WEAPON CLASSIFICATION USING BOUNDING BOX REGRESSION ALGORITHMS

by

Kathan Trivedi

A project report submitted in conformity with the requirements
for the degree of Master of Science, Information Technology

Department of Mathematical and Physical Sciences
Faculty of Graduate Studies
Concordia University of Edmonton

WEAPON CLASSIFICATION USING BOUNDING BOX REGRESSION ALGORITHMS

KATHAN TRIVEDI

Approved:

Supervisor: Baidya Nath Saha, Ph.D. Date

Committee Member Date

Dean of Graduate Studies: Alison Yacyshyn, Ph.D. Date

WEAPON CLASSIFICATION USING BOUNDING BOX REGRESSION ALGORITHMS

Kathan Trivedi
Master of Science, Information Technology

Department of Mathematical and Physical Sciences
Concordia University of Edmonton
2022

Abstract

Crime rates are increasing at a very high rate globally. The use of weapons in schools, airports, and streets is gaining popularity. To prevent this we have a surveillance system that is monitored by security officers and requires constant observation. This is a tiresome job and involves human errors regardless of the significance of the problem. This research studies different machine learning algorithms for weapon detection and classification from images and videos. We have classified the data into four segments: (1) Knife (2) Pistol (3) Rifle (4) Grenade. We have analyzed supervised learning models consisting of around 5000 images with manual labeling of each image. Due to the lack of images for every class, the images are selected from the internet manually. This study involves the analysis of deep learning algorithms such as VGG16, VGG19, ResNet50, DenseNet201, and MobileNetV3Large to predict the potential threat from the given input and identify the particular weapon with a bounding box. The training of the model involves pre-processing of images and extraction of various features of the image. Extracted features are then passed through convolutional layers to predict the output. Different performance metrics are used to compare the performance of the deep learning detectors. The result of this study will help the government to establish an automated and reliable surveillance system.

Keywords:Weapon Classification, Crime Detection, Security, Weapon Data, Bounding Box Regression, Deep Learning, Supervised learning, Performance Metrics

Contents

1 Introduction 1

1.1 Background 1

1.2 Problem Statement 3

1.3 Contribution of the thesis 4

1.4 Organization of the thesis 5

2 Literature Review 6

2.1 Machine Learning Algorithms 6

2.2 Deep Learning Algorithms 7

2.2.1 Convolutional Neural Networks 7

2.2.2 Very Deep Networks 8

3 Weapon Detection 10

3.1 Deep Architectures 11

3.1.1 VGG16 and VGG19 11

3.2 Very Deep Architectures 12

3.2.1 ResNet50 12

3.2.2 DenseNet 12

3.3 Light-Weight Architecture 13

3.3.1 MobileNetV3 13

4 Data Collection and Pre-processing 15

4.1 Data Generation 15

4.1.1 Data Collection 16

4.1.2 Data Annotation 17

5 Results and Discussions 21

5.1 Qualitative Evaluation 21

5.1.1 Visualization of Weapon Detection and Classification 21

5.2 Quantitative Evaluation 21
5.2.1 Accuracy-Loss Over Epochs for Different Deep Learning Algorithms 21
5.2.2 Classification Reports 28
5.2.3 Precision-Recall Curve 28
5.2.4 ROC Curve 30
5.2.5 Confusion Matrix 31
5.2.6 Evaluation Metrics 31

6 Conclusion and Future Works 35

Bibliography 36

List of Tables

4.1 Dataset Description 16
4.2 .TXT File Format Description 17
4.3 COCO JSON annotation file description [28] 18

5.1 Weapon Classification Results for training dataset 28
5.2 Weapon Classification Results for validation dataset 28

List of Figures

1.1 Homicides rate from firearms, 2019 [1] 2
1.2 CCTV cameras workflow [2] 2

3.1 VGG16 Architecture [19] 11
3.2 ResNet Architecture [20] 12
3.3 DenseNet Architecture with growth rate of 4 [21] 13
3.4 MobileNet Architecture [22] 14

4.1 Four Classes for Weapon Classification 16
4.2 Original image with groundtruth 18
4.3 Pascal VOC format 19
4.4 JSON annotation file format 20

5.1 Prediction of Grenade Images 23
5.2 Prediction of Knife Images 24
5.3 Prediction of Pistol Images 25
5.4 Prediction of Rifle Images 26
5.5 Training-Validation accuracy over the epochs for VGG16, VGG19, ResNet50, DenseNet201, and MobileNetV3Large 27
5.6 Training-Validation loss over the epochs for VGG16, VGG19, ResNet50, DenseNet201, and MobileNetV3Large 27
5.7 Precision-Recall Curve for training and validation dataset of VGG16 . 28
5.8 Precision-Recall Curve for training and validation dataset of VGG19 . 29
5.9 Precision-Recall Curve for training and validation dataset of ResNet50 29
5.10 Precision-Recall Curve for training and validation dataset of DenseNet201 29
5.11 Precision-Recall Curve for training and validation dataset of MobileNetV3Large 29
5.12 ROC curves of VGG16 for training and validation dataset 30
5.13 ROC curves of VGG19 for training and validation dataset 30
5.14 ROC curves of ResNet50 for training and validation dataset 30

5.15 ROC curves of DenseNet201 for training and validation dataset . . . 31
5.16 ROC curves of MobileNetV3Large for training and validation dataset 31
5.17 Confusion matrix of VGG16 for training and validation dataset . . . 32
5.18 Confusion matrix of VGG19 for training and validation dataset . . . 32
5.19 Confusion matrix of ResNet50 for training and validation dataset . . 32
5.20 Confusion matrix of DenseNet201 for training and validation dataset 33
5.21 Confusion matrix of MobileNetV3Large for training and validation dataset 33
5.22 Dice score for deep learning algorithms 34
5.23 Jaccard coefficient for deep learning algorithms 34

Chapter 1

Introduction

1.1 Background

Homicide, a preventable and horrific event, has gained popularity in the last decade. The world has suffered a lot because of this nuisance. Moreover, society is suffering from events of murders at a very high rate. Human beings are influenced very negatively by this situation. Firearms are the major contributor to homicides globally. A current problem in regard to human rights is gun violence. The right to life, which is the most fundamental human right, is vanishing because of the extreme use of firearms in public places. People's right to education and right to healthcare facilities are being compromised as these deadly incidents can take place in schools, colleges, hospitals, and airports. People have started to feel unsafe in public places because of unwanted violence caused by firearms. Weapon violence is a major catastrophe that has an impact on people's lives all around the world. Every day, gun violence claims the lives of more than 500 people. The use of guns and knives for murders has started gaining popularity in many countries like the USA, Brazil, Mexico, Columbia, Afghanistan, Russia, and India. The figure 1.1 shows the comparison of the homicides rate worldwide caused by firearms in 2019.

For the security of human lives, the early detection of firearms is very essential. One potential solution to the problem can be video surveillance. In many public places, Closed Circuit Television (CCTV) cameras are actively tracking the moments of everyone. The CCTV cameras are mainly used for post-crime analysis as a piece of evidence. The typical workflow of the CCTV cameras is shown in figure 1.2. These cameras are usually present in public places such as shopping malls, schools, colleges, and airports. This surveillance system is usually monitored by one group of people. The monitoring can be error-prone as they are monitored by human eyes which involves human efforts to a very great extent. State-of-the-art surveillance

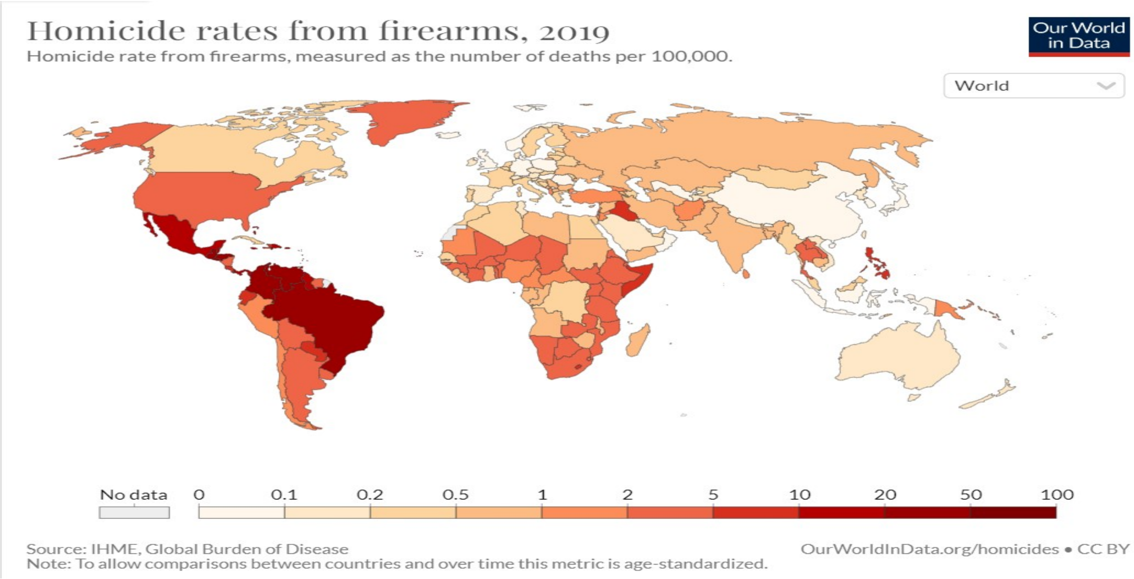

Homicide rates from firearms, 2019
Homicide rate from firearms, measured as the number of deaths per 100,000.
Our World in Data
World
No data
0
0.1
0.2
0.5
1
2
5
10
20
50
100
Source: IHME, Global Burden of Disease
Note: To allow comparisons between countries and over time this metric is age-standardized.
OurWorldInData.org/homicides • CC BY

Figure 1.1: Homicides rate from firearms, 2019 [1]

methods in high-security zone such as airports, and the embassy include manual screening and semi-automated image-based surveillance. Additionally, it requires constant monitoring which is not time effective solution for such a sensitive issue with a poor technique.

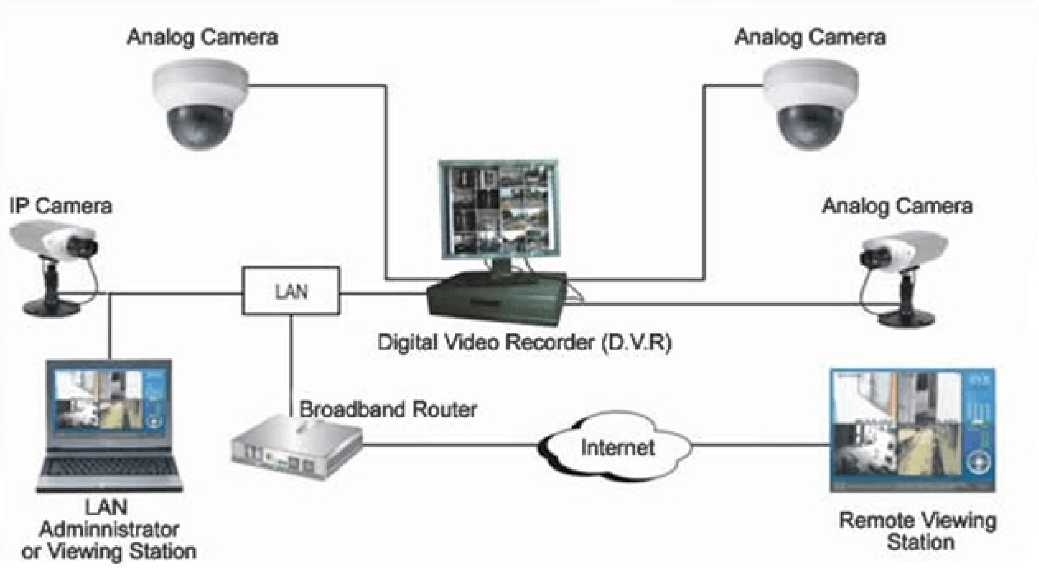


Figure 1.2: CCTV cameras workflow [2]

The solution to implementing CCTV is not much efficient to address this sensitive issue. Active surveillance cannot be achieved by manual monitoring. Implementation of any automated system can be used to eliminate the human efforts from this process. By automating this process, high efficiency in security surveillance can be achieved.

Elimination of tiresome and tedious human efforts can increase the accuracy of video surveillance which can be advantageous for human lives. We need an automated system that is highly configured and accurate to detect firearms from the photo or video which can be implemented in the existing CCTV system installed in all public places. Automating the process of firearms detection can also prevent any potential threat to society. The security staff can get alerted by the system if the system detects any threat at any given time in real-time. Thus, automation in firearms detection can help in preserving the fundamental rights of society. Algorithms developed using Machine Learning and Artificial Intelligence have become very popular to replace humans because of their long-term support, promising results, and high efficiency.

[3] The study and use of deep learning evolved from "artificial neural nets" in 1980. In the initial years of its development, deep learning was modeled as a human brain which consists of 100 billion neurons. Over the past few years, the number of real-world applications of deep learning has increased and developed handwriting recognition, language translation, automatic game playing, object detection, object classification, and many other fields.

Machine Learning (ML) algorithms, Artificial Intelligence, and Image Processing techniques are widely used in the automation industry [3]. There has been a dramatic use observed in recent years in many industries. Technology giants like Google, Microsoft, Facebook, and IBM have invested their resources heavily to work in this direction. All the industries have applied their hands to Deep Learning and have realized its potential. Deep learning algorithms are being applied in many sectors such as security, automation, healthcare, and education. Many organizations have developed and implemented machine learning concepts to detect weapons, firearms, and potentially threatening objects by using CCTV cameras or images. Some organizations such as Google, Facebook, Omnilert, ZeroEyes, and PatriotOne Technologies have used the concepts of ML algorithms for security surveillance.

1.2 Problem Statement

Deep learning algorithms have delivered some promising results in video analytics and video surveillance which are challenging problems for the image processing community. There are some areas for improvement and security systems still can be breached. So threat detection and weapon classification is still an open problem in the community which needs to be addressed and requires more research. Moreover, the elimination of False-Negative from the deep learning algorithms is highly challenging. Since the sensitivity of the matter is high, we need to develop an algorithm that is accurate and efficient.

We are using deep learning algorithms in our research for weapon detection and classification. Deep learning algorithms can be trained by learning from data. This research involves the collection of data and pre-processing them. In this research, we are working on image segmentation and image classification. By using deep learning algorithms, we can classify the weapon from the image and also identify the exact location of the weapon in the image itself. This study mainly detects four classes of weapons which are (1) Grenade (2) Knife (3) Pistol (4) Rifle. The primary objective of this research is to identify the appropriate algorithm for different places and environments. This research will study different deep learning algorithms, compare the experimental results and give a conclusion.

1.3 Contribution of the thesis

- The research work consists of a collection of data by scrapping web from resources such as GitHub, Kaggle, Data Science, and Computational Intelligence Institute (DaSCI). We are using supervised learning for the research. Supervised learning requires accurate labeling of the images. We have labeled each image manually to prepare them for different machine learning algorithms.

- The prepared dataset consists of different images of grenades, knives, pistols, and rifles. The dataset involves some handheld weapons, plain weapon images, and also toy weapons to increase its performance in threat detection.

- Deep learning algorithms are performed on a part of a dataset which is a training dataset and are used to train the model to detect and identify the object on a different dataset. This research studies six different algorithms for weapon detection and classification such as VGG16, VGG19, ResNet50, DenseNet, and MobileNetV3Large. At the end of the training, we compared the results of all the algorithms and analysis of the performance.

- For plotting the results and comparison of different algorithms, Receiver Operating Characteristics (ROC) Curve, Confusion Matrix, F1 Score, precision, accuracy, Jaccard score, and dice coefficient is calculated. The model is tested on the test dataset for classification and bounding box prediction.

1.4 Organization of the thesis

We have divided the research work into mainly six (6) chapters. Chapter 1 demonstrates the introduction of this study which includes background information, the existing homicides numbers globally, justification of why this research is important, the problem statement, and how we are planning to use the dataset and deep learning algorithms for weapon detection, and classification. Chapter 2 shows the Literature Review which describes and discusses different algorithms such as convolutional neural networks, deep convolutional neural networks, and lightweight convolutional neural networks. Chapter 3 describes the methodology used in this research and gives information about how deep learning algorithms are used in this subject. Chapter 4 is the Data Collection and Processing which gives information about how the data is collected for training and what changes are made to make the dataset ready for training. Chapter 5 talks about the results recorded by different deep learning algorithms. Chapter 6 concludes this research work and also depicts the future work involved in the research.

Chapter 2

Literature Review

Threat detection from images or videos using image processing algorithms is a very sensitive and challenging problem for the image processing community. Therefore, many researchers have been working on weapon detection and classification using machine learning algorithms and deep learning algorithms. Mainly the image classification process involves categorization and labeling groups of vectors or pixels present in an image based on specific rules [4]. In this research work, related work involves deep learning algorithms that fall under categories such as Convolutional Neural Networks, Very Deep Convolutional Networks (VGG16, VGG19), and Deep Residual Learning (ResNet50), Densely Connected Convolutional Networks (DenseNet), and Lightweight Machine Learning Algorithm (MobileNetV3). Additionally, this study also gives information about past work done for weapon classification using orthodox machine learning algorithms. This section explains different algorithms and the results achieved in previous work. All algorithms are suitable for different scenarios so it is essential to consider the previous experimental results in this research. Furthermore, we have analyzed the two approaches to addressing the image classification problem, one is through machine learning algorithms and another is deep learning algorithms.

2.1 Machine Learning Algorithms

Nadhir Ben Halima et al [5] proposed an automatic surveillance system for detecting firearms in a cluttered scene. Initially, SIFT features are calculated and extracted from the collection of images. After that, the K-means algorithm is used for clustering the features. Then, a histogram is used by counting the numbers of extracted clusters in every image. The histogram will be used as an input to the proposed SVM model. In the end, the SVM will generate the classification output whether the image contains a class or not. The proposed model gave high accuracy and achieved better results.

Descriptors are imposed on the image when a weapon is detected. Orange and yellow descriptors are plotted on the image for weapon detection. In this research work, we have categorized the deep learning algorithms into three subsections such as very deep networks, residual deep networks, dense networks, and mobile networks.

Tiwari and Verma [6] proposed a way for automatic surveillance by using a visual gun detection framework. In their research, they are using color-based segmentation and a harris interest point detector for the detection of guns in images. Because of the long processing time, their proposed approach is not suitable for real-time detection.

2.2 Deep Learning Algorithms

2.2.1 Convolutional Neural Networks

Fraol Gelana and Arvind Yadav [7] studied different machine learning algorithms for weapon identification from images by training 1869 weapon images and 4000 non-weapon images. They proposed a Convolutional Neural Network (CNN) approach to identify firearms from CCTV recordings. Their proposed model is primarily divided into six major phases. Firstly, they convert the RGB images to grayscale images to reduce the complexity of each frame and to subtract background information efficiently. For subtracting background information, three approaches are used the Visual Background Extractor [8], Improved Gaussian Mixture Model [9], and the difference of frame background subtraction algorithm. Due to high noise in the foreground object, they used Dilation and Erosion operations for removing noise elements. After that, they used Canny [10] for edge detection from the foreground object. After that, the sliding window approach was chosen for frame-by-frame evaluation for threatening object detection. Finally, in the last step classification is performed using the Tensorflow-based CNN algorithm. The proposed approach achieved 97.78% accuracy. However, due to heavy processing requirements, this model is only applicable for high-quality videos and also can detect only one class of firearms.

Muhammad Tahir Bhatti et al [11] manually collected the images of weapons and non-weapons from scrapping web and images captured by them. In their proposed work, they are using the pistol as a reference class for the binary classification approach. They compared and analyzed the results of the algorithms such as VGG16, Inception-V3, Inception-ResnetV2, SSDMobileNetV1, FRCNN Inception-ResnetV2, YOLOv3, and YOLOv4. They observed the best results are achieved using the YOLOv4 model. However, the models trained in this study could just predict one class of firearms and the number of False-Positive was very high.

Harsh Jain et al [12] collected the data on weapons and proposed two approaches for weapon classification. One of the approaches is to use Single Shot Detector (SSD), and another is to use Faster-RCNN (FRCNN). They prepared the CSV file from the XML annotation files. They are using different algorithms to address different parameters such as accuracy and speed. Their results suggest that for better accuracy is FRCNN is the ideal choice whereas in the cases where execution time should be less, SSD is the appropriate option.

Roberto et al [13] trained the FRCNN model on the pistol dataset which can detect the weapon and also identify the region from the image. They also developed an Alarm Activation Timer per Interval (AATpI) to assess the performance of a detection model in videos. This research proposed two methods for the problem. One approach is the Sliding Window approach, which divides the image into different windows and scans through each window, and identifies the weapon if it is present. For this, The Histogram of Oriented Gradients (HOG) [14] based model uses the HOG descriptor to predict the object. The other concept used in this study is region proposals. Choosing all the windows of the images for identification is tedious, instead of all, this method selects actual candidate regions by detection proposal methods. By using, FRCNN, features can be extracted from the images. The promising results are achieved using the FRCNN approach for this problem which is efficient enough for real-time detection also. The authors are still trying to reduce the False-Positive for the proposed approach.

Maddula et al [15] carried out their research for weapon and face detection from low-resolution images. They used CNN for this research. They proposed a novel convolutional neural network for face recognition and Haar cascade for face and weapon detection.

2.2.2 Very Deep Networks

Neelam Dwiwedi et al [16] collected the images from the internet and captured some of the images in their lab. They prepared the dataset of three classes such as knife, gun, and no-weapon. No-weapon class contains the images of human beings, cars, and chairs. Due to the small dataset, they shrank the fully connected layer to reduce the number of parameters. In their research, VGG16 architecture was used as a base model and they added three Fully Connected layers at the end with different neuron values. By changing the neuron values in three layers, Model A and Model B were generated for classification. By using the concept of transfer learning, both models were trained on the training dataset. After training, Model A got nearly 98% of accuracy which contains 1024, 512, and 3 neurons in FC1, FC2, and FC3 respectively. This research depicts that accuracy does not always increase by increasing the neurons.

Additionally, it was also concluded that average accuracy decreases with an increased dropout rate.

Volkan Kavya et al [17] prepared the dataset of seven classes which contains images of assault rifles, bazookas, grenades, hunting rifles, knives, pistols, and revolvers. The authors proposed a new model based on VGGNet architecture to classify weapon images. The author proposed a model with 25 layers. It contains 7 convolutional, 4 pooling, 4 dropouts, 7 ReLU, 1 flattened, 1 fully connected, and 1 classification layer. Finally, the comparison of a proposed approach with VGG-16, ResNet-50, and ResNet-101 is given. Experimentally results and output is compared and analyzed. This research work achieves high accuracy of 98.40%. The proposed study not only addresses the weapon classification but also the image segmentation. The classification of seven classes with a region proposal approach makes this model ideal for surveillance. However, there are some areas to be improved such as dataset distribution and real-life scenes of weapons.

Arif Warsi et al [18] categorized two main approaches for handgun and knife classification, one of them is Non-Deep Learning Algorithms, and the other Deep Learning Algorithms. Non-deep learning algorithms involve methods such as color-segmentation, interest points, shapes, and edge detectors. Whereas deep learning algorithms are FRCNN, SSD, CNN, YOLO, and many others as this field is growing very quickly. The authors compared and analyzed various algorithms and showed the results in their study to explain the differences. According to this study, high accuracy and high-speed results can be achieved using deep learning algorithms. In conclusion, non-deep learning algorithms are ideal for X-ray-based images where processing time does not matter much and images are already in binary format. Deep learning algorithms have delivered very good results in color-based images and it is recommended by authors.

Chapter 3

Weapon Detection

This research work is addressing two essential problems in the image processing community, one of them is classification and another is regression. In this research work, we have identified the weapon class and also the exact location of the weapon present in the image. A rectangle is plotted with the class name on the image. As this research is focused more on real-time detection from the images and videos, finding the exact location becomes extremely important in this field.

The classification problem is an open and challenging problem for decades. The classification process involves the prediction of the discrete class label and assigning a particular category to an image or a part of an image. It mainly divides the images or dataset into more than one category for identification. Classification algorithms calculate various parameters from the image to categorize the image into any particular class. This research work has implemented and analyzed various deep learning algorithms for the classification of images.

Regression is a challenging problem when there are many parameters to calculate in the dataset. It mainly finds the correlation between dependent and independent variables. It predicts continuous variables such as stock market value and price prediction. In this research, while training the model, we are also passing the bounding box details to generate and predict the object localization. By providing image and bounding box details we can get the exact location of the weapon present in the image. In this research work, we have implemented three major types of deep learning models such as deep architectures (VGG16 and VGG19), very deep architectures (ResNet50 and DenseNet), and lightweight architecture (MobileNetV3). The results of these architectures are used for comparison and evaluation of the different models for weapon classification and detection.

3.1 Deep Architectures

3.1.1 VGG16 and VGG19

[19] VGG16 is an advanced version of the Convolutional Neural Network (CNN) model. The founders of the VGG16 increased its depth by adding tiny (3*3) convolution filters which ultimately increased its depth and accuracy in predicting the output. It is still one of the best computer vision models for the image processing community. VGG16 contains 16 weight layers, whereas VGG19 [19] contains 19 weight layers. It is a deeper version of ConvNets. The authors of the paper explained that they increased the total trainable parameters to 138 parameters by increasing their depth. They observed a dramatic increase in the model's accuracy after increasing its depth. VGG16 can also be used for transfer learning. The fig. 3.1 shows the architecture of the VGG16. Typically, the VGG16 model contains 16 weight layers, five Max Pooling layers, and three Dense layers. It accepts the images of the RGB channel. The input to ConvNets is a fixed size of 224 * 224 with an RGB channel. In comparison with traditional CNN, VGG16 is trained for classification and localization both. There are some drawbacks to using VGG16, one of them is, that it is very slow for training. The trained model consumes a big size of the memory in the disk. An increasing number of trainable parameters ultimately leads to vanishing gradient problems in the deeper networks.

In this research, we have trained the VGG16 and VGG19 models for classification and regression purposes. We are passing the annotations and the images to the model and achieved good results for image classification and identification.

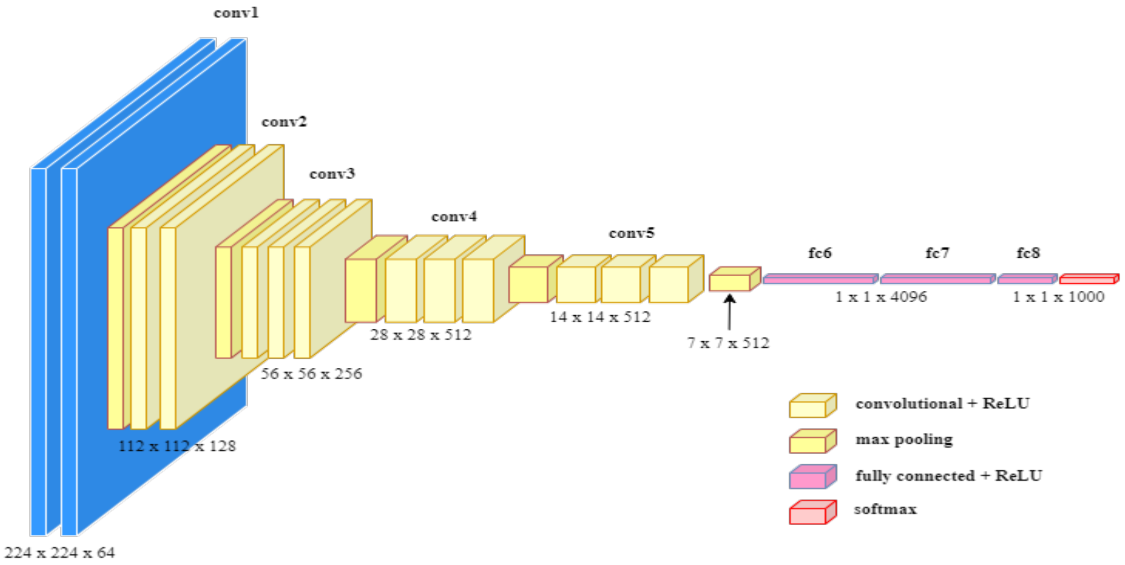


Figure 3.1: VGG16 Architecture [19]

3.2 Very Deep Architectures

3.2.1 ResNet50

[20] Resnet50 is trained on the ImageNet database which contains more than a million images. ResNet50 is a CNN model which has 50 layers in its architecture. ResNet50 was developed to overcome the difficulties raised in deeper neural networks like VGG16 and VGG19. ResNet50 is deeper than VGG16 and VGG19 but it eliminates the complexity of the model by using residual functions reference. With its efficiency in the image processing problem, it managed to secure 1st place in ILSVRC 2015. It accepts the input images with the size of 224*224. ResNet50 eliminates the vanishing gradient problem by using the skip connection technique. In the deeper neural networks, the vanishing gradient problem occurred and the training was not done properly. To avoid the training issues, and get better accuracy, the authors of ResNet-50 used the skip connection technique. The figure 3.2 shows the architecture of the ResNet50 model.

This research work consists of the implementation of the ResNet50 model on the weapon dataset. This research also involves the analysis of results achieved using ResNet50 compared to other deep learning algorithms.

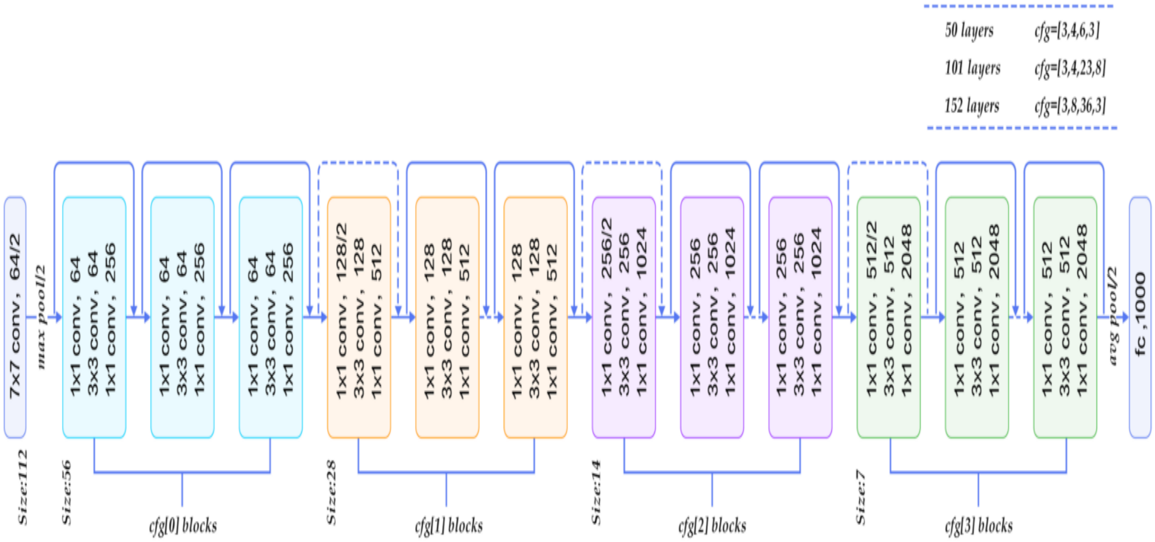


Figure 3.2: ResNet Architecture [20]

3.2.2 DenseNet

[21] The efficiency of the convolutional networks can be increased if the distance between layers close to input and output can be decreased. DenseNet is developed using the same concept. It is the advanced version of CNN. DenseNets have the primary

advantage of re-utilizing the features, eliminating of vanishing gradient problem, and reducing the parameters. DenseNet is trained on the ImageNet database. Because feature maps are sent to each layer from their corresponding preceding layers, the network can be thinner and compact. Because of this architecture, the computational power and accuracy of the model are higher. The figure 3.3 shows the architecture of the DenseNet model.

We have included the DenseNet algorithm in our research work for analysis. The DenseNet model is trained on the weapon dataset.

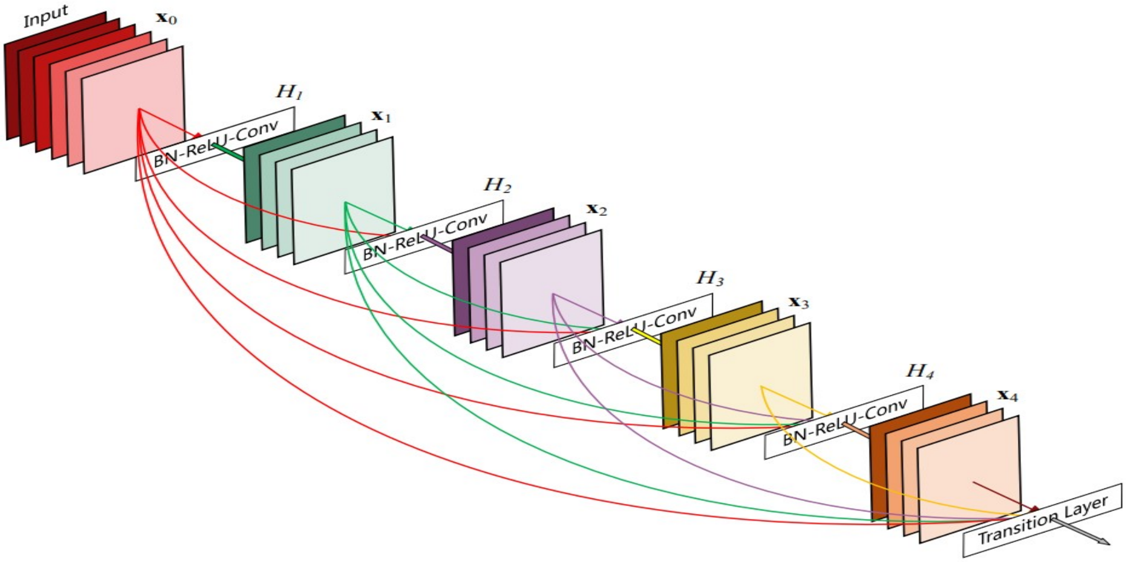


Figure 3.3: DenseNet Architecture with growth rate of 4 [21]

3.3 Light-Weight Architecture

3.3.1 MobileNetV3

[22] MobileNets are specifically developed for mobile devices and embedded vision applications. MobileNet is trained on the ImageNet database and uses the streamed-line architecture. The first layer of the MobileNet is the convolution and it is built on depth-wise separable convolutions. In modern object detection systems, MobileNet can also be deployed as an effective base network. The authors explained how to build smaller and faster MobileNets with the use of multiplier and resolution multiplier by trading off the accuracy. The figure 3.4 shows the architecture of the MobilenNet. This research work implemented the MobileNetV3Large on the weapon dataset.

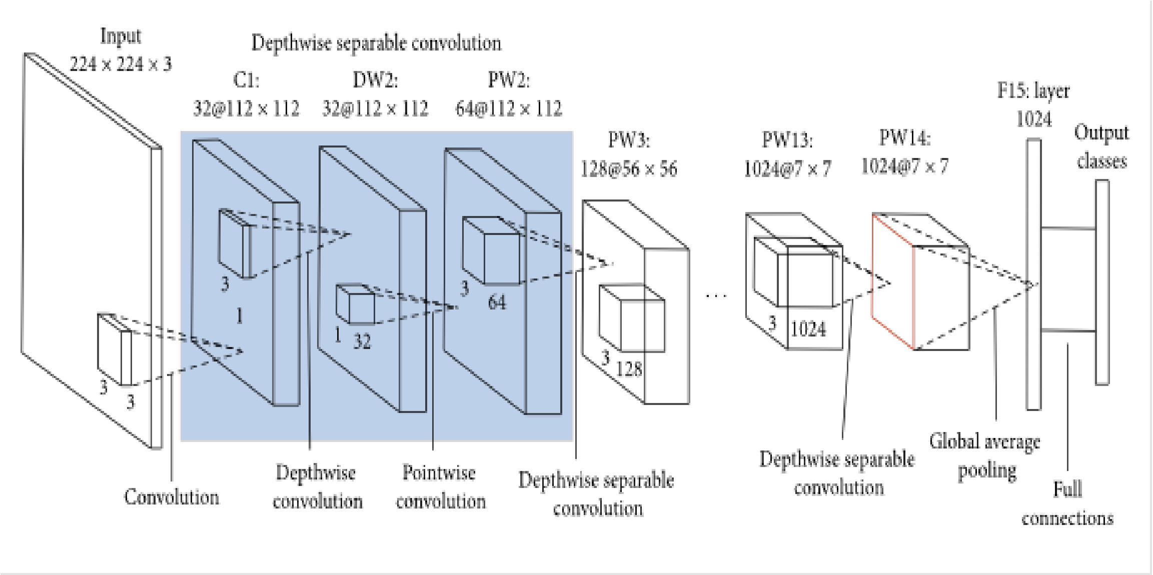


Figure 3.4: MobileNet Architecture [22]

Chapter 4

Data Collection and Pre-processing

4.1 Data Generation

In recent times, due to the installation of CCTV cameras, human activities can be tracked quickly and easily. Data collected from the CCTV footage involves illegal weapons use in public places. In this research, we have used the combination of the images from CCTV footage available on the internet and handheld firearms images. Therefore, the proposed model can be implemented to provide real-time results and detection of weapons. Additionally, data is collected using various web resources such as Kaggle, GitHub, google images, and DaSCI.es [23][24],

[25][26]Supervised learning is the concept that mainly requires human intervention in the early stage of preparing the data. It adjusts the correct classification value by constantly making predictions on training data. Using labeled inputs and outputs model can learn over the period and increase its accuracy. Supervised learning usually addresses two types of problems such as classification, and regression. In classification, the algorithms are used to categorize the test data and assign the images to a particular class. Whereas in regression, the predicted output is the relationship between dependent and independent variables. In contrast, unsupervised learning is used for analyzing and clustering the unlabelled data. Unsupervised learning typically involves Clustering, Association, and Dimensionality Reduction tasks.

In this research work, supervised learning is taken into consideration for training different models. For training our model, labeling of data is required. Data labeling is a tedious task and requires a lot of human effort. Getting the labeled data for this research work was not easy, so the data is labeled manually in this research work. For labeling the images, a bounding box around the particular area of the image needs to be plotted and given a class. While training, these parts of the images will be taken into consideration for training. Therefore, it is a tiring job and requires accuracy.

Otherwise, the output of the trained model might be inaccurate.

4.1.1 Data Collection

We collected the data from different sources and prepared the training and validation dataset. The table 4.1 gives information about the collected dataset and the number of images contained for training, validation, and testing. Figure 4.1 shows different class images of weapon dataset.

Table 4.1: Dataset Description

Classes	Class_ID	Train Data	Validation Data	Test Data
Grenade	0	1120	480	160
Knife	1	700	300	100
Pistol	2	1398	599	103
Rifle	3	866	371	123

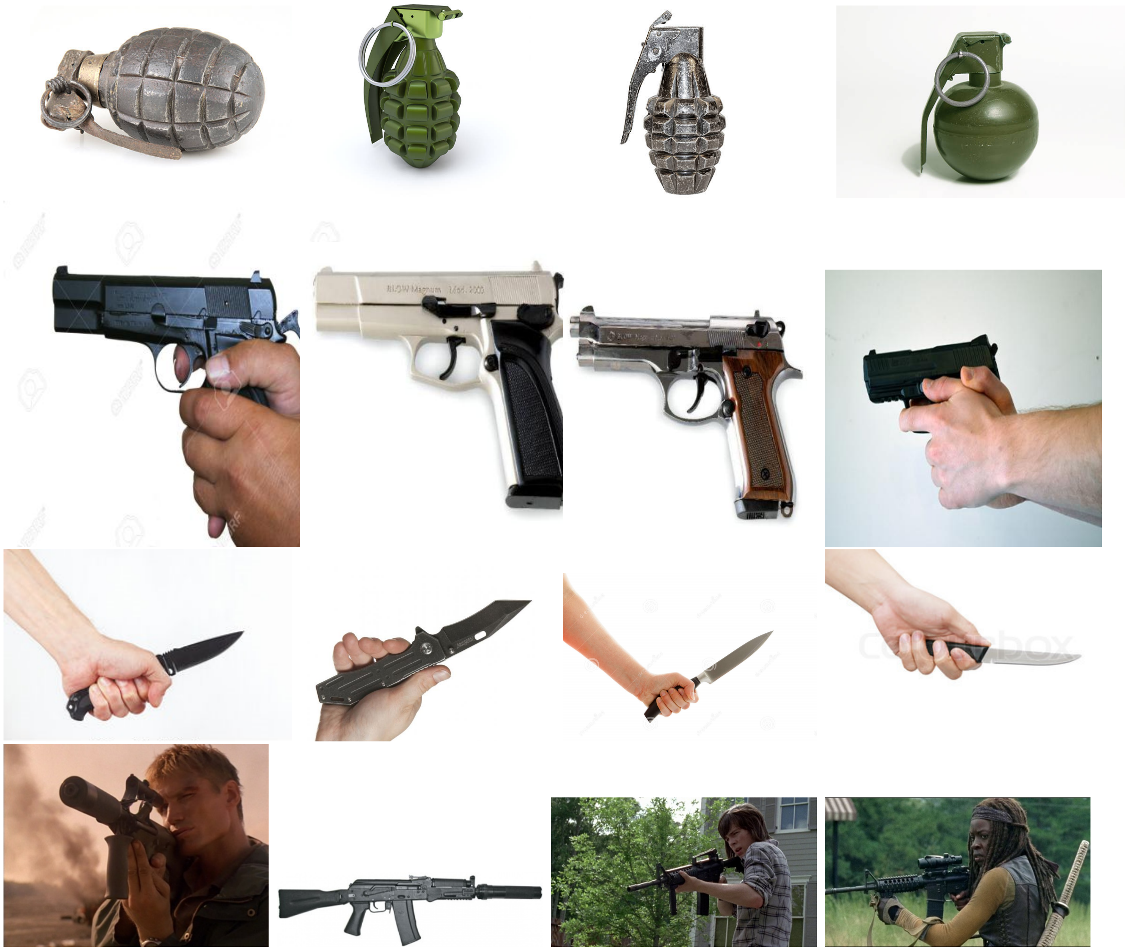


Figure 4.1: Four Classes for Weapon Classification

4.1.2 Data Annotation

For labeling our dataset, we used the labeling tool [27]. This tool offers various label formats for generating label files such as TXT format, JSON format, and PASCAL VOC (Visual Object Classes) format. The detailed architecture of the label files is explained in subsections. To label the image, one has to plot the rectangle around the specified area of the image and give the appropriate class. The primary goal to prepare the dataset with all the annotation formats is to train different machine learning algorithms efficiently because different algorithms use different annotation files.

TXT format

The filename of the label file is exactly as same as the image filename. This format contains different fields to store the annotation. The fields are separated by space in the .txt file. Table 4.2 lists the fields and description of every field.

Table 4.2: .TXT File Format Description

Field	Description
Class_ID	This field describes the class in which this particular area of the image belongs to
Start-X	X-axis start location of the rectangle
Start-Y	Y-axis start location of the rectangle
End-X	X-axis end location of the rectangle
End-Y	Y-axis end location of the rectangle

Therefore, from the labeled file, we can plot the rectangle around the particular area of the image.

PASCAL VOC format

The Pascal VOC format is nothing but an XML file that contains all the necessary information of the image such as folder, filename, bounding box information, and class name. Some of the deep learning algorithms use Pascal VOC format labels for training. Figure 4.3 gives an idea about the PASCAL VOC format annotation file and figure 4.2 shows the bounding box plotted on the corresponding image.

JSON format

The JSON file format is also a popular annotation file format used in supervised learning methodologies. JSON (JavaScript Object Notation) file contains the necessary information in a different way than the above-mentioned files. Figure 4.4 gives

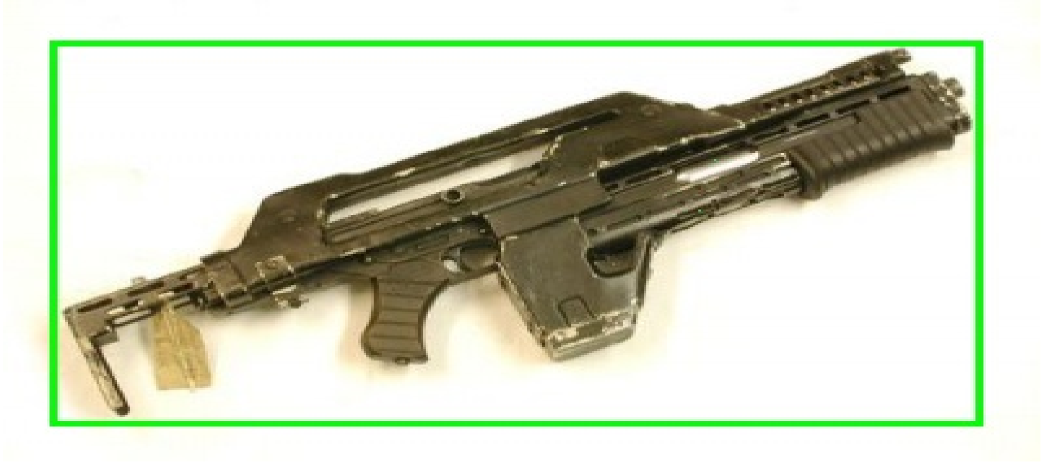


Figure 4.2: Original image with groundtruth

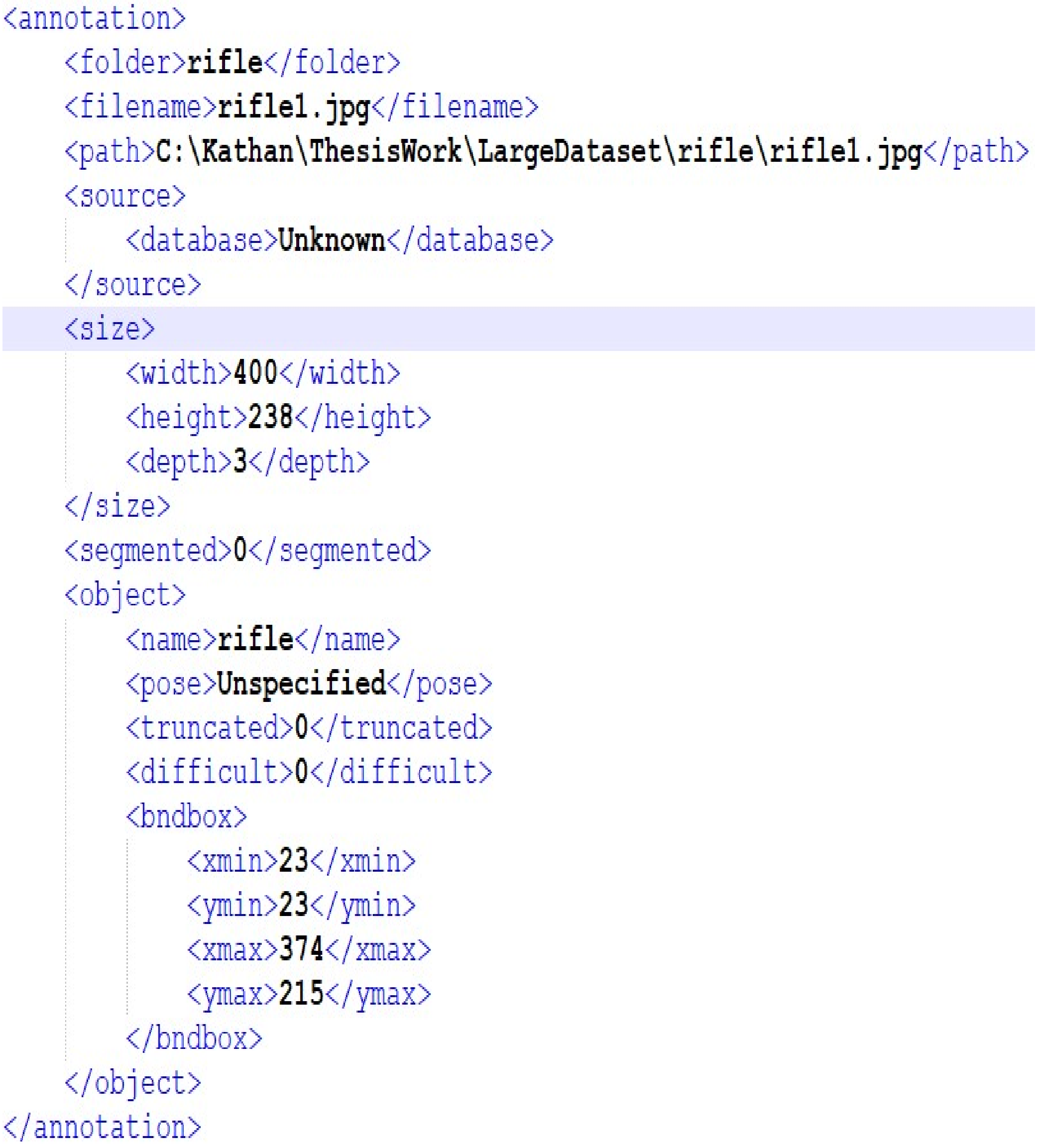
information about the JSON file format. Instead of X-axis and Y-axis end location, it takes the height and width of the rectangle.

COCO JSON format

Apart from the labeling tool, we used the Roboflow tool to convert the annotation files from PASCAL VOC XML to COCO JSON format. Machine learning algorithms such as YOLO (You Only Look Once) use the COCO JSON format for training. In the future, we are planning to work on the latest YOLO model to get the image segmentation and classification in real-time with high efficiency. There will be just one file created for COCO JSON format for all the images and it contains necessary details in just one file in different tags of JSON file. Table 4.3 gives the details of all the fields present in the COCO JSON format annotation file.

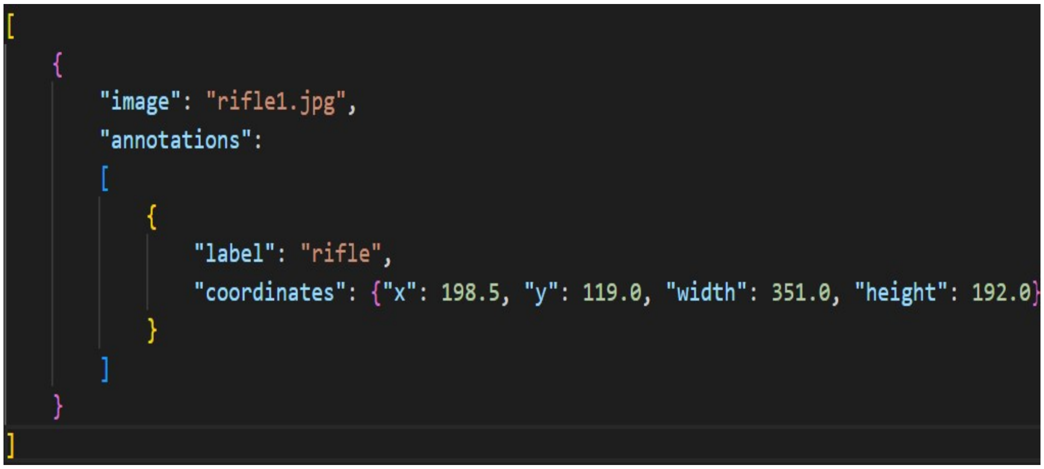
Table 4.3: COCO JSON annotation file description [28]

Field	Description
Info	It usually depicts the high level information of the prepared dataset
Licenses	It provides the list of licenses under which the images are licensed
Categories	Contains list of classes
Images	Stores the image information in the dataset with unique image id
Annotations	List of every individual object annotation from every image in the dataset



```
<annotation>
    <folder>rifle</folder>
    <filename>rifle1.jpg</filename>
    <path>C:\Kathan\ThesisWork\LargeDataset\rifle\rifle1.jpg</path>
    <source>
        <database>Unknown</database>
    </source>
    <size>
        <width>400</width>
        <height>238</height>
        <depth>3</depth>
    </size>
    <segmented>0</segmented>
    <object>
        <name>rifle</name>
        <pose>Unspecified</pose>
        <truncated>0</truncated>
        <difficult>0</difficult>
        <bndbox>
            <xmin>23</xmin>
            <ymin>23</ymin>
            <xmax>374</xmax>
            <ymax>215</ymax>
        </bndbox>
    </object>
</annotation>
```

Figure 4.3: Pascal VOC format



```
[
    {
        "image": "rifle1.jpg",
        "annotations":
        [
            {
                "label": "rifle",
                "coordinates": {"x": 198.5, "y": 119.0, "width": 351.0, "height": 192.0}
            }
        ]
    }
]
```

Figure 4.4: JSON annotation file format

CSV file format

CSV (Comma Separated Value) files are very common formats for annotations. Many machine learning algorithms read the ground-truth information from the CSV file format. For easy access and reading the information about the ground truth, in this research work, we have prepared the CSV file from the PASCAL VOC annotations. The CSV file was prepared using a Python script. We stored the necessary details in the CSV file extracted from an XML file. The CSV file contains filename, width, height, class name, X-axis start location, Y-axis start location, X-axis end location, and Y-axis end location.

Chapter 5

Results and Discussions

This chapter gives information about different results plotted by various deep learning algorithms. The results are divided into major subsections such as visualization of weapon detection, comparative analysis of classification results, and comparative analysis of detection results.

5.1 Qualitative Evaluation

5.1.1 Visualization of Weapon Detection and Classification

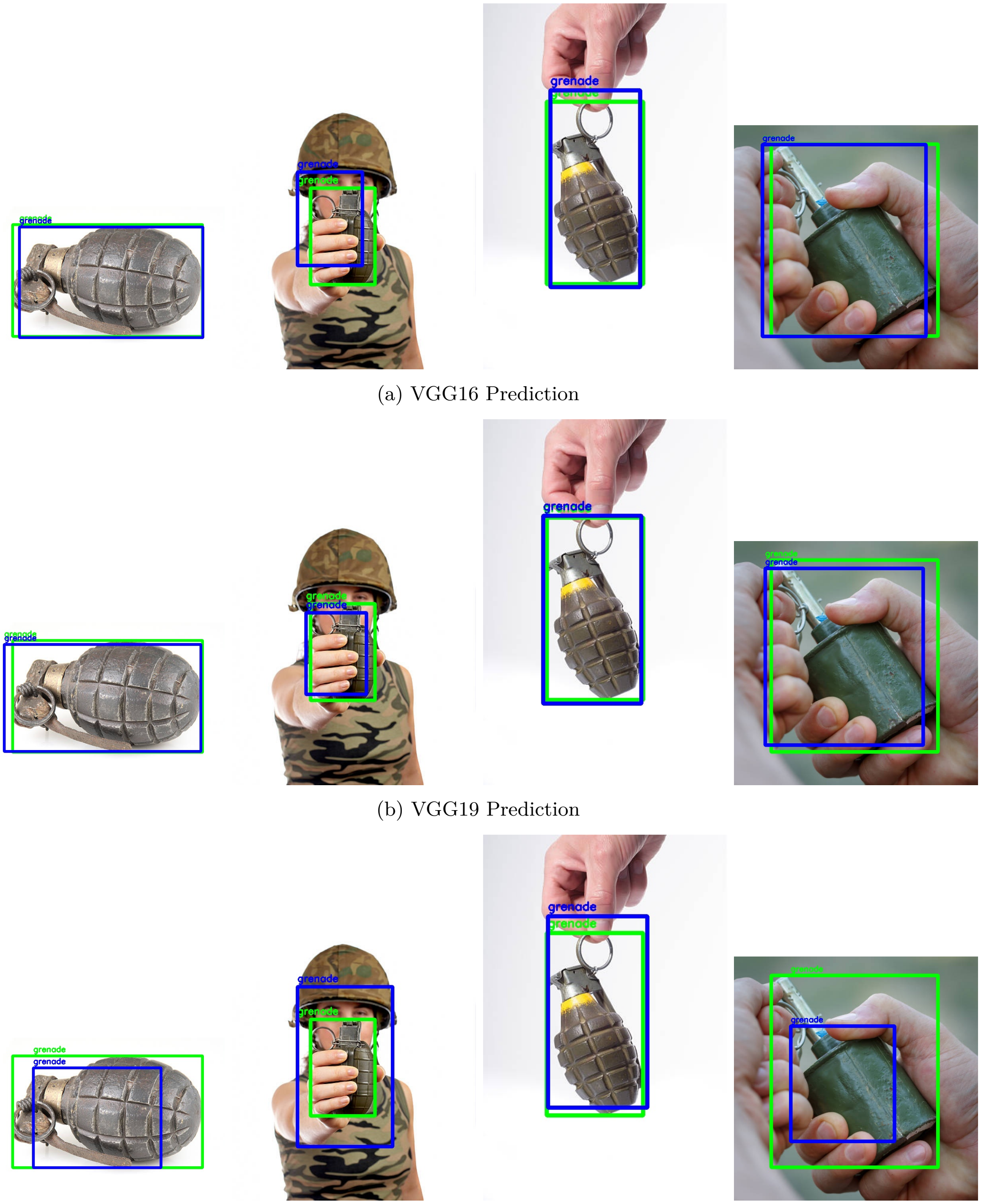
The figures 5.1, 5.2, 5.3, and 5.4 shows the predicted class and bounding box plotted on the image for VGG16, VGG19, ResNet50, DenseNet201, and MobileNetV3Large respectively. The results are plotted using the Matplot library. Bounding boxes plotted in green color depict the ground truth and rectangles in blue color show the actual prediction by the deep learning algorithm.

5.2 Quantitative Evaluation

Different algorithms are compared through several evaluation metrics which are demonstrated in this section.

5.2.1 Accuracy-Loss Over Epochs for Different Deep Learning Algorithms

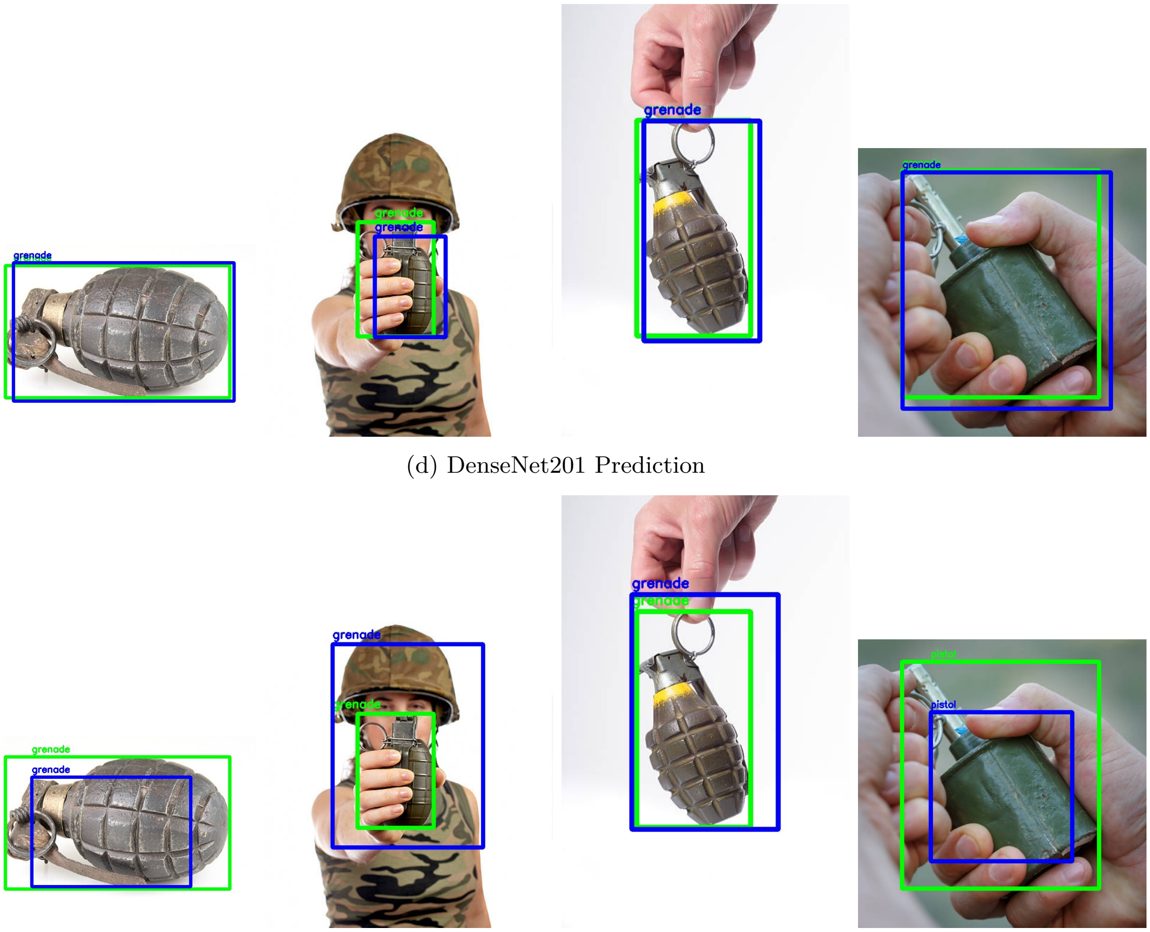
This section gives the visualization of loss and accuracy in graphical form over the epochs of every deep learning algorithm. The figure 5.5 shows the training and validation accuracy performance of deep learning algorithms in graphical form. The figure 5.6 shows the training and validation loss performance of deep learning algorithms.



(a) VGG16 Prediction

(b) VGG19 Prediction

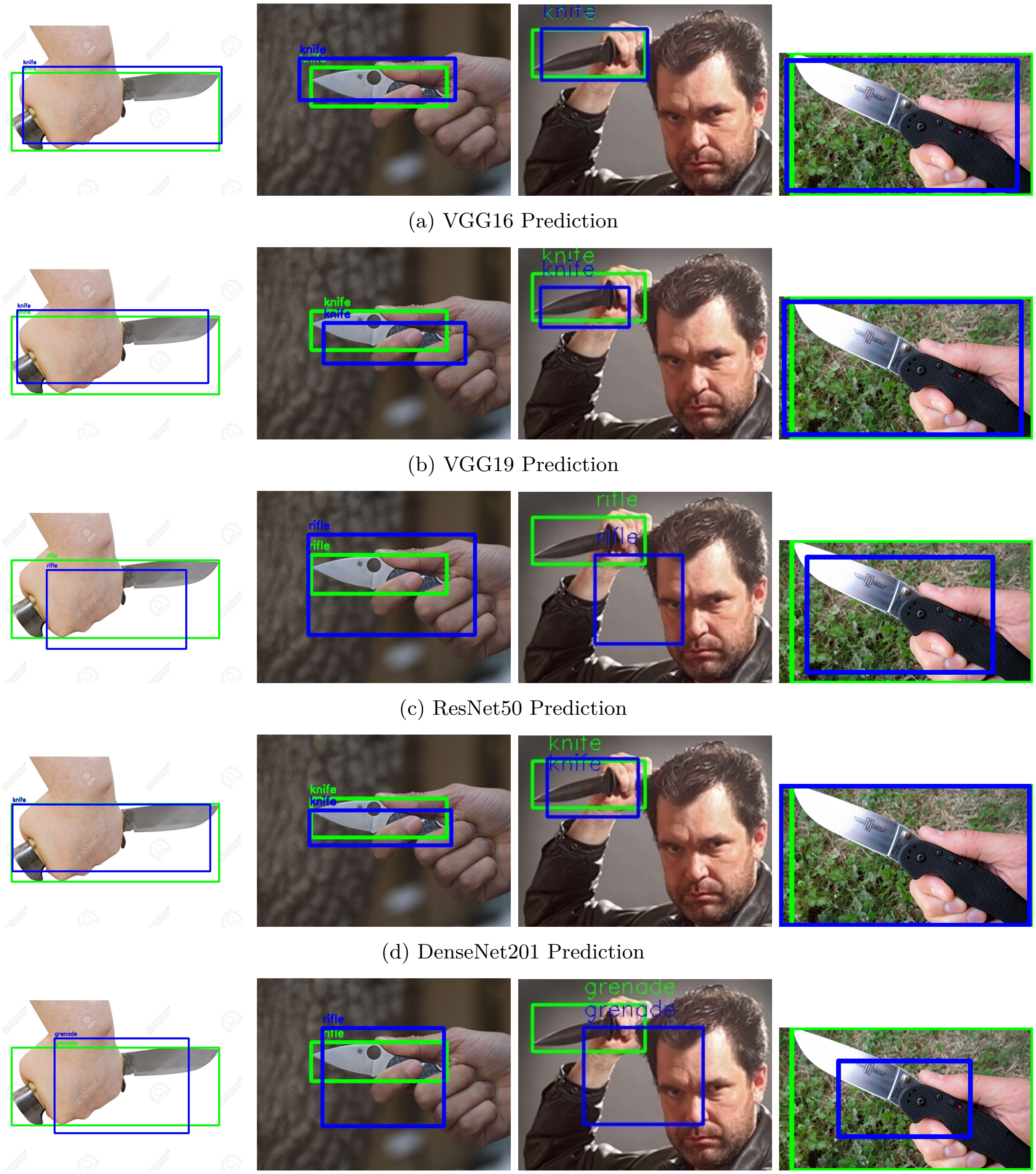
(c) ResNet50 Prediction



(d) DenseNet201 Prediction

(e) MobileNetV3Large Prediction

Figure 5.1: Prediction of Grenade Images



(a) VGG16 Prediction

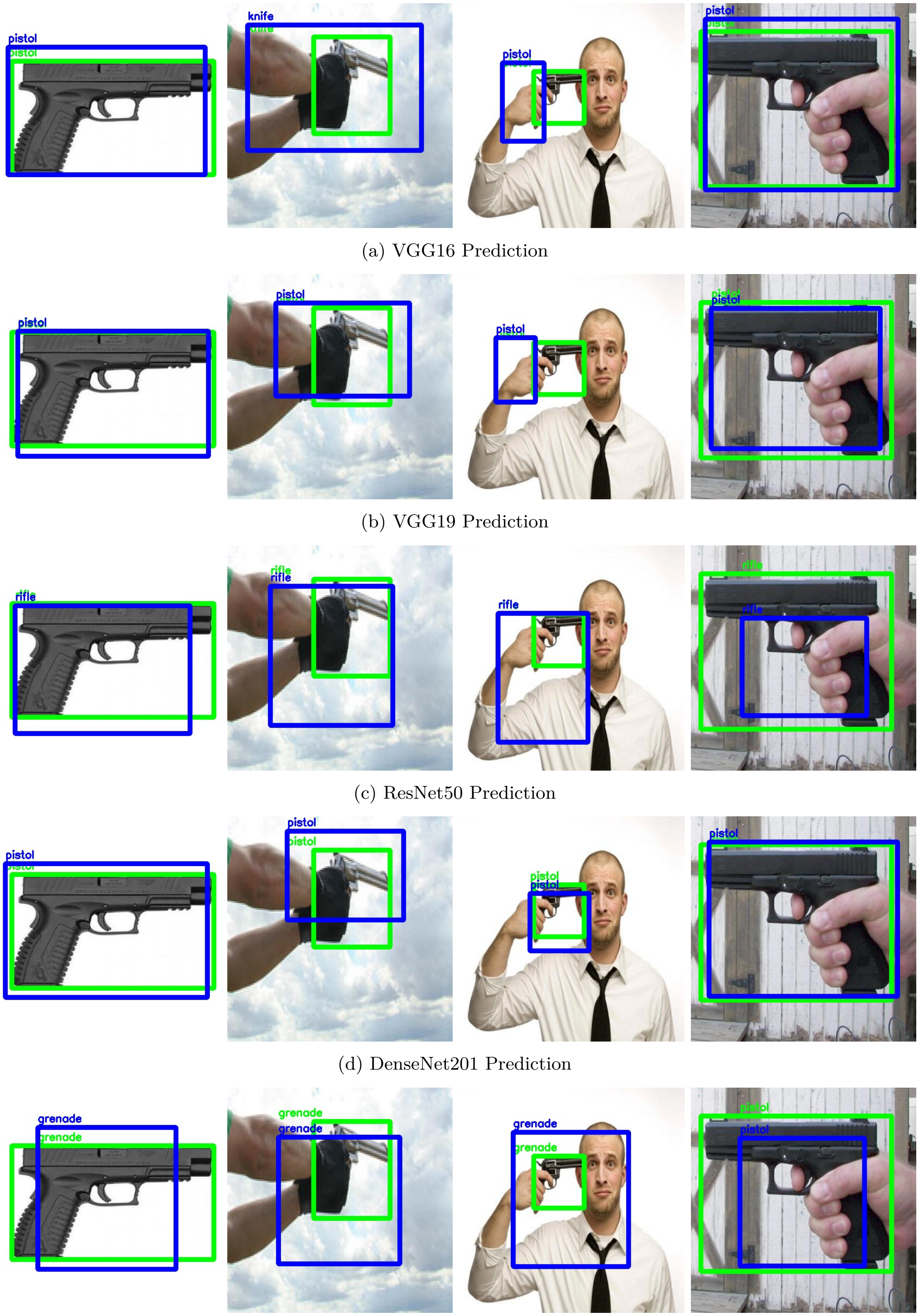
(b) VGG19 Prediction

(c) ResNet50 Prediction

(d) DenseNet201 Prediction

(e) MobileNetV3Large Prediction

Figure 5.2: Prediction of Knife Images



(a) VGG16 Prediction

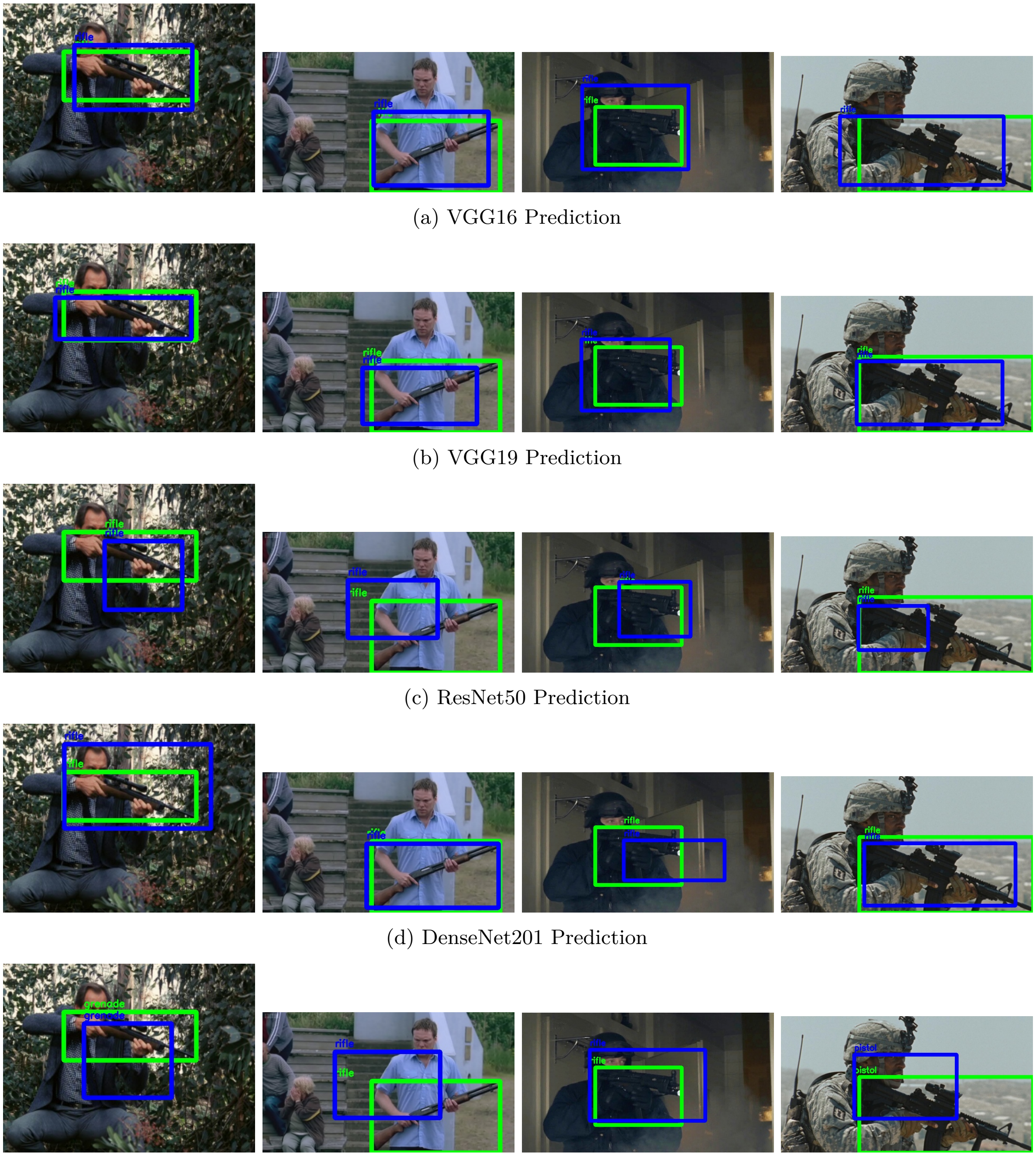
(b) VGG19 Prediction

(c) ResNet50 Prediction

(d) DenseNet201 Prediction

(e) MobileNetV3Large Prediction

Figure 5.3: Prediction of Pistol Images



(a) VGG16 Prediction

(b) VGG19 Prediction

(c) ResNet50 Prediction

(d) DenseNet201 Prediction

(e) MobileNetV3Large Prediction

Figure 5.4: Prediction of Rifle Images

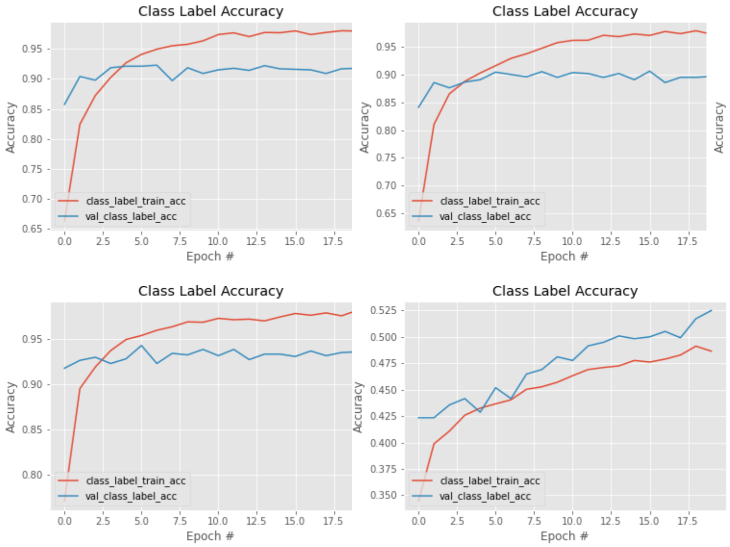


Figure 5.5: Training-Validation accuracy over the epochs for VGG16, VGG19, ResNet50, DenseNet201, and MobileNetV3Large

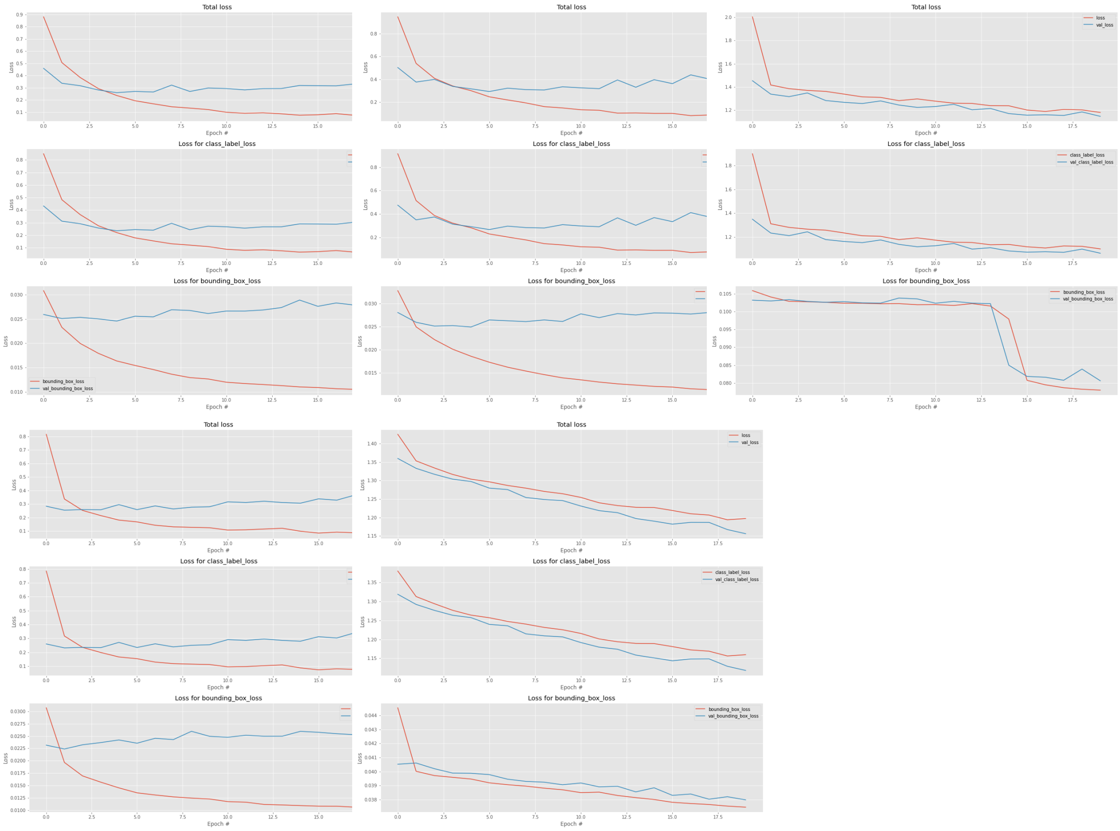


Figure 5.6: Training-Validation loss over the epochs for VGG16, VGG19, ResNet50, DenseNet201, and MobileNetV3Large

5.2.2 Classification Reports

This section shows the classification report of various deep learning algorithms. The classification report involves accuracy, precision, and F1-score. The classification report shows the efficiency of the classification of the deep learning algorithm. Table 5.1 shows the classification report for training dataset and table 5.2 shows the classification report for validation dataset.

Table 5.1: Weapon Classification Results for training dataset

Classifier	Accuracy	Precision	Recall	F1
VGG16	0.99	0.99	0.99	0.99
VGG19	0.99	0.99	0.99	0.99
ResNet50	0.56	0.43	0.56	0.44
DenseNet201	0.99	0.99	0.99	0.99
MobileNetV3Large	0.51	0.49	0.51	0.44

Table 5.2: Weapon Classification Results for validation dataset

Classifier	Accuracy	Precision	Recall	F1
VGG16	0.92	0.92	0.92	0.92
VGG19	0.90	0.90	0.90	0.89
ResNet50	0.54	0.47	0.54	0.42
DenseNet201	0.94	0.94	0.94	0.94
MobileNetV3Large	0.52	0.50	0.52	0.45

5.2.3 Precision-Recall Curve

This section demonstrates the precision-recall curves for the training and testing dataset for all the deep learning algorithms. It depicts the trade-off of precision and recall of an algorithm. The figures 5.7, 5.8, 5.9, 5.10, and 5.11 show precision-recall curves for VGG16, VGG19, ResNet50, DenseNet201, and MobileNetV3Large respectively.

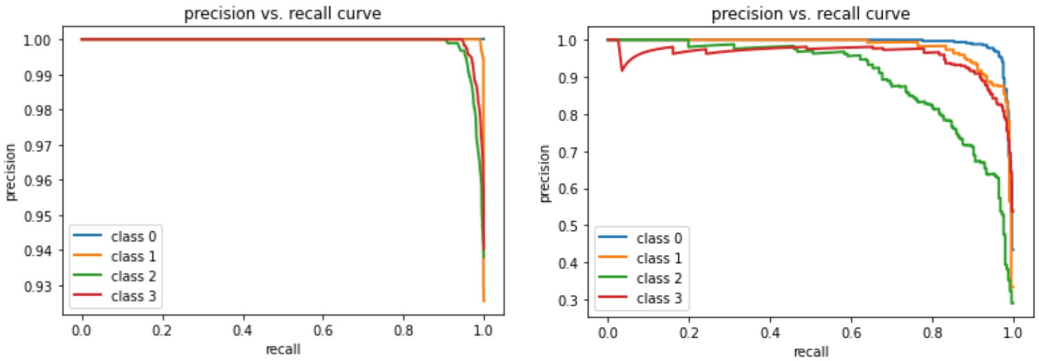


Figure 5.7: Precision-Recall Curve for training and validation dataset of VGG16

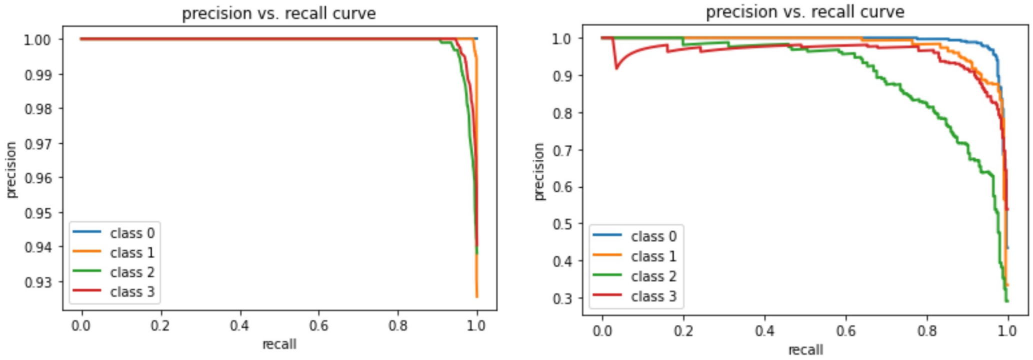


Figure 5.8: Precision-Recall Curve for training and validation dataset of VGG19

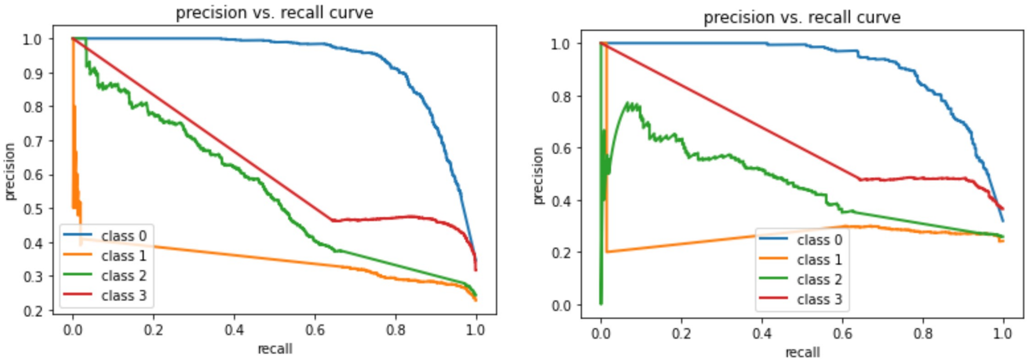


Figure 5.9: Precision-Recall Curve for training and validation dataset of ResNet50

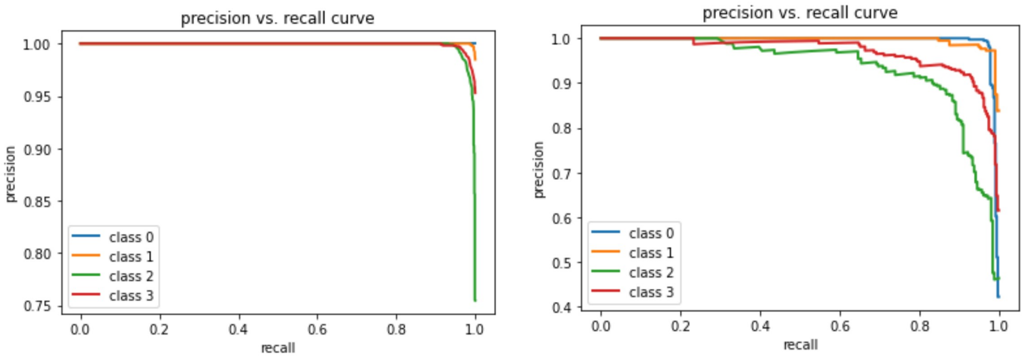


Figure 5.10: Precision-Recall Curve for training and validation dataset of DenseNet201

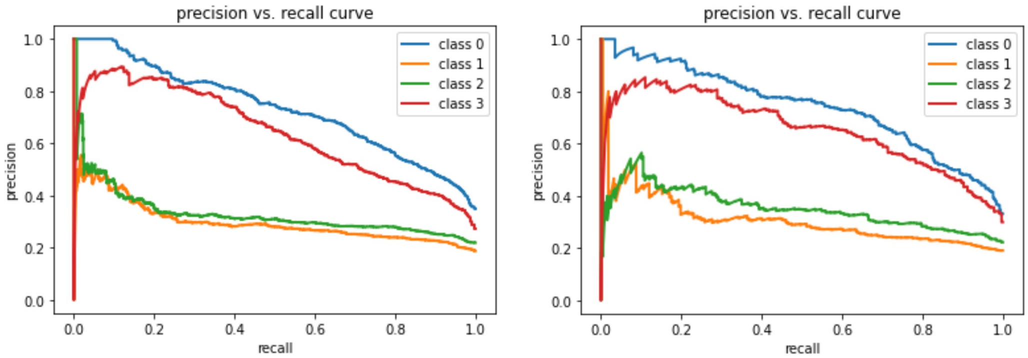


Figure 5.11: Precision-Recall Curve for training and validation dataset of MobileNetV3Large

5.2.4 ROC Curve

This section shows the results of the Receiver Operating Characteristic (ROC) curve of various deep learning algorithms. ROC curve is an evaluation metric for the classification algorithms. It shows the relationship between clinical sensitivity and specificity for every possible cut-off. It is represented in a graphical form. The figures 5.12, 5.13, 5.14, 5.15, and 5.16 show ROC curves for VGG16, VGG19, ResNet50, DenseNet201, and MobileNetV3Large respectively.

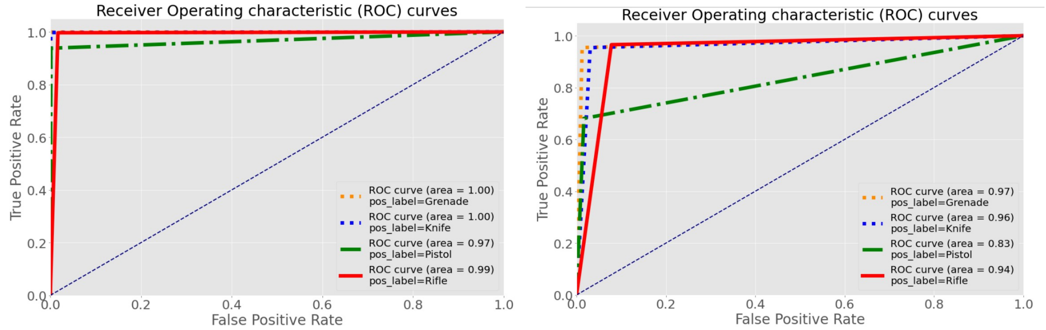


Figure 5.12: ROC curves of VGG16 for training and validation dataset

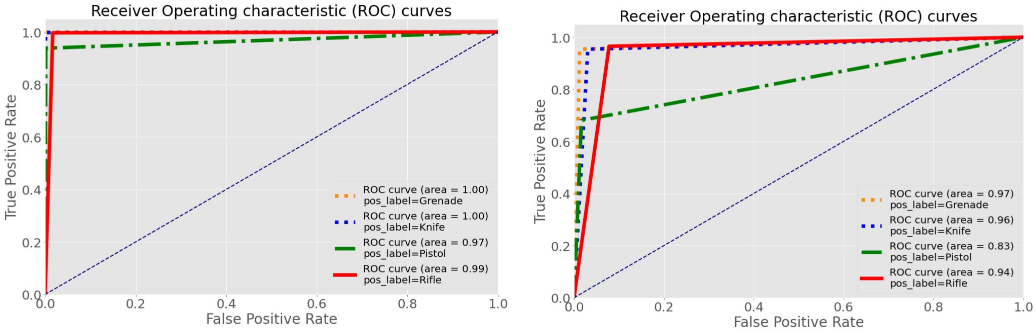


Figure 5.13: ROC curves of VGG19 for training and validation dataset

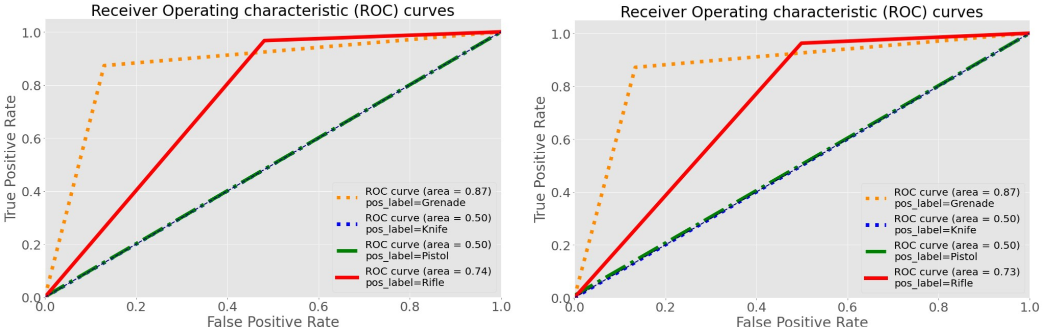


Figure 5.14: ROC curves of ResNet50 for training and validation dataset

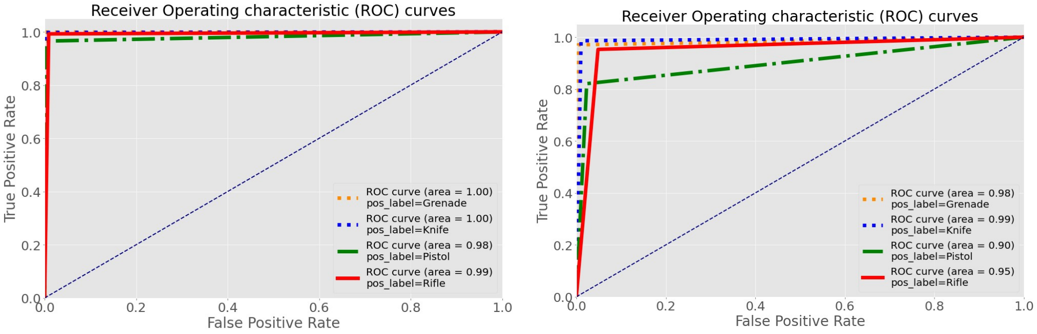


Figure 5.15: ROC curves of DenseNet201 for training and validation dataset

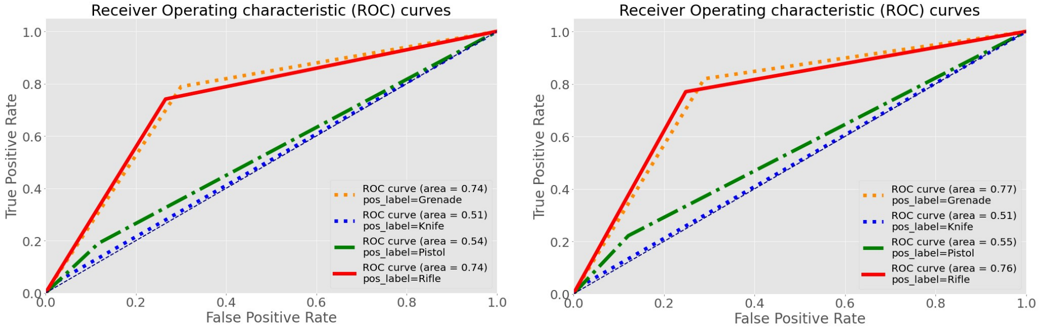


Figure 5.16: ROC curves of MobileNetV3Large for training and validation dataset

5.2.5 Confusion Matrix

This section gives information about the number of images that are predicted accurately. The confusion matrix shows the numbers of True-Positive, True-Negative, False-Positive, and False-Negative of the predicted images. The figures 5.17. 5.18, 5.19, 5.20, and 5.21 show confusion matrix for VGG16, VGG19, ResNet50, DenseNet201, and MobileNetV3Large respectively.

5.2.6 Evaluation Metrics

This section gives information about the efficiency of the bounding box plotted by the algorithms. Jaccard similarity index shows the number from 0 to 1, which depicts the similarity of ground truths and predicted rectangle. Dice coefficient is also one evaluation matrix for image segmentation which shows the similarity between two datasets. Figure 5.22 shows the Dice Score and figure 5.23 shows the Jaccard Score for different deep learning algorithms.

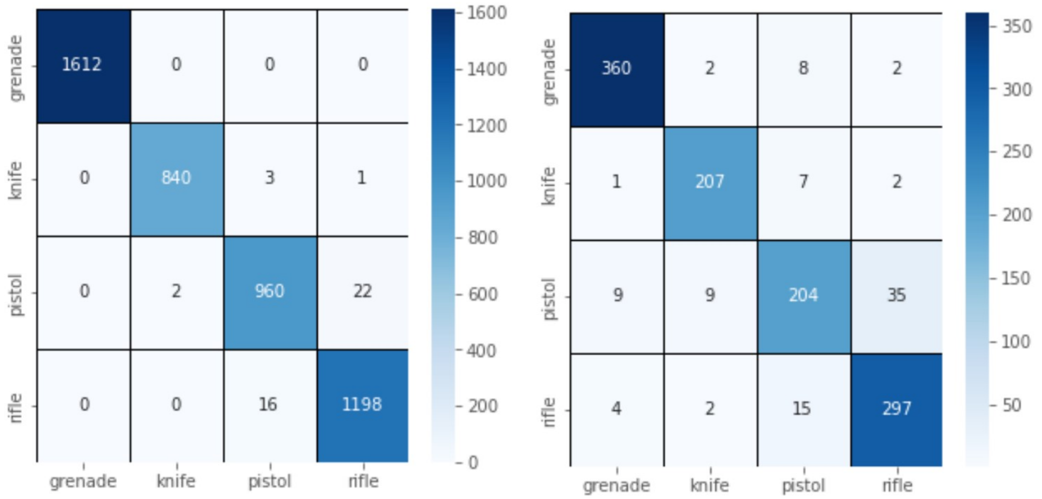


Figure 5.17: Confusion matrix of VGG16 for training and validation dataset

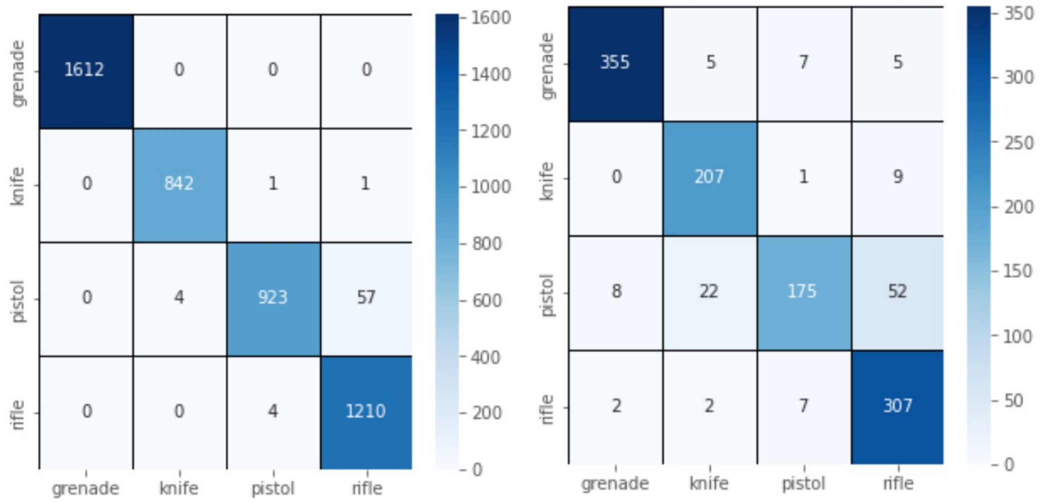


Figure 5.18: Confusion matrix of VGG19 for training and validation dataset

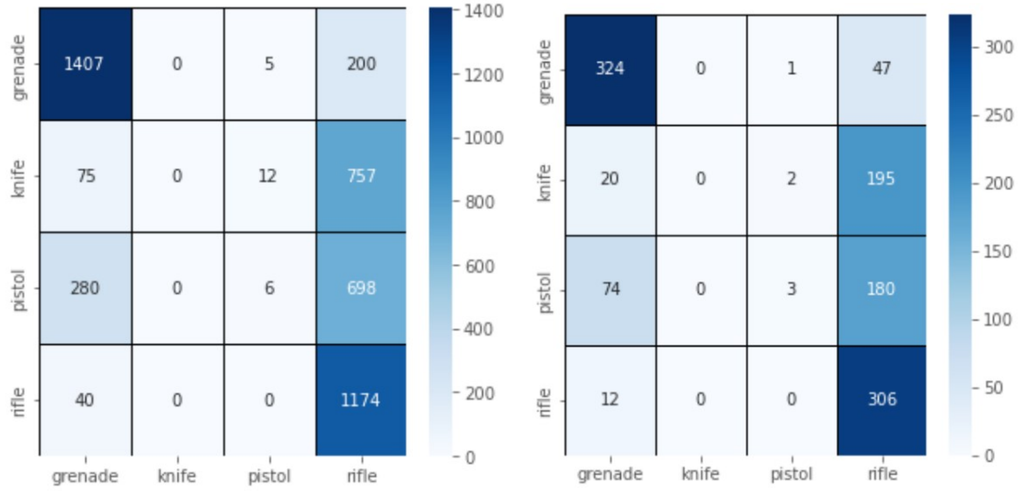


Figure 5.19: Confusion matrix of ResNet50 for training and validation dataset

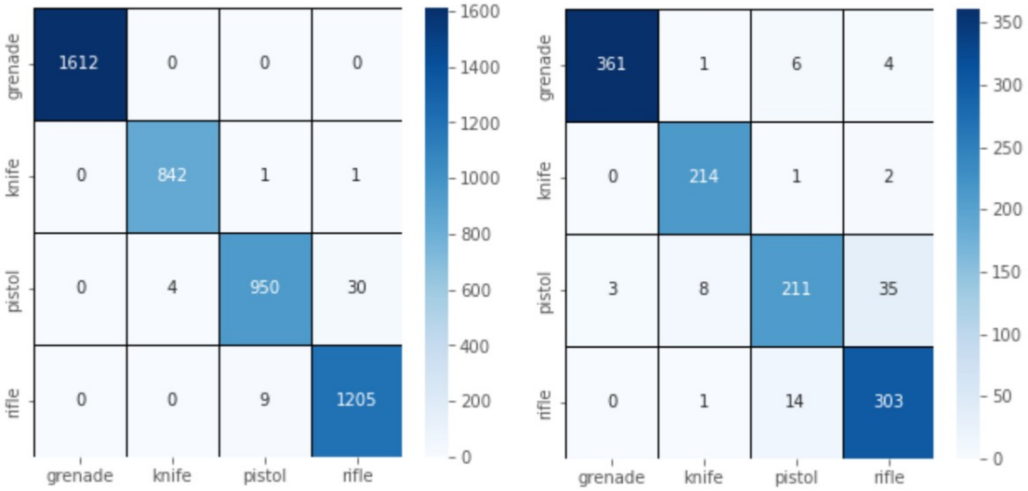


Figure 5.20: Confusion matrix of DenseNet201 for training and validation dataset

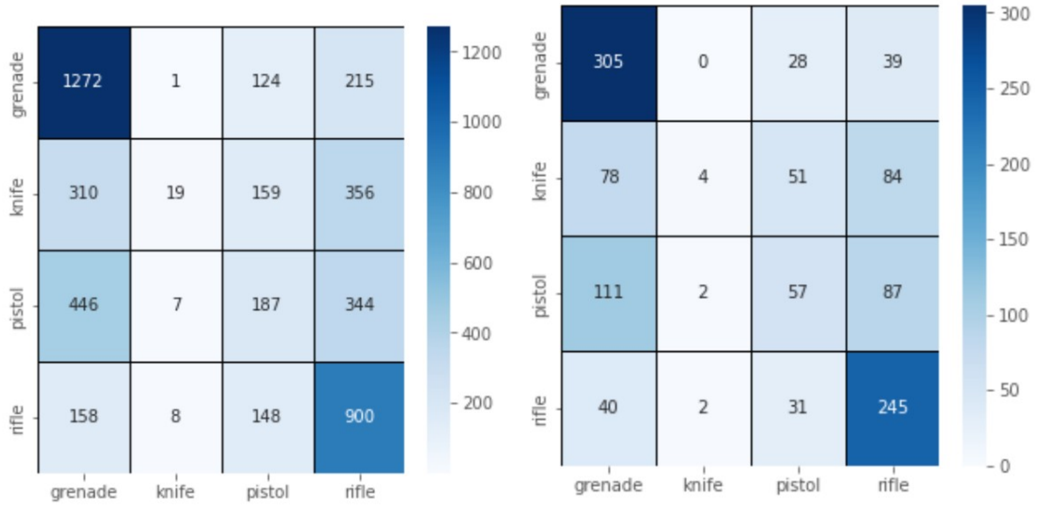


Figure 5.21: Confusion matrix of MobileNetV3Large for training and validation dataset

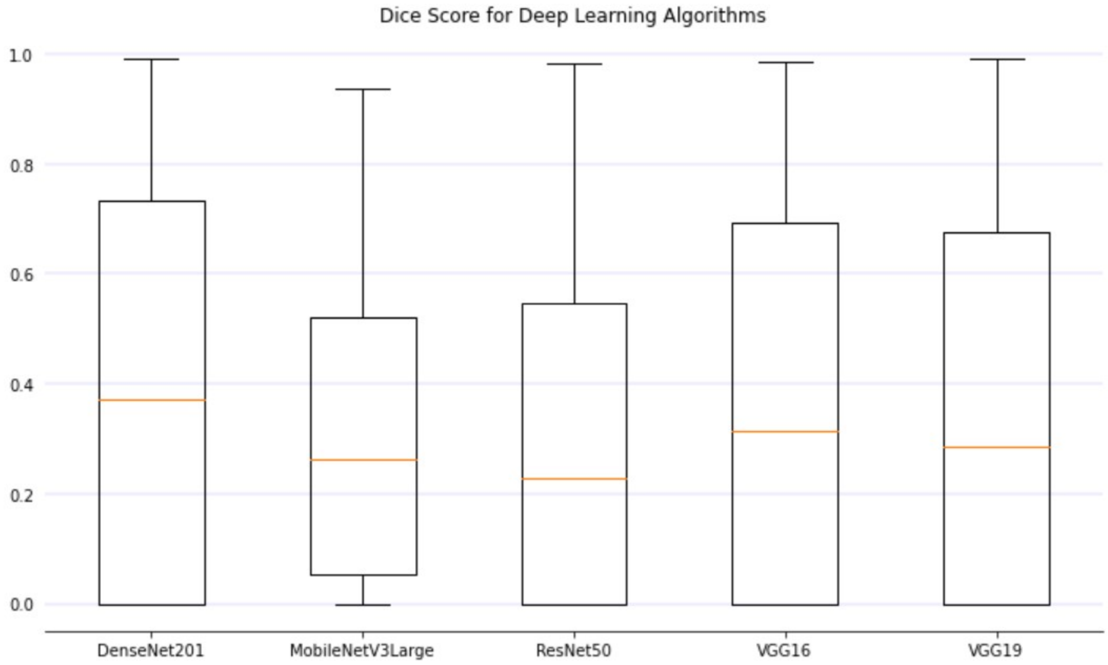


Figure 5.22: Dice score for deep learning algorithms

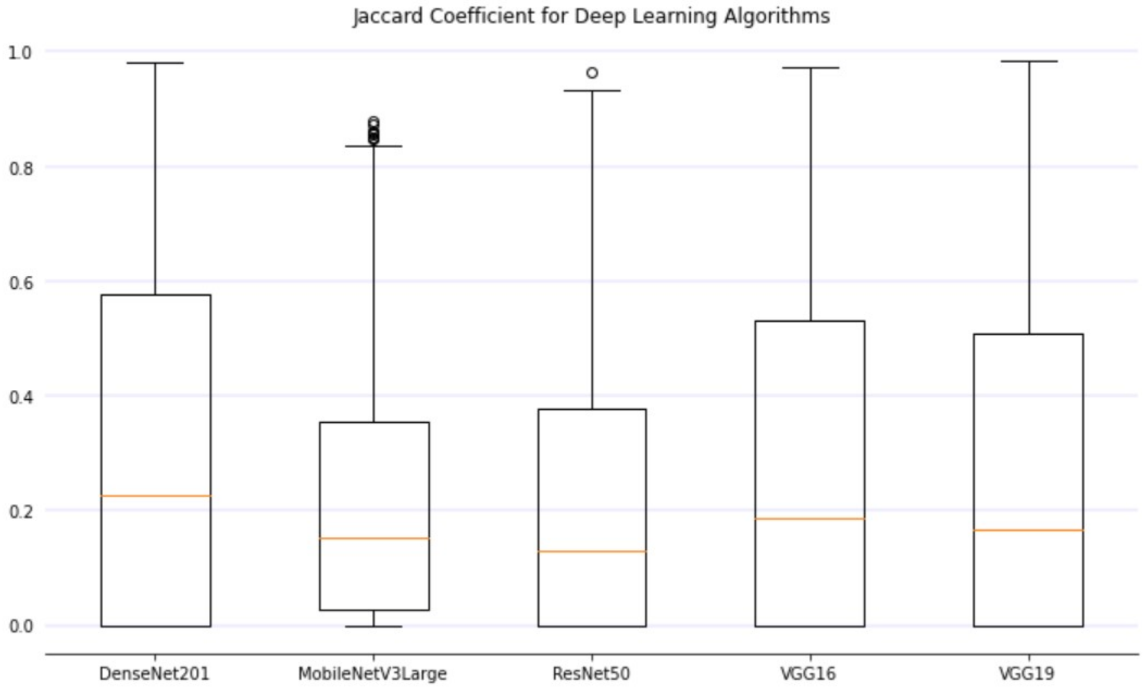


Figure 5.23: Jaccard coefficient for deep learning algorithms

Chapter 6

Conclusion and Future Works

There has been a massive rise in crime rates because of firearms and knives. This research work focuses on weapon detection and classification problem by using images and CCTV footage. Research has shown that there has been an increase in the use of weapons globally. Over the period, to increase the security of civilians governments and private companies have installed CCTV cameras for better surveillance. Installation of cameras is useful to some extent but the major drawback of this system is manual monitoring of the surveillance cameras. This research focuses on giving an approach to implementing automation in the existing system by using deep learning algorithms. We implemented deep learning algorithms which can predict and identify the weapon from the videos or images. The algorithms can also classify the image/video among four classes.

By using deep learning algorithms, we classified the four classes of weapons. In this research, we used the supervised learning approach to classify and detect the weapons from the images. We used various evaluation metrics to identify the best deep learning algorithm for the weapon classification problem. Moreover, we concluded that simple deep learning algorithms like VGG16 and VGG19 are more efficient and accurate than lightweight architectures and residually connected architectures. The essential part of this research is the algorithms are also trained on toy images of the weapons so that they can identify them as well. The problem is challenging and sensitive for humans so it becomes very important to detect False-Negative images as well.

This research work will help government and private security companies to increase the layer of security. This research is still in progress as we still analyzing the deep learning algorithms for real-time detection and classification. Thus, security officers can take appropriate actions on time. This research work is still on to protect the fundamental rights of humans.

Bibliography

[1] *Homicides by firearms, 2019*, 2022. [Online]. Available: https://ourworldindata.org/homicides.

[2] I. Geng, *How does video surveillance system work?, 2021*, 2022. [Online]. Available: https://blog.router-switch.com/2021/10/how-does-video-surveillance-system-work/.

[3] P. Sun, *Deep learning technology applications for video surveillance, 2021*, 2022. [Online]. Available: https://www.securityinformed.com/insights/deep-learning-technology-applications-video-surveillance-co-14319-ga.21460.html/.

[4] N. Wani and K. Raza, "Chapter 3 - multiple kernel-learning approach for medical image analysis," in *Soft Computing Based Medical Image Analysis*, N. Dey, A. S. Ashour, F. Shi, and V. E. Balas, Eds., Academic Press, 2018, pp. 31–47, ISBN: 978-0-12-813087-2. DOI: https://doi.org/10.1016/B978-0-12-813087-2.00002-6. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780128130872000026.

[5] N. Ben Halima and O. Hosam, "Bag of words based surveillance system using support vector machines," *International Journal of Security and Its Applications*, vol. 10, pp. 331–346, Apr. 2016. DOI: 10.14257/ijsia.2016.10.4.30.

[6] R. K. Tiwari and G. K. Verma, "A computer vision based framework for visual gun detection using harris interest point detector," *Procedia Computer Science*, vol. 54, pp. 703–712, 2015, Eleventh International Conference on Communication Networks, ICCN 2015, August 21-23, 2015, Bangalore, India Eleventh International Conference on Data Mining and Warehousing, ICDMW 2015, August 21-23, 2015, Bangalore, India Eleventh International Conference on Image and Signal Processing, ICISP 2015, August 21-23, 2015, Bangalore, India, ISSN: 1877-0509. DOI: https://doi.org/10.1016/j.procs.2015.06.083. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050915014076.

[7] F. Gelana and A. Yadav, "Firearm detection from surveillance cameras using image processing and machine learning techniques: Proceedings of icsiccs-2018," in Jan. 2019, pp. 25–34, ISBN: 978-981-13-2413-0. DOI: 10.1007/978-981-13-2414-7_3.

[8] O. Barnich and M. Van Droogenbroeck, "Vibe: A universal background subtraction algorithm for video sequences," *IEEE Transactions on Image Processing*, vol. 20, no. 6, pp. 1709–1724, 2011. DOI: 10.1109/TIP.2010.2101613.

[9] Z. Zivkovic, "Improved adaptive gaussian mixture model for background subtraction," *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, vol. 2, 28–31 Vol.2, 2004.

[10] S. Brutzer, B. Höferlin, and G. Heidemann, "Evaluation of background subtraction techniques for video surveillance," in *CVPR 2011*, 2011, pp. 1937–1944. DOI: 10.1109/CVPR.2011.5995508.

[11] M. T. Bhatti, M. G. Khan, M. Aslam, and M. J. Fiaz, "Weapon detection in real-time cctv videos using deep learning," *IEEE Access*, vol. 9, pp. 34 366–34 382, 2021. DOI: 10.1109/ACCESS.2021.3059170.

[12] H. Jain, A. Vikram, Mohana, A. Kashyap, and A. Jain, "Weapon detection using artificial intelligence and deep learning for security applications," in *2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)*, 2020, pp. 193–198. DOI: 10.1109/ICESC48915.2020.9155832.

[13] R. Olmos, S. Tabik, and F. Herrera, "Automatic handgun detection alarm in videos using deep learning," *Neurocomputing*, vol. 275, pp. 66–72, 2018, ISSN: 0925-2312. DOI: https://doi.org/10.1016/j.neucom.2017.05.012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925231217308196.

[14] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1, 2005, 886–893 vol. 1. DOI: 10.1109/CVPR.2005.177.

[15] M. J. S. K. Asrith, K. P. Reddy, and Sujihelen, "Face recognition and weapon detection from very low resolution image," in *2018 International Conference on Emerging Trends and Innovations In Engineering And Technological Research (ICETIETR)*, 2018, pp. 1–5. DOI: 10.1109/ICETIETR.2018.8529108.

[16] N. Dwivedi, D. K. Singh, and D. S. Kushwaha, "Weapon classification using deep convolutional neural network," in *2019 IEEE Conference on Information and Communication Technology*, 2019, pp. 1–5. DOI: 10.1109/CICT48419.2019.9066227.

[17] V. Kaya, S. Tuncer, and A. Baran, "Detection and classification of different weapon types using deep learning," *Applied Sciences*, vol. 11, p. 7535, Aug. 2021. DOI: 10.3390/app11167535.

[18] A. Warsi, M. Abdullah, M. N. Husen, and M. Yahya, "Automatic handgun and knife detection algorithms: A review," in *2020 14th International Conference on Ubiquitous Information Management and Communication (IMCOM)*, 2020, pp. 1–9. DOI: 10.1109/IMCOM48794.2020.9001725.

[19] K. Simonyan and A. Zisserman, *Very deep convolutional networks for large-scale image recognition*, 2014. DOI: 10.48550/ARXIV.1409.1556. [Online]. Available: https://arxiv.org/abs/1409.1556.

[20] K. He, X. Zhang, S. Ren, and J. Sun, *Deep residual learning for image recognition*, 2015. DOI: 10.48550/ARXIV.1512.03385. [Online]. Available: https://arxiv.org/abs/1512.03385.

[21] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, *Densely connected convolutional networks*, 2016. DOI: 10.48550/ARXIV.1608.06993. [Online]. Available: https://arxiv.org/abs/1608.06993.

[22] A. G. Howard, M. Zhu, B. Chen, *et al.*, *Mobilenets: Efficient convolutional neural networks for mobile vision applications*, 2017. DOI: 10.48550/ARXIV.1704.04861. [Online]. Available: https://arxiv.org/abs/1704.04861.

[23] R. Olmos, S. Tabik, and F. Herrera, "Automatic handgun detection alarm in videos using deep learning," *Neurocomputing*, vol. 275, Feb. 2017. DOI: `10.1016/j.neucom.2017.05.012`.

[24] A. Castillo Lamas, S. Tabik, F. Pérez, R. Olmos, and F. Herrera, "Brightness guided preprocessing for automatic cold steel weapon detection in surveillance videos with deep learning," *Neurocomputing*, vol. 330, Nov. 2018. DOI: `10.1016/j.neucom.2018.10.076`.

[25] S. Bansal, *Supervised and unsupervised learning*, July 7, 2022. [Online]. Available: `https://www.geeksforgeeks.org/supervised-unsupervised-learning/`.

[26] J. Delua, *Supervised and unsupervised learning*, March, 2015. [Online]. Available: `https://www.ibm.com/cloud/blog/supervised-vs-unsupervised-learning`.

[27] Tzutalin, *Labelimg*, 2015. [Online]. Available: `https://github.com/tzutalin/labelImg`.

[28] R. Khandelwal, *Coco and pascal voc data format for object detection*, 2019. [Online]. Available: `https://towardsdatascience.com/coco-data-format-for-object-detection-a4c5eaf518c5`.