, A. W. M., McManus, K. J., & Burn, S. (2008). Prediction models for serviceability deterioration of stormwater pipes. *Structure and Infrastructure Engineering*, *4*(4), 287–295. https://doi.org/10.1080/15732470600792236

Tran, D. H., Ng, A. W. M., Perera, B. J. C., Burn, S., & Davis, P. (2006). Application of probabilistic neural networks in modelling structural deterioration of stormwater pipes. *Urban Water Journal*, *3*(3), 175–184. https://doi.org/10.1080/15730620600961684

Trybula, W. J. (1994). Building simulation models without data. *Proceedings of IEEE International Conference on Systems, Man and Cybernetics*. https://doi.org/10.1109/ICSMC.1994.399838

Voulodimos, A., Doulamis, N., Doulamis, A., & Protopapadakis, E. (2018). Deep Learning for Computer Vision: A Brief Review. *Computational Intelligence and Neuroscience*, *2018*. https://doi.org/10.1155/2018/7068349

Vu, D. H., Muttaqi, K. M., & Agalgaonkar, A. P. (2015). A variance inflation factor and backward elimination based robust regression model for forecasting monthly electricity demand using climatic variables. *Applied Energy*, *140*, 385–394. https://doi.org/10.1016/J.APENERGY.2014.12.011

Wagstaff, K., Cardie, C., Rogers, S., & Schrödl, S. (2001). Constrained K-means Clustering with Background Knowledge. *International Conference on Machine Learning ICML*, *pages*, 577–584. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.90.4624&rep=rep1&type=pdf

Wang, M., Wright, J., Buswell, R., & Brownlee, A. (2013). A Comparison Of Approaches To Stepwise Regression For Global Sensitivity Analysis Used With Evolutionary Optimization. *13th Conference of International Building Performance Simulation Association*, 2551–2558.

Water Research Centre (WRc). (2013). *Manual of Sewer Condition Classification 5th Edition (MSCC)*. WRc.

Wirahadikusumah, R., Abraham, D., & Iseley, T. (2001). Challenging issues in modeling deterioration of combined sewers. *Journal of Infrastructure Systems*, *7*(2), 77–84. https://doi.org/10.1061/(ASCE)1076-0342(2001)7:2(77)

Wirahadikusumah, Reini, Abraham, D. M., Iseley, T., & Prasanth, R. K. (1998). Assessment technologies for sewer system rehabilitation. *Automation in Construction*, *7*(4), 259–270. https://doi.org/10.1016/S0926-5805(97)00071-X

Wu, L., Ji, W., & AbouRizk, S. M. (2020). Bayesian Inference with Markov Chain Monte Carlo–Based Numerical Approach for Input Model Updating. *Journal of Computing in Civil

Engineering, *34*(1), 04019043. https://doi.org/10.1061/(asce)cp.1943-5487.0000862

Wu, W., Liu, Z., & He, Y. (2015). Classification of defects with ensemble methods in the automated visual inspection of sewer pipes. *Pattern Analysis and Applications*, *18*(2), 263–276. https://doi.org/10.1007/s10044-013-0355-5

Yang, M. Der, & Su, T. C. (2009). Segmenting ideal morphologies of sewer pipe defects on CCTV images for automated diagnosis. *Expert Systems with Applications*, *36*(2 PART 2), 3562–3573. https://doi.org/10.1016/j.eswa.2008.02.006

Yang, J., Gunaratne, M., Lu, J. J., & Dietrich, B. (2005). Use of Recurrent Markov Chains for Modeling the Crack Performance of Flexible Pavements. *Journal of Transportation Engineering*, *131*(11), 861–872. https://doi.org/10.1061/(asce)0733-947x(2005)131:11(861)

Yang, M. D., & Su, T. C. (2008). Automated diagnosis of sewer pipe defects based on machine learning approaches. *Expert Systems with Applications*, *35*(3), 1327–1337. https://doi.org/10.1016/j.eswa.2007.08.013

Yin, X., Chen, Y., Bouferguene, A., Zaman, H., Al-Hussein, M., & Kurach, L. (2019). A deep learning-based framework for an automated defect detection system for sewer pipes. *Automation in Construction*.

Yin, X., Chen, Y., Bouferguene, A., Zaman, H., Al-Hussein, M., & Russell, R. (2020a). Data-driven framework for modeling productivity of closed-circuit television recording process for sewer pipes. *Journal of Construction Engineering and Management*, *146*(8), 04020093. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001885

Yin, X., Chen, Y., Bouferguene, A., Zaman, H., Al-Hussein, M., Russell, R., & Kurach, L. (2019). Standard closed-circuit television (CCTV) collection time extraction of sewer pipes with machine learning algorithm. *Proceedings of the 36th International Symposium on Automation and Robotics in Construction, ISARC 2019*, (Isarc), 107–113. https://doi.org/10.22260/isarc2019/0015

Yin, X., Chen, Y., Bouferguene, A., & Al-Hussein, M. (2020b). Data-driven bi-level sewer pipe deterioration model: Design and analysis. *Automation in Construction*, *116*, 103181. https://doi.org/10.1016/j.autcon.2020.103181

Yin, X., Chen, Y., Bouferguene, A., Zaman, H., Al-Hussein, M., & Kurach, L. (2020c). A deep learning-based framework for an automated defect detection system for sewer pipes. *Automation in Construction*, *109*(August 2019), 102967.

https://doi.org/10.1016/j.autcon.2019.102967

Younis, R., & Knight, M. A. (2010). A probability model for investigating the trend of structural deterioration of wastewater pipelines. *Tunnelling and Underground Space Technology*, *25*(6), 670–680. https://doi.org/10.1016/j.tust.2010.05.007

Zabinsky, Z. B. (2011). Random Search Algorithms. In *Wiley Encyclopedia of Operations Research and Management Science* (pp. 335–364). https://doi.org/10.1002/9780470400531.eorms0704

Zaman, H. (2016). *A Framework for Improving the Productivity of Operational Preventive Maintenance Activities for Wastewater Collection System* (University of Alberta). https://doi.org/https://doi.org/10.7939/R38P5VJ0N

Zaman, H., Bouferguene, A., Al-Hussein, M., & Lorentz, C. (2015). Framework for Modeling On-Site Productivity of Preventive Maintenance Activities for Wastewater Collection Systems. *Journal of Infrastructure Systems*. https://doi.org/10.1061/(asce)is.1943-555x.0000252

Zaman, H., Bouferguene, A., Al-Hussein, M., & Lorentz, C. (2017). Improving the productivity of drainage operations activities through schedule optimisation. *Urban Water Journal*. https://doi.org/10.1080/1573062X.2015.1112409

Zayed, T. M., & Halpin, D. W. (2005). Pile construction productivity assessment. *Journal of Construction Engineering and Management*, *131*(6), 705–714. https://doi.org/10.1061/(ASCE)0733-9364(2005)131:6(705)

Zeferino, J. A., Antunes, A. P., & Cunha, M. C. (2009). An efficient simulated annealing algorithm for regional wastewater system planning. *Computer-Aided Civil and Infrastructure Engineering*. https://doi.org/10.1111/j.1467-8667.2009.00594.x

Zouaoui, F., & Wilson, J. R. (2004). Accounting for input-model and input-parameter uncertainties in simulation. *IIE Transactions (Institute of Industrial Engineers)*, *36*(11), 1135–1151. https://doi.org/10.1080/07408170490500708

APPENDIX A.

Table A-1:Summary of publications in selected journals regarding the topic of computer vision-based techniques for sewer pipe defect detection.

Time	Reference	Detection method	Title	Journal	Defect type
1999	Moselhi & Shehab-Eldeen, 1999	neural networks	Automated detection of surface defects in water and sewer pipes	Automation in Construction	Cracks, joint displacements
2000	Moselhi & Shehab-Eldeen, 2000	Back-propagation neural networks	Classification of defects in sewer pipes using neural networks	Journal of Infrastructure Systemsture system	Cracks, joint displacements, reduction of cross-sectional area, and spalling
2002	Sinha & Karray, 2002	Neuro-Fuzzy Algorithm	Classification of underground pipe scanned images using feature extraction and neuro-fuzzy algorithm	IEEE Transactions On Neural Networks	Crack and hole.
2003	Sinha & Karray, 2003	Neuro-fuzzy algorithm	Computer vision techniques for automatic structural assessment of underground pipes	Computer-Aided Civil and Infrastructure Engineering	Crack
2005	Iyer & Sinha, 2005	Morphological approach	A robust approach for automatic detection and segmentation of cracks in underground pipeline images	Image and Vision Computing	Crack
2005	Shehab & Moselhi, 2005	Back-propagation neural network	Automated Detection and Classification of Infiltration in Sewer Pipes	Journal of Infrastructure system	Infiltration
2006	Iyer & Sinha, 2006	Morphological approach	Segmentation of Pipe Images for Crack Detection in Buried Sewers	Computer-Aided Civil and Infrastructure Engineering	Crack
2006	Sinha & Fieguth, 2006	Neuro-fuzzy network	Neuro-fuzzy network for the classification of buried pipe defects	Automation in Construction	cracks and holes, broken joints and laterals
2006	Sinha & Fieguth, 2006a	Statistical filters	Automated detection of cracks in buried concrete pipe images	Automation in Construction	Crack
2007	Duran et al., 2007	multilayer perceptron (MLP)	Automated Pipe Defect Detection and Categorization Using Camera/Laser-Based Profiler and Artificial Neural Network	IEEE Transactions On Automation Science And Engineering	holes, cracks, joint, and obstacle
2008	Yang & Su, 2008	Back-propagation neural network, radial basis network, support vector machine	Automated diagnosis of sewer pipe defects based on machine learning approaches	Expert Systems with Applications	open joint, crack, broken pipe, and fracture
2009	Guo, et al., 2009	Change detection approach	Automated defect detection for sewer pipeline inspection and condition assessment	Automation in Construction	cracks, corrosion areas, connections and joints

2009	Yang & Su, 2009	Radial basis network	Segmenting ideal morphologies of sewer pipe defects on CCTV images for automated diagnosis	Expert Systems with Applications	Broken, Crack, open joint
2010	Mashford et al., 2010	Morphological approach	A morphological approach to pipe image interpretation based on segmentation by support vector machine	Automation in Construction	hole, corrosion, pipe connection, deposit and tree root
2012	Kirstein et al., 2012	Canny edge detection, Hough transform for straight lines and cost minimization using Dijkstra's shortest path algorithm	Robust adaptive flow line detection in sewer pipes	Automation in Construction	flow line
2015	Wu et al., 2015	Adaptive boosting, Random forest, Rotation forest, RotBoost	Classification of defects with ensemble methods in the automated visual inspection of sewer pipes	Pattern Analysis and Applications	crack, collapse, tree root penetration
2018	Safari et al., 2018	multilayer perceptron (MLP), curve completion method	Detection and Isolation of Interior Defects Based on Image Processing and Neural Networks: HDPE Pipeline Case Study	Journal of Pipeline Systems Engineering and Practice	short crack, medium crack, long crack, small perforation, medium perforation, big perforation
2018	Kumar et al., 2018	convolutional neural networks	Automated defect classification in sewer closed circuit television inspections using deep convolutional neural networks	Automation in Construction	root intrusions, deposits, cracks, infiltration, debris, connections, material change
2018	Cheng & Wang, 2018	Faster region-based CNN	Automated detection of sewer pipe defects in closed-circuit television images using deep learning techniques	Automation in Construction	tree root intrusion, deposit, infiltration and crack.
2019	Li et al., 2019	CNN, hierarchical classification	Sewer damage detection from imbalanced CCTV inspection data using deep convolutional neural networks with hierarchical classification	Automation in Construction journal	deposits settlement, joint offset, broken, obstacles, water level stag, deformation, etc.

Table A-2:Detailed structure of the YOLO network.

layer	Name	filters	size	input	output
0	conv	32	3x3/1	416x416x3	416x416x32
1	conv	64	3x3/2	416x416x32	208x208x64
2	conv	32	1x1/1	208x208x64	208x208x32
3	conv	64	3x3/1	208x208x32	208x208x64
4	res 1			208x208x64	208x208x64
5	conv	128	3x3/2	208x208x64	104x104x128
6	conv	64	1x1/1	104x104x128	104x104x64
7	conv	128	3x3/1	104x104x64	104x104x128
8	res 5			104x104x128	104x104x128
9	conv	64	1x1/1	104x104x128	104x104x64
10	conv	128	3x3/1	104x104x64	104x104x128

11	res 8			104x104x128	104x104x128
12	conv	256	3x3/2	104x104x128	52x52x256
13	conv	128	1x1/1	52x52x256	52x52x128
14	conv	256	3x3/1	52x52x128	52x52x256
15	res 12			52x52x256	52x52x256
16	conv	128	1x1/1	52x52x256	52x52x128
17	conv	256	3x3/1	52x52x128	52x52x256
18	res 15			52x52x256	52x52x256
19	conv	128	1x1/1	52x52x256	52x52x128
20	conv	256	3x3/1	52x52x128	52x52x256
21	res 18			52x52x256	52x52x256
22	conv	128	1x1/1	52x52x256	52x52x128
23	conv	256	3x3/1	52x52x128	52x52x256
24	res 21			52x52x256	52x52x256
25	conv	128	1x1/1	52x52x256	52x52x128
26	conv	256	3x3/1	52x52x128	52x52x256
27	res 24			52x52x256	52x52x256
28	conv	128	1x1/1	52x52x256	52x52x128
29	conv	256	3x3/1	52x52x128	52x52x256
30	res 27			52x52x256	52x52x256
31	conv	128	1x1/1	52x52x256	52x52x128
32	conv	256	3x3/1	52x52x128	52x52x256
33	res 30			52x52x256	52x52x256
34	conv	128	1x1/1	52x52x256	52x52x128
35	conv	256	3x3/1	52x52x128	52x52x256
36	res 33			52x52x256	52x52x256
37	conv	512	3x3/2	52x52x256	26x26x512
38	conv	256	1x1/1	26x26x512	26x26x256
39	conv	512	3x3/1	26x26x256	26x26x512
40	res 37			26x26x512	26x26x512
41	conv	256	1x1/1	26x26x512	26x26x256
42	conv	512	3x3/1	26x26x256	26x26x512
43	res 40			26x26x512	26x26x512
44	conv	256	1x1/1	26x26x512	26x26x256
45	conv	512	3x3/1	26x26x256	26x26x512
46	res 43			26x26x512	26x26x512
47	conv	256	1x1/1	26x26x512	26x26x256
48	conv	512	3x3/1	26x26x256	26x26x512
49	res 46			26x26x512	26x26x512
50	conv	256	1x1/1	26x26x512	26x26x256
51	conv	512	3x3/1	26x26x256	26x26x512
52	res 49			26x26x512	26x26x512
53	conv	256	1x1/1	26x26x512	26x26x256

54	conv	512	3x3/1	26x26x256	26x26x512
55	res 52			26x26x512	26x26x512
56	conv	256	1x1/1	26x26x512	26x26x256
57	conv	512	3x3/1	26x26x256	26x26x512
58	res 55			26x26x512	26x26x512
59	conv	256	1x1/1	26x26x512	26x26x256
60	conv	512	3x3/1	26x26x256	26x26x512
61	res 58			26x26x512	26x26x512
62	conv	1024	3x3/2	26x26x512	13x13x1024
63	conv	512	1x1/1	13x13x1024	13x13x512
64	conv	1024	3x3/1	13x13x512	13x13x1024
65	res 62			13x13x1024	13x13x1024
66	conv	512	1x1/1	13x13x1024	13x13x512
67	conv	1024	3x3/1	13x13x512	13x13x1024
68	res 65			13x13x1024->	13x13x1024
69	conv	512	1x1/1	13x13x1024->	13x13x512
70	conv	1024	3x3/1	13x13x512	13x13x1024
71	res 68			13x13x1024->	13x13x1024
72	conv	512	1x1/1	13x13x1024->	13x13x512
73	conv	1024	3x3/1	13x13x512	13x13x1024
74	res 71			13x13x1024->	13x13x1024
75	conv	512	1x1/1	13x13x1024->	13x13x512
76	conv	1024	3x3/1	13x13x512	13x13x1024
77	conv	512	1x1/1	13x13x1024->	13x13x512
78	conv	1024	3x3/1	13x13x512	13x13x1024
79	conv	512	1x1/1	13x13x1024->	13x13x512
80	conv	1024	3x3/1	13x13x512	13x13x1024
81	conv	36	1x1/1	13x13x1024->	13x13x36
82	detection				
83	route 79				
84	conv	256	1x1/1	13x13x512	13x13x
85	upsample		2x	13x13x256	26x26x256
86	route 85 61				
87	conv	256	1x1/1	26x26x768	26x26x256
88	conv	512	3x3/1	26x26x256	26x26x512
89	conv	256	1x1/1	26x26x512	26x26x256
90	conv	512	3x3/1	26x26x256	26x26x512
91	conv	256	1x1/1	26x26x512	26x26x256
92	conv	512	3x3/1	26x26x256	26x26x512
93	conv	36	1x1/1	26x26x512	26x26x36
94	detection				
95	route 91				
96	conv	128	1x1/1	26x26x256	26x26x128

97	upsample		2x	26x26x128	52x52x128
98	route 97 36				
99	conv	128	1x1/1	52x52x384	52x52x128
100	conv	256	3x3/1	52x52x128	52x52x256
101	conv	128	1x1/1	52x52x256	52x52x128
102	conv	256	3x3/1	52x52x128	52x52x256
103	conv	128	1x1/1	52x52x256	52x52x128
104	conv	256	3x3/1	52x52x128	52x52x256
105	conv	36	1x1/1	52x52x256	52x52x36
106	detection				