

Scalable Solutions to Image Abnormality Detection and Restoration using Limited Contextual Information

by

Subhayan Mukherjee

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

Doctor of Philosophy

Department of Computing Science

University of Alberta

© Subhayan Mukherjee, 2020

Abstract

Detecting and interpreting image abnormalities and restoring images are essential to many processing pipelines in diverse fields. Challenges involved include randomness and unstructured nature of image artefacts (from signal processing perspective) and performance constraints imposed by resource-constrained systems (computational perspective). This thesis studies three such scenarios involving different sensor modalities, and proposes novel approaches to address such challenges using computer vision and machine learning techniques. The three scenarios are: GPU-friendly debanding for mobile HDR, MRI abnormality detection, and InSAR signal recovery for WAM. Below, we summarize each scenario, and outline the unique challenges involved in solving them.

In the first scenario, in High-Dynamic-Range (HDR) imaging domain, we propose how to perceptually eliminate quantization artefacts resulting from dynamic range conversion, without distorting image contents, by adding noise patterns in the mobile Graphics Processing Unit (GPU) computing environment with limited computational resources. Traditional filtering methods are computationally non-ideal due to limited access to neighborhood information. We introduce a pixel based solution which does not rely on any neighborhood information. We impose a real-world transmission scenario where the receiving end cannot access the un-quantized (original) image. Most existing methods assume access to the original image, and this makes the image restoration problem easier to address. Our challenge is designing noise patterns that are perceptually pleasant, i.e. blended into the image content naturally based on

the intensity profile and dynamic range conversion characteristics of the image.

In the second scenario, we apply our methods to medical imaging: Magnetic Resonance (MR) images of preterm infant brains. We propose abnormality detection without using tissue priors (atlas) from a single acquisition sequence (T1-weighted). The rate of preterm births is increasing worldwide at an alarming rate. Preterm infant brains are at extremely high risk of developing abnormalities, which deter neuro-development. Traditional segmentation-based lesion detection methods rely on brain atlases to guide segmentation. For rapidly evolving preterm infant brains, such reliable atlases are unavailable, and low signal-to-noise ratio of small preterm infant brains complicates the restoration of tissue intensities. This motivates our solution design using more generic structural assumptions and heuristics for atlas-free WMI detection.

In the third scenario, for Wide Area Monitoring (WAM) via ground movement prediction from noisy interferometric images, e.g. nation-wide monitoring for earthquake prediction or landslide prediction at mining sites, we propose image detail recovery through unsupervised learning-based filtering and filter output confidence prediction due to unavailability of clean training data. Filtering and confidence estimation are crucial steps in interferometric image processing pipelines. Challenges arise due to atmospheric factors, e.g. moisture, corrupting the images during acquisition. Also, corresponding pixels in time series images get decorrelated due to other undesired movements on the ground, e.g. moving vehicles on roads, vegetation and flowing water. Intermediate results of the filtering and confidence estimation steps can also influence the quality of the pipeline’s final result. Traditional algorithms in this area were not designed by prioritizing scalability via parallelism, without sacrificing accuracy, which are crucial for WAM, and addressed in this thesis.

Thus, our main contribution is proposing novel solutions in scenarios where traditional image abnormality detection / restoration approaches are inade-

quate to address the signal processing and computational challenges involved.

Preface

I started my PhD studies by applying my image processing knowledge and skills to investigate how WMI in preterm newborns could impact their developmental defects at a later age. However, our collaborator did not have sufficient data to perform further study. Furthermore, in order to obtain more reliable analysis, longitudinal data over a 10 years period is preferred. Therefore, the research was put on hold.

Then I got an internship opportunity in Dolby, California on HDR research from 2016 September to 2017 April. However, when I returned to UofA, only limited work could be conducted because the research outcome needs to be validated on HDR monitors, which was too expensive for the lab to purchase.

From 2017 July, I started to work on the CARIC funded InSAR project with industrial partner 3vG Vancouver. In this project, I applied my computer vision skills combined with machine learning approaches.

The contents of this thesis have mostly been published or currently under review in peer-reviewed international conferences or journals/letters. Chapter 3 describes the first algorithm to detect WMI and PVH from single sequence (T1) MRI in preterm infants without using brain atlas, as published in *Springer - Medical and Biological Engineering and Computing* and *International Conference on Smart Multimedia (ICSM) (Springer LNCS proceedings)*. Chapter 4 describes the first application of CNNs to InSAR phase filtering and coherence estimation and has been published in *IEEE Sensors*, while its extension to generative modelling as described in the same chapter, is currently under review with *IEEE Geoscience and Remote Sensing Letters*. Chapter 5 describes the first de-banding method which uses dithering directly on the quantized image without accessing any neighborhood information, as published in *ICSM*.

I write this thesis in first person plural to acknowledge and honor the contribution of my advisors and collaborators.

Ethics approval

Informed consent was obtained from all individual participants of the debanding subjective experiments. For the WMI/PVH work, all images were provided by Dr. Steven Miller's group at SickKids Hospital, Toronto, Canada from a research project for which parents provided consent. They followed the "Standard Protocol Approvals, Registration, and Patient Consents" at the BC Children's Hospital in Vancouver. A written informed consent from the legal guardian of each participating neonate was obtained. This study was reviewed and approved by the Clinical Research Ethics Board at the University of British Columbia and BC Children's and Women's Hospitals.

Related First-Author Publications during PhD study

- S. Mukherjee, I. Cheng, S. Miller, T. Guo, V. Chau, and A. Basu, “A fast segmentation-free fully automated approach to white matter injury detection in preterm infants,” *Medical & Biological Engineering & Computing*, vol. 57, no. 1, pp. 71–87, Jul. 2018.
- S. Mukherjee, I. Cheng, and A. Basu, “Atlas-Free Method of Periventricular Hemorrhage Detection from Preterm Infants’ T1 MR Images,” in *Smart Multimedia, Lecture Notes in Computer Science*, Springer International Publishing, Toulon, 2018, vol. 11010, pp. 157–168.
- S. Mukherjee, A. Zimmer, N. K. Kottayil, X. Sun, P. Ghuman and I. Cheng, “CNN-Based InSAR Denoising and Coherence Metric,” *IEEE SENSORS*, New Delhi, 2018, pp. 1–4
- S. Mukherjee, A. Zimmer, X. Sun, P. Ghuman and I. Cheng, “CNN-based InSAR Coherence Classification,” *IEEE SENSORS*, New Delhi, 2018, pp. 1–4.
- S. Mukherjee, A. Zimmer, X. Sun, and I. Cheng, “A Novel Generative Neural Approach for InSAR Joint Phase Filtering and Coherence Estimation,” *IEEE Geoscience and Remote Sensing Letters*. (submitted)
- S. Mukherjee, N. K. Kottayil, X. Sun, and I. Cheng, “CNN-Based Real-Time Parameter Tuning for Optimizing Denoising Filter Performance,” in *Image Analysis and Recognition, Lecture Notes in Computer Science*, Springer International Publishing, Waterloo, 2019, vol. 11662, pp. 112–125.
- S. Mukherjee, G.-M. Su, and I. Cheng, “Adaptive Dithering Using Curved Markov-Gaussian Noise in the Quantized Domain for Mapping SDR to HDR Image,” in *Smart Multimedia, Lecture Notes in Computer Science*, Springer International Publishing, Toulon, 2018, vol. 11010, pp. 193–203.

Acknowledgements

I would like to thank my advisors, Professors Irene Cheng and Anup Basu for their support, motivation and enthusiasm which enriched my PhD journey.

My thanks to Dr. Steven Miller and his team at SickKids Hospital, Toronto, Canada for contributing the annotated images which helped us validate our WMI/PVH work. I am grateful to Dr. Guan-Ming Su for the internship opportunity at Dolby Laboratories, Sunnyvale, USA where I first experienced industrial research. I thank Prof Irene and all my colleagues and mentors at Dolby for their encouragement and insights which enriched my internship work. My thanks to Parwant Ghuman for the internship opportunity at 3vGeomatics, Vancouver, Canada, to Prof Irene, and my 3vG mentor, Aaron Zimmer for helping navigate the steep learning curve to understanding InSAR images and valuable inputs to algorithm design and performance evaluation, which contributed immensely to the success of the 3vG internship. I thank the Consortium of Aerospace Research and Innovation in Canada (CARIC) and the Mitacs Accelerate programme for contributing in funding this internship.

I gratefully acknowledge the Alberta Government for my Alberta Graduate Excellence Scholarship (AGES). I am also thankful for financial support received directly or indirectly from CIHR, NeuroDevNet, Alberta Innovates (iCORE) Research Chair program, NSERC and DND during my PhD studies.

I thank my thesis committee Professors Nilanjan Ray, Ehab Elmallah and Witold Pedrycz for their insights and feedback in improving my thesis writing.

My heartfelt thanks to my mother, Krishna Mukherjee for her continuous encouragement and support, and those of my UofA colleagues / friends: Nava-neeth, Xinyao, Sara, Housam, Amir, Nasim, Gabriel, Chenqiu, Ajin, Krishna, Rohith, Megha, Sucheta, Sankalp, Sanket, Soumik, Ananda, and many others.

Contents

1	Introduction	1
1.1	Image abnormality paradigms	2
1.2	Signal paradigm problems	3
1.2.1	MRI abnormality detection	3
1.2.2	InSAR signal recovery for WAM	5
1.3	Artifact paradigm problem	8
1.3.1	GPU-friendly debanding for mobile HDR	8
1.4	Organization of the thesis	10
2	Segmentation via Parametric and Structural Optimization	11
2.1	Snakes	12
2.2	Active Contours	13
2.3	Level Sets	14
2.4	Chan-Vese algorithm	15
3	MRI abnormality detection	16
3.1	Introduction	16
3.1.1	MRI basics	16
3.1.2	Motivation and Clinical Significance	18
3.1.3	Computational and Signal Processing Challenges	20
3.1.4	Our Computational and Clinical Contributions	22
3.1.5	Differences with Related Work	23
3.2	Proposed Method	27
3.2.1	Coarse Detection phase	27
3.2.2	Fine Detection phase	34
3.3	Results	35
3.3.1	Quantitative Results	40
3.3.2	Qualitative Results	44
3.4	Discussion	49
3.4.1	Slice Thickness	51
3.4.2	Applicability of general assumptions regarding WM lesion characteristics	52
3.5	Conclusion	53
4	InSAR signal recovery for WAM	54
4.1	Introduction	54
4.2	Proposed Methods	57
4.2.1	CNN-InSAR: CNN-based phase filtering and coherence estimation	57
4.2.2	GenInSAR: CNN-based generative modeling for phase filtering and coherence estimation	60
4.3	Results and Discussion	64
4.3.1	Experiments using satellite InSAR images	65

4.3.2	Experiments using simulated InSAR images	65
4.3.3	Key observations from few other experiments	68
4.4	Conclusion	71
5	GPU-friendly debanding for mobile HDR	72
5.1	Introduction	72
5.1.1	Motivation	77
5.2	Proposed Method	78
5.2.1	Markov Gaussian Noise	79
5.2.2	Enhancement of Markov Gaussian Noise	81
5.3	Results and Discussion	89
5.3.1	Dataset	89
5.3.2	Quality of Dithered Images	89
5.3.3	Subjective Testing	89
5.4	Conclusion	92
6	Conclusion and Future Work	94
	References	96

List of Tables

3.1	MRI Acquisition Parameters	35
3.2	Quantitative Comparison (for each slice) of Proposed Method using varying parameter values in [20]	41
3.3	Accuracy & Execution Time comparison of Our Method with Method [20] by varying its parameter values	42
3.4	Wilcoxon Rank-Sum Test comparing Proposed Method’s sensitivity with Method [20]	42
3.5	Effectiveness of the Size and Distance Constraints on Proposed Method	43
3.6	Proposed Method’s Reduced performance with Minimum Lesion Size constraint [14]	43
3.7	Periventricular hemorrhage detection quantitative results	46
4.1	Quantitative evaluation of GenInSAR, CNN-InSAR, and four existing methods, and scalability of GenInSAR over increasing GPU counts	71
5.1	BLUT-Aware Variance and Noise Pattern	88
5.2	Image pair displayed randomly to subjects	91
5.3	DMOS scores from Subjective Test results	92

List of Figures

2.1	PSO deforms a rough initial contour (left) through successive iterations (centre) such that it finally delineates the object of interest (right).	12
3.1	3D reconstruction of adult human head from MRI.	17
3.2	Typical MRI scanner with magnetic strength = 3 Tesla.	17
3.3	Comparison of image qualities to illustrate the challenge in segmenting low resolution preterm neonate brain MR images.	25
3.4	An overview of our proposed method.	28
3.5	Visualization of representative Ground Truth WMIs on coronal cross-sections of subjects.	36
3.6	An illustration of the various steps in our proposed method.	39
3.7	Result of blob detection (top row) and refinement (bottom row) to locate ventricles using the MSER algorithm on Slices 1 through 7. Each detected blob has been shown using a different colour for ease of visualization.	39
3.8	Comparison of Proposed Method and [20] using Slice 9.	45
3.9	Comparison of Proposed Method and [20] using Slice 10.	46
3.10	Comparison of Proposed Method and [20] using Slice 11.	47
3.11	Periventricular hemorrhage detection using proposed method on DICOM slices obtained from SickKids Hospital, Toronto. Red squares mark True Positives.	48
3.12	True and false positive WMI detection performance of proposed method relative to lesion size: top row shows small lesions, bottom row shows big lesions.	50
4.1	InSAR image capture by two passes of the satellite.	56
4.2	Autoencoder CNN for denoising InSAR images.	58
4.3	CNN for InSAR coherence estimation.	59
4.4	Preprocess raw 11×11 coherence to remove artefacts.	60
4.5	Architecture of GenInSAR.	61
4.6	Filtered phase and coherence outputs for satellite InSAR images processed by GenInSAR, CNN-InSAR, and four existing methods. Visualizations for phase are coloured between $-\pi$ (blue) to $+\pi$ (red), and coherence between 0 (black: low) to 1 (white: high) respectively.	67
4.7	Cropped interferometric phase images generated by GenInSAR for the same noisy input. Visualizations coloured from $-\pi$ (blue) to $+\pi$ (red).	67

4.8	Filtered phase and coherence outputs for simulated InSAR images processed by GenInSAR, CNN-InSAR, and four existing methods. Visualizations for phase are coloured between $-\pi$ (blue) to $+\pi$ (red), and coherence between 0 (black: low) to 1 (white: high) respectively.	70
5.1	A visualization of how HDR compares against SDR	73
5.2	(a, b): Quantized Image and its Cropped region with banding; (c, d): Output from Proposed Method and the same Cropped; (e): Circular Noise, compared with (f): Curved Noise, which shows better blending result. Our method de-bands quantized images effectively. Note that (a-f) are tone-mapped in order to visualize banding/noise in printed form or LDR displays. This is a very challenging image in our dataset.	74
5.3	(a) BLUT for Y channel from a sunset scene b) Extraction of square block of circular pattern noise	79
5.4	Two-state Markov-Gaussian Noise generator	80
5.5	Two-state Markov Gaussian Noise with intra-state transition probabilities 0.6 (top-left), 0.7 (top-right), 0.8 (bottom-left), and 0.9 (bottom-right)	81
5.6	Convert (a) Circular to (b) Curved Markov-Gaussian Noise; Block size = 200×200 ; Trans. Prob. = 0.815	83
5.7	Voronoi Tessellation of arbitrary matrix; adjacent cells: shown in different colors; sites (*): red	84
5.8	Google Forms Questionnaire for Subjective Tests	90
5.9	Confidence interval plots at significance level $\alpha = 0.05$ showing superior performance of proposed adaptive method.	93

Glossary

Adaptive Balloon Snake (ABS)

A technique of demarcating the outline of an object from an image by deforming an initial contour based on energy function minimization.

Backward Look-Up Table (BLUT)

HDR code-words indexed by SDR code-words in a 1:1 correspondence.

CerebroSpinal Fluid (CSF)

A clear body fluid found in the brain and spinal cord which provides immunity and mechanical protection to the brain inside the skull.

Computer Aided Detection (CAD)

Computational methods to reduce observational oversights (false negatives) by physicians in interpreting medical images for disease diagnosis.

Contrast-to-Noise Ratio (CNR)

An image quality measure similar to SNR except that it subtracts the bias from the signal before computing the ratio.

Convolutional Neural Network (CNN)

A type of neural network characterized by shared weights and translation invariance, usually applied in image analysis.

Cumulative Distribution Function (CDF)

A function of a random variable which takes a real value and outputs the probability of the variable taking values lesser than that value.

Differential Mean Opinion Score (DMOS)

Difference of mean opinion scores for reference and processed images.

Digital Elevation Model (DEM)

A three dimensional graphical representation of the surface of a terrain.

Digital Imaging and Communications in Medicine (DICOM)

A medical imaging communication and management standard.

Electro-Optical Transfer Function (EOTF)

A function for converting display data to a particular brightness (nits).

Enhanced Dynamic Range (EDR)

A synonym for HDR.

Expectation Maximization (EM)

An iterative approximation of the likelihood function for incomplete or missing data or latent variables.

FLuid-Attenuated Inversion Recovery (FLAIR)

An MRI sequence used to null signals originating from certain types of tissues in order to emphasize others.

Genetic Algorithm (GA)

Optimization inspired by natural selection principles like mutation.

Graphics Processing Unit (GPU)

Electronic circuits specially designed for rapid memory access to accelerate generation of images for display output. Their structure is highly parallel, suited for algorithms that benefit from parallel data processing.

Grey Matter (GM)

Nerve cells of the human brain.

High Dynamic Range (HDR)

Image capturing/processing/display technologies which achieve higher range of luminosity using more than 8 bits per pixel.

Interferometric Synthetic Aperture Radar (InSAR)

A radar-based remote sensing technique which uses phase difference of waves returning to the satellite while capturing successive SAR images to generate surface elevation or deformation maps.

inverse Tone-Mapping Operator (iTMO)

A function to expand an LDR image into HDR for display.

Joint Photographic Experts Group (JPEG)

A lossy still image compression standard developed and maintained by an international committee having the same name.

Low Dynamic Range (LDR)

Traditional image capturing/processing/display technologies which produce lower range of luminosity than HDR using 8 bits per pixel.

Magnetic Resonance (MR)

Quantum mechanical phenomenon where a magnetic element switches its energy states when tickled with a magnetic field of appropriate frequency.

Magnetic Resonance Imaging (MRI)

A radiology technique for capturing the anatomy and physiological processes of the human body using strong magnetic fields.

Markov Random Field (MRF)

A set of random variables represented as an undirected graph having Markov property.

Maximally Stable Extremal Regions (MSER)

A method to detect blobs in images by extracting co-variant regions as stable connected components of the image's gray-level sets.

Montreal Neurological Institute (MNI)

A neuroscience research and medical center located in Montreal, Canada.

Periventricular-IntraVentricular Hemorrhage (PIVH)

The pathology associated with a PVH extending to the space between the two ventricles of the human brain.

PeriVentricular Hemorrhage (PVH)

Blood vessel ruptures in the vicinity of the ventricles in the human brain.

Probabilistic Fuzzy C-Means (PFCM)

A modification of the fuzzy clustering algorithm for normally distributed data objects, which uses a probabilistic distance for clustering.

Rectified Linear Unit (ReLU)

An activation function used in neural networks which outputs its input if positive and zero otherwise.

Self-Organizing Map (SOM)

A dimensionality reduction method using neural networks trained in an unsupervised manner to create a low dimensional discrete representation of the input space.

Signal-to-Noise Ratio (SNR)

Ratio between signal and background noise power, expressed in decibels.

Standard Dynamic Range (SDR)

A synonym for LDR.

Synthetic Aperture Radar (SAR)

A type of radar which achieves fine imaging resolution by utilizing the motion of the antenna over the region of interest.

White Matter (WM)

Fibrous tissue carrying signals between nerve cells in the human brain.

White Matter Injury (WMI)

Preterm birth-related injuries to the white matter tissue, often caused due to blood pressure fluctuations or inflammations. Such injuries severely increase risks of neuro-developmental deficits in later stages of life.

Wide Area Monitoring (WAM)

Monitoring land displacement over an area spanning multiple provinces or territories for safety of infrastructure like oil and natural gas pipelines.

Chapter 1

Introduction

Image abnormality detection aims to detect, quantify, or interpret aberrations or artefacts in an image, based on prior domain knowledge of how a corresponding clean or normal image should look like. On the other hand, image restoration deals with improving the quality of a distorted image, to get back the original image by removing the aberrations or artefacts present. In a broad sense, the general goal of image restoration is to improve the usability of the image to the end user. Depending on the application area and acquisition modality, the specific goals can be to discover patterns in the image which got hidden due to the distortion, to improve the perceptual quality of the image to the human observer, etc. Similarly, depending on the application area, image abnormality detection can help assess potential perceptual degradation of the image due to artefacts, the usefulness of the image in a processing pipeline, or the presence of features of concern to the particular imaging application (target identification). While image abnormality detection and restoration in general have been sufficiently explored by researchers over the past several decades, often, particular use cases suffer from lack of contextual information and/or call for scalable solutions, not addressed by general-purpose methods.

In this thesis, we explore such gaps in literature and show using specific use cases, how challenges unique to acquisition modalities and application areas can be solved using novel computer vision and machine learning approaches.

1.1 Image abnormality paradigms

Image abnormality estimation and restoration can be viewed in the broader context of image quality analysis, which has been an actively researched area. However, its focus paradigms have changed over time, or newer paradigms have come up and coexisted with traditional ones [79]. There are four such types of paradigms speaking broadly, and in the present day and age, one can observe that all of them co-exist, although the amount of research focus varies across paradigms. Also, depending on the type of application, one paradigm may be more suitable than others for image analysis and restoration:

1. Signal paradigm: This includes methods which operate based on the properties of the image signal, sometimes through spectral analysis.
2. Artifact paradigm: This includes methods which tie image quality to the defects or degradation which happen as a result of, for example compression, like JPEG blocking artifacts. Sometimes, they also take into account the image's perceptual quality.
3. Information paradigm: The third category includes methods which consider the image capturing process as information transfer through a channel between the source (real world) and the receiver (camera sensor) often using information theoretic tools like entropy and channel capacity for image quality analysis and restoration.
4. Aesthetics paradigm: This is arguably the most difficult category and aims to quantify and enhance the aesthetics of the image. This is an interdisciplinary exercise at the intersection of diverse fields including psychology, biology and philosophy.

We consider few challenging problems belonging to the first two paradigms, provide a short summary of their significance and challenges and summarize why existing methods are insufficient or unsuitable for addressing those problems. Note that, the research to tackle any particular problem in its entirety

can span across more than a single paradigm. Still, the research can be characterized by its core functionality, to be dominantly aligned to one paradigm. Next, we summarize the scope of application of our research to address the aforementioned challenges, leading to a summary of our contributions. We conclude by describing the organization of subsequent chapters of this thesis.

1.2 Signal paradigm problems

1.2.1 MRI abnormality detection

Significance and challenges due to scalability requirements and limited contextual information availability

The first problem we investigate in the *signal* paradigm deals with Magnetic Resonance Imaging (MRI) of preterm infant brains. Preterm infants are those born prematurely, before the normal gestational period of 36 weeks. They are at high risk of developing brain abnormalities (lesions) like White Matter Injury (WMI) or Periventricular Hemorrhage (PVH) which can cause severe neuro-developmental deficits later in life. Detecting brain injuries in adults is a well researched problem because of readily available brain tissue segmentation maps called brain atlases. However, preterm infants have underdeveloped brains whose structure changes rapidly. Thus, it is difficult to obtain reliable brain atlases. Ideally, we have to be able to detect the aforementioned brain injuries without using any atlas. The sheer volume of brain MRI data obtained as hundreds of slices for a single scan of a single patient makes complete manual inspection by radiologists unfeasible. Apart from these clinical challenges, the signal processing challenges arise from the small size, patient movements during scanning etc. which yields a very low signal-to-noise ratio (SNR). The resulting distortions make it very difficult for a computational algorithm to detect the patterns in the MRI signal indicative of brain damage.

Key weaknesses of existing MRI abnormality detection approaches

Due to the practical problems mentioned above, although previous research has extensively studied automated feature extraction and segmentation in adult

brain images [60], [81], [105], the same is not true for preterm or even term neonates. Other recent works on detecting brain lesions from MRIs [63], [106] are not specific to preterm neonates. Some publicly available medical diagnostic software packages [113], [115] often have popular brain lesion detection methods embedded in them, but require multiple scans of the same subject using multiple modalities [113], [115] which are impractical to obtain due to the practical problems mentioned before. The work from Griffis et al. [47] for detecting ischemic stroke lesions can function based on a single sequence (T1) scan. T1 scans are characterized by short time intervals between successive radio-frequency pulse sequences (short Repetition Time, TR) and quick receipt of the echo after RF pulse delivery (short Echo Time, TE). However, Griffis' is a supervised method and requires training on expert annotated scans, which are difficult to obtain for preterm infants. Also, Griffis' work performs Montreal Neurological Institute (MNI) brain template based segmentation, which is possible for adult brains having a fixed structure, whereas registering preterm infant brains to fixed atlases is impractical due to their rapid structural changes. Our earlier algorithm [20] used a stochastic process to estimate intensity variation likelihood in target pixels belonging to a WMI. However, it requires an already segmented White Matter (WM) region and worked for high resolution and noise-free slices, thus rendering it impractical for preterm neonate MR scans. Moreover, attempts at detecting another related type of injury, Periventricular-Intraventricular Hemorrhage (PIVH) in preterm infants has recently started gaining increased attention. But such attempts have resorted to (complete or partial) manual examination by radiologists or subject matter experts [2], [5], [57], [77], [95]. However, huge volume of MRI data makes manual examination impractical, thus requiring efficient automated methods, which further motivates our research.

Scope of application of our research

Our research to address the aforementioned challenges amounts to a CAD system which can help physicians detect WMI/PVH from T1 sequence 2D DICOM MRI slices of preterm infant brains, without requiring any brain at-

las. Thus, our contribution is purely on the MRI processing/analysis using our developed software algorithms, and not concerned with improving MRI hardware components or optimizing MRI acquisition parameters to improve lesion detection/diagnosis. The MRI slices were provided by our collaborator in SickKids Hospital, Toronto.

Our research contributions

We developed the first fully automated single-sequence (T1) MRI WMI/PVH detection approach suitable for preterm infants which bypasses the requirement of traditional brain atlas-based accurate WM segmentation in the WMI/PVH detection pipeline. As explained earlier, since reliable preterm infant brain atlases are unavailable, this is an important contribution for preterm WMI/PVH detection, and also drastically reduces the execution time by skipping the computationally expensive WM segmentation and using heuristics instead.

1.2.2 InSAR signal recovery for WAM

Significance and challenges due to scalability requirements and limited contextual information availability

The second problem we investigate in the *signal* paradigm deals with Interferometric Synthetic Aperture Radar (InSAR) used to measure movement of the earth surface from satellite microwave imagery. Here, the challenges arise from the unavailability of clean InSAR images due to the contamination of the reflected microwave signals during their passage through the atmosphere. Several factors like turbulence and moisture are responsible for such contamination. Yet another factor is loss in correlation between chronologically ordered images of the same location on the earth's surface due to human activities like passing vehicles or construction work or natural phenomenon like random scattering of microwaves by vegetation and water bodies. Thus, in order to improve the usability of the highly contaminated microwave images, we need to recover the underlying true signal and quantify the extent of signal recovery for each point of the image (coherence). We attempt to develop a solution for monitoring ground movements for a wide area (WAM) which refers to a

landmass having area comparable to a province (state) or even a country. For such applications, data volume becomes a prime concern. Even for monitoring a localized area like a mining site (few square kilometers), the images have billions of pixels, thus mandating scalable solutions via parallel processing.

Key weaknesses of existing InSAR signal recovery approaches

Research in this area started with the simple boxcar filter, and is still widely used today. Boxcar filtering involves computing moving average over a rectangular window. However, InSAR-acquired ground imagery are non-stationary (constantly change) due to topography variations and line-of-sight land displacements. This makes sample average methods like boxcar unsuitable for InSAR denoising [114]. Additional limitations arise from strong smoothing effect of filtering and loss in spatial resolution, ultimately resulting in noticeable errors in phase and coherence estimation near signal discontinuities. Subsequent methods which tried to address this problem may be broadly classified as spatial methods, e.g., Lee [70], and frequency based methods, e.g., Goldstein [44] and their numerous improvements [17], [41], [61], [120], [133], [137]. They are adaptive to the local fringe direction and/or local noise. Frequency based methods gradually evolved into the wavelet domain [10], [75], [139] as wavelet based filtering helps preserve spatial resolution. Local modeling based [12], [53] methods and Markov Random Field (MRF) based methods [31], [37] have also been explored, although MRF requires modeling the prior distribution, which is a challenging problem. Recently researchers have achieved promising results in non-local filtering of various types of SAR images such as amplitude, interferometric phase, polarimetric and multitemporal stack images [28], [117]. However, for wide area monitoring using InSAR images, we have to process a huge volume of data which necessitates parallel processing. In recent years, efficient GPU-based software architectures have made extremely fast inference possible for neural network models. For image processing problems in particular, Convolutional Neural Networks have shown promising results in visual pattern recognition. While despeckling of SAR images [129], [134], [135], [140] and geo-localization accuracy improvement of optical satellite images [82] us-

ing neural networks have been attempted, CNN-based learning approaches to InSAR images have not been explored. This motivates us to design CNN-based scalable solutions for InSAR phase filtering and coherence estimation to exploit the scalability of massively parallel GPU architectures for WAM.

Scope of application of our research

Our research to address the aforementioned challenges amounts to replacing the filtering and coherence estimation (using traditional methods) stage of the InSAR processing pipeline with our developed scalable software algorithms based on unsupervised learning using CNNs, and subsequently, generative modelling. Thus, our direct contribution is to one particular stage of the InSAR processing/analysis, and not concerned with improving InSAR hardware components or optimizing InSAR acquisition parameters to improve processing. However, filtering and coherence estimation are among the most critical stages of the pipeline, as outputs of our developed algorithms (filtered phase and coherence) are input to the next stage (unwrapping), which computes the distance of ground points from the InSAR sensor mounted on the satellite. Thus, it has significant influence on later stages of the pipeline and final outputs, like accurate DEM generation or estimation of geophysical parameters.

Our research contributions

We developed the first CNN-based phase filtering and coherence estimation solution for InSAR and subsequently improved it to develop a generative modelling based approach. Both developed algorithms are based on unsupervised learning on noisy InSAR data, and thus do not require training on clean InSAR images which are unavailable as such. Traditional algorithms in this area had not been designed by prioritizing scalability, which is a primary requirement for our specific use case (WAM). Also, generative modelling is an improvement over traditional approaches in that, for the input noisy image, the trained model outputs a Gaussian distribution for the predicted clean image. Thus, we can sample from this distribution to generate new images, which are slight variations of the clean image. This can be used in InSAR machine

learning for data augmentation [116], or to test InSAR processing chains by running the same chain all the way through but with slightly different images (interferograms) to measure the variance of the outputs of the complete processing chain. This can potentially be a good error analysis method.

1.3 Artifact paradigm problem

1.3.1 GPU-friendly debanding for mobile HDR

Significance and challenges due to scalability requirements and limited contextual information availability

The problem we investigate in the *artifact* paradigm deals with perceptual masking of banding artifacts (due to quantization) in High Dynamic Range (HDR) displays, tailored to the mobile Graphics Processing Unit (GPU) environment. HDR displays are becoming ubiquitous nowadays, even in handheld consumer electronic devices like mobile phones. HDR images have 10+ bits per pixel, compared to the traditional Low Dynamic Range (LDR) devices which have 8 bits per pixel. However, for backward compatibility with the majority of available visual content (which is still largely LDR), bit depth conversion in software and hardware often result in quantization artifacts when the content is played back on HDR screens. The specific type of artifact we address is banding artifacts which are characterized by bands of color instead of smooth gradients, in low-texture image regions like the sunset sky. Traditional solutions to banding relied on neighborhood operations like filtering, which is not ideal for GPU environments. In GPUs, independent operations on individual pixels are preferred, rather than pixel neighborhood based filtering operations. Traditional pixel-based solutions like dithering assumed availability of the original, unquantized signal. This is impractical in many real-world scenarios where only the quantized signal is available. Thus, for debanding on the mobile GPU platform, we require a pixel-based solution which works directly on the quantized signal.

Key weaknesses of existing debanding approaches

A popular approach to debanding is adding noise (dithering). Traditional dithering methods can be classified as adaptive or non-adaptive depending on whether or not they use pixel neighborhood information, but both types require access to the original, unquantized (high bit-depth) pixel values, as they perform dithering before or during quantization [8], [38], [40], [51], [59], [62], [89], [90], [103], [125], [131], [132]. However, for certain applications which require image quantization or compression, only the quantized data is available at the dithering stage. Examples of such scenarios include a video decoder which decodes a 10-bit bit-stream but outputs a 8-bit signal, or when a 8-bit signal is received. If we have the unquantized data available, adding zero-mean noise can remove banding, but not from quantized signal. Also, for our application area of mobile GPUs, the computation capacity is limited and accessing neighborhood information is costly in terms of memory I/O and timing. Thus filtering and false contour detection based approaches are not suitable [9], [26], [55], [69], [126], [127].

Scope of application of our research

Since de-banding is a mature research area, it is important to note that our dithering solution is applicable to specific types of de-banding scenarios, as explained earlier. Also, our research does not concern itself with specific details of the iTMO, and can be used with any iTMO as such. We only assume that the iTMO can be formalized as a one-to-one mapping between its input and output code-words and can thus be pre-loaded and stored in memory as a BLUT, to be subsequently accessed by our developed algorithm at run time.

Our research contributions

We developed the first dithering solution to address de-banding use cases where one has access to *only* the quantized signal and must refrain from accessing neighborhood information due to architectural and computational constraints. Our solution is structured as two independent parts. The first part deals

with offline generation of novel noise patterns which are effective in perceptual masking of banding artefacts and blend well with the scene, so as to not cause visual irritation to human observers. The second part deals with online injection of those noise patterns modulated by the slope of the iTMO, besides other factors. The modular nature of our solution facilitates future researchers to use our noise patterns in conjunction with their own noise injection solutions, and thus potentially opens up new avenues of research.

1.4 Organization of the thesis

The rest of this thesis further motivates and describes our developed algorithms and validation results, and is organized as follows: Chapter 2 is a short summary of few parametric and structural optimization based image segmentation algorithms to help the reader understand some related literature reviewed in Chapter 3 and the segmentation algorithm used in Chapter 4. Chapter 3 describes the first published algorithm to detect WMI and PVH from single sequence (T1) MRI in preterm infants without using brain atlas. Chapter 4 describes the first published application of CNNs to InSAR phase filtering and coherence estimation and its extension to generative modelling. Chapter 5 describes the first de-banding method in literature which uses dithering directly on the quantized image without accessing any neighborhood information. Chapter 6 has our concluding remarks and directions of future research.

Chapter 2

Segmentation via Parametric and Structural Optimization

In this chapter, we briefly discuss few image segmentation algorithms based on parametric and structural optimization (hereafter referred to as PSO). Some of these concepts will help the reader understand the related work presented in Chapter 3 as well as the Chan-Vese segmentation algorithm used in the first method proposed in Chapter 4. We will also briefly discuss the rationale behind our choice of when to use (or not to use) PSO in our proposed algorithms.

PSO deviates from the simpler pixel-based (thresholding) segmentation algorithms like Otsu [94] in that PSO is performed by deforming a (set of) initial curve(s) based on certain criteria, rather than simply values of pixel intensities. Examples of such criteria include that the curve conforms to image edges and is a smoothly varying curve, as opposed to being extremely jagged. From this example, we can see that there is often a trade-off between the criteria, but overall, PSO is arguably more similar to how humans approach a manual segmentation task, by trying to draw shapes or contours around objects of interest in an image, rather than just trying to aggregate pixels having certain intensities. A pictorial depiction of how PSO works is shown in Fig. 2.1. In actual PSO implementations, the iterative process can be fixed to a certain maximum number of iterations, or allowed to continue till the shape of the curve does not change significantly, as defined by a tolerance value. At any iteration, an energy function quantifies how good or bad the curve is. The curve iteratively evolves to reduce/minimize this energy function, as explained

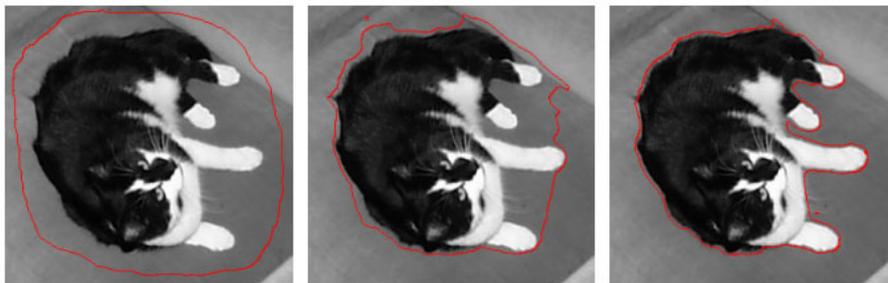


Figure 2.1: PSO deforms a rough initial contour (left) through successive iterations (centre) such that it finally delineates the object of interest (right).

in subsequent sections of this chapter, with reference to specific algorithms.

2.1 Snakes

Snakes start with the idea of a parametric representation of a curve. Instead of representing a curve conventionally as $y = f(x)$ (in the 2D case) where y and x are the dependent and independent variables respectively, snakes represent a curve as the x and y coordinates being functions $X(s)$ and $Y(s)$ respectively of a continuous parameter $s \in [0, 1]$ such that we can trace out a closed contour in the 2D plane by smoothly changing s . For example, if we wanted a circle of radius r , we would define $x(s) = r \cos(2\pi s)$ and $y(s) = r \sin(2\pi s)$. One benefit of this is that it allows us to represent arbitrarily shaped curves more efficiently than the conventional Cartesian representation. We define the Energy E as a function of the curve c to have an internal part which depends only on the shape of the curve and an external part which depends on image intensities, thus balancing between the two criteria we mentioned earlier, as in Eq. 2.1.

$$E(c) = E_{internal}(c) + E_{external}(c) \quad (2.1)$$

The internal energy of the curve can in turn be represented as the sum of the single and double derivatives of the curve. The former ensures that the curve is not stretched too much (low elastic energy) while the latter ensures that the curve does not change its curvature abruptly, as shown in Eq. 2.2, where $c(s)$ represents the curve and α and β balance the effects of the derivatives.

$$E_{internal}(c) = \int_0^1 \alpha \|c'(s)\|^2 + \beta \|c''(s)\|^2 ds \quad (2.2)$$

The external energy on the other hand is designed to take on low values when the curve coincides with strong edges in the image, as shown in Eq. 2.3.

$$\begin{aligned} E_{external}(c) &= \int_0^1 -\|\nabla I(c(s))\|^2 ds \\ &= \int_0^1 - \left[\left(\frac{\partial I}{\partial x}(x(s), y(s)) \right)^2 + \left(\frac{\partial I}{\partial y}(x(s), y(s)) \right)^2 \right] \end{aligned} \quad (2.3)$$

Deriving the details of the process of minimizing $E(c)$ involves variational calculus and is beyond the scope of this discussion. However, we present some intuitions of how the algorithm works in practice, as in the real world, we have the discrete pixel grid instead of the theoretical X-Y plane. A curve is thus parameterized by a set of K points on the grid (connected by straight lines) and the greater the number of points, the smoother the representation of the curve, at the cost of higher computational complexity. We thus express the curve as $c(s, t)$ where $s \in [1, 2, \dots, K]$ is the index of the point on the curve and t is the iteration number, as the curve evolves taking steps from $c(s, t-1)$ to $c(s, t)$ along the gradient of $E(c)$ until the boundary points stop changing.

However, the snake method does not detect strong edges that are at a distance from the curve, and it is sensitive to noise which creates small gradients and confuses the method. A solution is to slightly blur the image as on one hand, this decreases the effect of noise and on the other hand, distributes strong gradients (edges) to the adjacent image regions. However, this solution does not work when the initial contour is too far away from the edge.

2.2 Active Contours

A better solution to the problem described above is a technique called Gradient Vector Flow (GVF). Here, in addition to using the magnitude of the image gradient $e(x, y)$ (edge map), we create a new vector field over the image plane $v(x, y)$ with the x-component $v_x(x, y)$ and y-component $v_y(x, y)$ as in Eq. 2.4,

where v is defined to minimize the term shown in Eq. 2.5, with ∇_e giving the magnitude of the edge map, and μ serving as a tuning parameter.

$$v(x, y) = \begin{bmatrix} v_x(x, y) \\ v_y(x, y) \end{bmatrix} \quad (2.4)$$

$$\iint \mu \left[\left(\frac{\partial V_x}{\partial x} \right)^2 + \left(\frac{\partial v_x}{\partial y} \right)^2 + \left(\frac{\partial v_y}{\partial x} \right)^2 + \left(\frac{\partial v_y}{\partial y} \right)^2 \right] + \|\nabla_e\|^2 \|v - \nabla_e\|^2 dx dy \quad (2.5)$$

The intuition here is that when ∇_e has a large value (strong gradient, edges in image) v simply follows the edge gradient, and when ∇_e has a small value (weak gradient, flat image regions), v falls off smoothly. This effectively causes the strong edges in the image to get diffused out to guide the curve's evolution.

This method works better than the basic snakes method, but fails in cases where the immediate background of the object of interest has strong texture. There are also concerns which apply to both methods. Firstly, the method to initialize the curve is left unspecified. Secondly, in practical implementations, we have to keep track of the number and distribution of points on the curve and make sure that the points probe into the concave boundaries of the object of interest. Lastly, a single snake cannot segment (wrap into) multiple objects of interest at once, or segment the inner boundaries of objects with holes.

2.3 Level Sets

To solve the above problems in a better way, instead of parameterizing the curve with a set of points, we can consider the whole image plane as a discrete grid structure (x, y) and define a level-set function $\phi(x, y)$ and instead of evolving a curve, we evolve this function. Pixels (C) where $\phi(x, y) = 0$ implicitly define our objects of interest, and this can include multiple objects or holes inside objects, as in Eq. 2.6. The advantage of this is that since we evolve the whole function over the entire X-Y plane, we can change the topology of our target curve(s) on the fly, instead of starting off with initially assumed curve(s). This formulation helps discover our target curve(s) over iterations.

$$C = \{(x, y) \mid \phi(x, y) = 0\} \quad (2.6)$$

Thus, $\phi(x, y)$ can be a signed distance function: for points inside a region it would be negative, zero on the boundary and positive outside the region. In practical implementations, like earlier, we can evolve the function $\phi(x, y, t)$.

2.4 Chan-Vese algorithm

All the above (edge-based) methods assume that somewhere in the image there are strong edges, but in some practical scenarios this assumption might not hold, so we need a region-based algorithm like Chan-Vese. It assumes that the image can be partitioned into two regions, one inside the curve and the other outside, and has an energy function that ensures that the variance of pixels inside and outside the curve are low, and favors a short curve with small area, as in Eq. 2.7 where I denotes inside pixels with mean μ_I , O denotes outside pixels with mean μ_O , $I(x, y)$ denotes image intensity at coordinates (x, y) , and $L(c)$ and $A(c)$ denote the length and area of curve c respectively. The λ 's balance the effects of the four terms in the energy function $E(c)$. The first two terms in Eq. 2.7 try to maximize the difference between distributions of pixels inside and outside the curve. Thus, Chan-Vese succeeds even in those scenarios where the object of interest has a strongly textured background.

$$E(c) = \lambda_1 \int_I (I(x, y) - \mu_I)^2 dx dy + \lambda_2 \int_O (I(x, y) - \mu_O)^2 dx dy + \lambda_3 L(c) + \lambda_4 A(c) \quad (2.7)$$

Relevance to Thesis

All above methods are slow, as they evolve the curve over successive iterations based on multiple considerations. Chapter 3 shows how this affects our earlier WMI detection algorithm [20] based on level sets. However, Chapter 4 shows Chan-Vese can effectively segment fuzzy raw coherence maps while preparing the training data (offline) and not during the actual coherence prediction.

Chapter 3

MRI abnormality detection

3.1 Introduction

In this Chapter, we will show how earlier algorithms, e.g. level set, are insufficient to address the MRI abnormality detection problem.

3.1.1 MRI basics

Magnetic Resonance Imaging (MRI) is a medical imaging technology used by clinicians to see soft tissues such as muscles, blood vessels, ligaments and internal organs. MRI scanners capture sections through the body which can be used to make three-dimensional images. For example, Fig. 3.1 shows a 3D reconstruction of the structures inside an adult human head obtained using MRI. An MRI machine uses a powerful magnetic field to align the magnetization of some atoms in the body. Radio frequency fields systematically alter the alignment of this magnetization. This causes the nuclei to produce a rotating magnetic field detectable by the scanner. This information is recorded to construct an image of the body. Images are constructed when protons in different tissues return to equilibrium state at different rates. MRI machines look like a large block with a tube running through the middle of the machine, called the bore of the magnet. The bore is where the patient is located for the duration of the scan, as shown in Fig. 3.2. The MRI machine picks points in the patient's body, decides what type of tissue the points define, then compiles the points into 2D and 3D images. Once the 3D image is created, the MRI machine creates a model of the tissue. This allows the clinician to diagnose

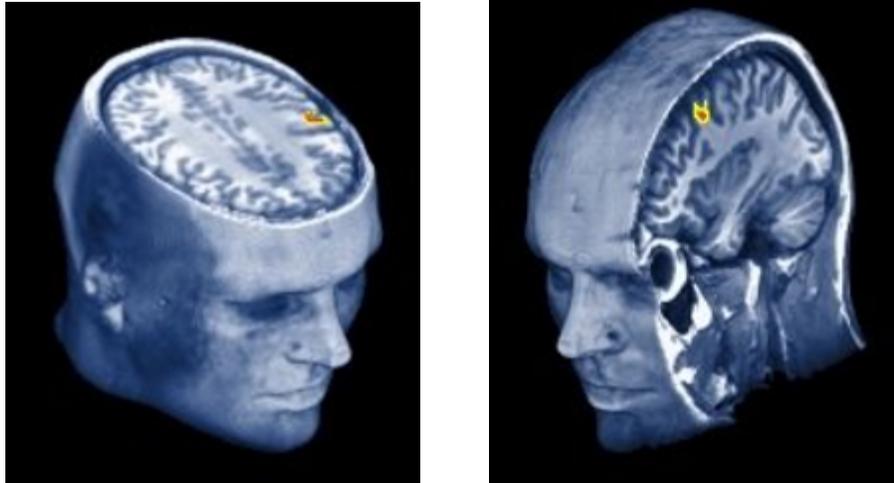


Figure 3.1: 3D reconstruction of adult human head from MRI.

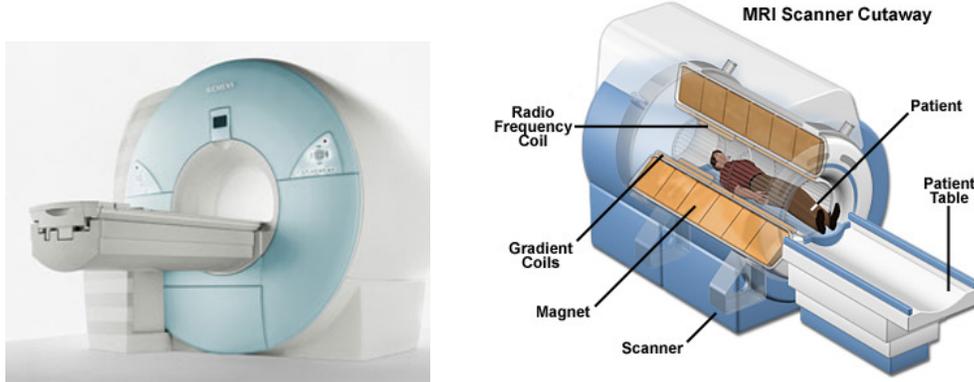


Figure 3.2: Typical MRI scanner with magnetic strength = 3 Tesla.

without the use of invasive surgery. MRI is considered ideally suited for soft tissue problems, such as diagnosis of multiple sclerosis (MS), brain tumours, spinal infections etc. Advantages of MRI include no radiation exposure, and good grey-white differentiation, which refers to the appearance of the interface between cerebral and cerebellar white matter and grey matter. A weakness of MRI is that scans require patients to hold very still for long periods of time, up to 90 minutes or more in some cases. The procedure is done in multiple parts, and takes time to switch between different scans and fields of view.

3.1.2 Motivation and Clinical Significance

The premature infant brain's immaturity makes it inherently more susceptible to injury [118] often leading to developmental deficits or even fatalities in subsequent stages of life [77]. Degree of prematurity of the infant is positively correlated with likelihood of brain damage. Near the center of brain, in each half, there exists a C-shaped cavity having cerebrospinal fluid, called the lateral ventricles. The area of the premature infant brain most vulnerable to injury is the periventricular area, which is a brain tissue rim lining the outside of each lateral ventricle. A dense network of thin, fragile (prone to rupture) blood vessels called capillaries are present in each periventricular area. Unfortunately, more premature infants have more of these capillaries [5]. During premature birth, the fetus is abruptly expelled from the controlled environment of the uterus to the hostile, highly stimulating environment outside. The resulting shock and physiological stress causes the capillaries to rupture. Initially, periventricular hemorrhage or PVH occurs as the immediate periventricular areas start bleeding. Continued bleeding creates an expanding volume of blood which severs the adjacent lateral ventricles, thus causing an intraventricular hemorrhage or IVH.

The periventricular areas serve two important purposes:

1. The brain's outer layers (cerebral cortex) is formed by new brain cells which form and develop in the periventricular areas, and gradually move to form the outer layers. The cerebral cortex handles several vital functions of the brain, including learning and intelligence, behavior and personality, and speech. It also has considerable influence on control and strength of muscles.
2. Periventricular areas serve as passages for motor nerve signals to the muscles originating from the cerebral cortex.

PVH/IVH may cause brain injury via several mechanisms. If damaged new brain cells developing in the periventricular area ultimately move outward and get embedded inside the cerebral cortex, they impair the functions and

distort the structure of the cerebral cortex. Unfortunately, the infant does not even start using various parts of the cerebral cortex for months or even years after birth. Thus, the developmental issues arising out of the aforementioned damages to the cerebral cortex are not even apparent until later stages of life. Nonetheless, the implications of cerebral cortex damage are debilitating: learning and language difficulties, arrested mental development, behavioral and personality disorders, convulsions, etc.

MR studies of neonatal periventricular-intraventricular hemorrhage (PIVH) are scarce. Acute MR findings of adult and neonatal hemorrhage are considered similar. However, the medical community believes that, due to rapid myelination and maturation of the premature infant brain, subacute and chronic MR appearance of hemorrhagic evolution in premature infants are possibly quite distinct from those in adults [2].

All these emphasizes the need for both early detection and long-term developmental follow-up for high-risk infants [57], [95].

Early detection is possible by examining images of high-risk infants' brains acquired using scanning devices like Magnetic Resonance Imaging (MRI). To make this possible, the different types of brain tissues in an image need to be identified in order to look for abnormalities.

Another recent research investigated the clinical significance of Injuries to the White Matter (WM) tissues of the brain (WMI) in 216 pre-term neonates [48] where correlations were found between WMI volume and location and neurological developmental disorders. WMI volume and location was quantified in the test subjects' MRI following manual WMI segmentation. Then, the subjects were tracked over an 18-month period via assessment of their motor, cognitive and language abilities. The chief findings from their study were:

1. Irrespective of lesion location, greater WMI volumes predicted poor motor outcomes.
2. Greater WMI volumes in frontal, parietal, and temporal lobes have adverse motor outcomes.
3. Only frontal WMI volumes predicted adverse cognitive outcomes.

4. Frontal lobe lesions predict adverse cognitive and language development.
5. Adverse motor outcomes are predicted by widespread injury.

An important conclusion from this study was that the volume of frontal lobe lesions were the most predictive of the neuro-developmental outcomes. Hence, it is important to detect and localize WMI in pre-term neonates.

3.1.3 Computational and Signal Processing Challenges

Brain tissue segmentation generally refers to the separation of the brain into three functional components; namely, Grey Matter (GM), White Matter (WM) and Cerebro-Spinal Fluid (CSF). Segmentation is often performed as the first step in detection of physiological abnormalities, in volumetric study and diagnostic analysis [32]. Seizures, strokes, brain infections and injuries are often hard to determine by manual examination of image scans, since significant features may not be obvious on a 2D DICOM slice. Also, manual examination can be subjective; and interpretation may vary from one expert to another. Even the same expert may make different decisions at different times. Thus, the commonly adopted gold standard in the clinical community is to obtain at least two consistent interpretations out of three expert opinions. However, getting datasets of preterm neonates with ground truth annotation by multiple clinical experts is not feasible in practice.

To address this issue, Computer Aided (or Assisted) Detection (CAD) has gained increasing attention in clinical research. For example, semi-automated approaches, like the Clusterize algorithm [22], have been applied to identify brain lesions in stroke victims and have been shown to significantly speed up lesion demarcation without loss of precision and reproducibility [49]. Computer programs designed for CAD [15] aim to detect potential abnormalities, such as Brain Tumors and Multiple Sclerosis brain lesions [11] by identifying suspected features on the image for further inspection by a radiologist. The benefit of CAD is two-fold: (1) When there is a backlog, e.g., in looking through X-ray films, due to an insufficient number of radiologists, CAD is used to pre-screen and thus reduce the workload of radiologists; and (2) when a large sequence of

image scans need to be compared or examined as an integrated volume, CAD helps to detect patterns that are difficult if not impossible for human eyes to comprehend. CAD software can pin-point the areas of concern. Radiologists can then focus on these identified regions to arrive at the diagnosis.

Automated feature extraction and segmentation in adult brain images have been extensively studied over the past years [60], [81], [105]. However, the same is not true for neonates, whether preterm or term, because of several practical challenges in obtaining and analyzing MR images of neonate brains, including:

1. Lack of a reliable anatomical map, or atlas for an neonate brain, to guide the segmentation process in areas of low contrast and help distinguish tissues of similar intensities. Even when such atlases are available, they need to be registered onto the test MR image, which is a difficult process, because the neonate brain undergoes rapid structural and physiological changes during maturation.
2. Neonates tend to move during the MRI scan process, which is highly sensitive to patient movements. Motion artifacts, blurring, etc. degrade the quality of the MR images.
3. Neonate brains are small in size and the duration for which a neonate is scanned is also shorter than adults. This results in a low Contrast-to-Noise ratio (CNR), low Signal-to-Noise ratio (SNR) and low spatial resolution.
4. Contrast between grey matter (GM) and white matter (WM) in both T1- and T2-weighted images (T1w and T2w) is different from the adult brain. Most parts of the neonate brain are non-myelinated at birth, where WM appears less intense in T1 images and more intense in T2 images, whereas this trend is reversed for a fully myelinated adult brain (contrast inversion).

Thus, a WMI detection algorithm for pre-term infants should not be overly dependent on an atlas. It should work on noisy, low-resolution images with

motion artefacts, and be automated as much as possible (least human intervention), so that the algorithm can integrate well with the CAD pipeline.

In preterm neonates, the characteristic of brain injury is multi-focal WMI in the first weeks after birth, where using T1 weighted (T1w) MR images of the neonate’s brain for detection is more effective [18], [85]. Hence, in this work we adopt the atlas-free approach to analyze T1 weighted (T1w) MR images. We anticipate that this method will also be relevant to detecting WMI in elderly patients with leukoariorosis [24].

3.1.4 Our Computational and Clinical Contributions

Our most important advancement with respect to related work and our earlier pre-term WMI detection algorithm [20] (both described in the next section) is to eliminate the dependence of the WMI detection step on a pre-segmented WM region. As explained earlier, this gives primarily two important benefits:

1. The execution time for WMI detection decreases drastically, since we do not perform the time-consuming segmentation of the white matter region. This is the main contribution of our present work from the computational perspective.
2. Since we bypass segmentation, no brain atlas is required. For pre-term infants, it is difficult to obtain reliable brain atlases. Thus, obviating the requirement of brain atlases for pre-term infant WMI detection is our clinical contribution.

Moreover, although PIVH in preterm infants has recently started gaining increased attention, attempts at detecting such abnormalities have been (fully or partly) through manual examination by radiologists or subject matter experts [2], [5], [57], [77], [95]. However, such manual examination has its own limitations, as outlined earlier.

Our proposed approach is the first fully automated method to detect periventricular hemorrhage from preterm infants’ brain MR images.

It should be noted that our proposed method is different from existing atlas-free segmentation methods in literature which use local contrast and

geometric traits, brain morphology and tissue connectivity to guide segmentation. The strength of atlas-free methods is that they can accommodate changes in anatomy of the developing (neonate) brain, as they are not bound by constraints imposed by the atlas. However, their main weakness is the computational complexity of the segmentation process itself. We bypass segmentation to overcome this weakness.

3.1.5 Differences with Related Work

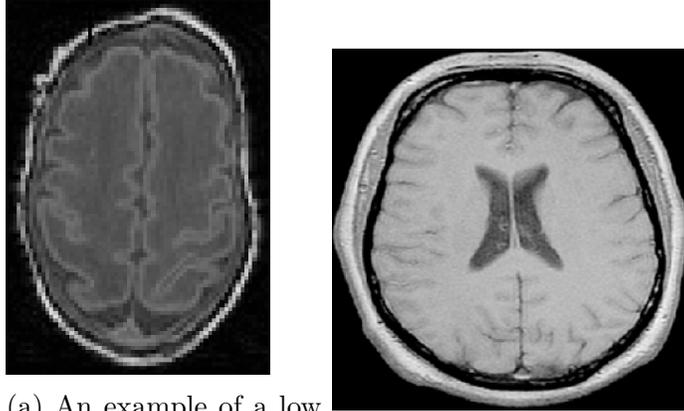
As described later, one of the steps in our proposed method involves the localization of ventricles as a collection of blobs detected using the Maximally Stable Extremal Regions (MSER) algorithm [78] and optimized using Genetic Algorithms (GAs) [43]. MSER or its modified forms have been used earlier to detect various retinopathy pathologies [110], segment ultrasound liver images [141], localize cell nuclei in microscopic images [121], isolate fetal brain tissues from maternal anatomy during fetal brain in-utero MR imaging [64] and for 3D segmentation of simulated brain MR images [33]. However, they have not been tested in preterm brain WMI detection. The main advantage of the MSER algorithm is that there is no need to specify an initial contour, which is necessary and often drawn manually in other algorithms. For example, brain tissue segmentation approaches based on Active Contour Models [72], [86], [98], [108], [109] require an initial contour. Furthermore, the region stability of MSER is constrained by local information obtained in the neighbourhood and can accommodate large intra-image variations [121].

Medical images often have poor image contrast and are associated with artifacts that result in missing or diffuse organ/tissue boundaries. The resulting search space is therefore often noisy with a multitude of local optima. Genetic Algorithms (GAs) benefit medical image segmentation [80] as they are less prone to get stuck in a local optima. GAs have been used in a learning-based approach to segment and label numerous neuroanatomic structures including left/right and third ventricles [122]. Their approach was based on observer-defined contours of neuroanatomic structures, which were used as a priori knowledge. However, in the context of preterm neonate brain WMI

detection, it is not possible to obtain sufficient number of expert-annotated training images (for learning or validation). A variant of GAs called parallel genetic algorithms have been used earlier for volumetric segmentation of lateral ventricles [35] on simulated Brainweb images, but not on preterm brain images. Their strategy for choosing the initial population involves deriving an initial surface by segmenting the ventricle slice-by-slice (using a 2D method), and then solving an evolution equation (formulated using that initial surface) using a finite-difference method, whose result is used to generate the initial population for the GA. In contrast, our initial population for the GA includes all regions detected by MSER on individual (2D) slices.

Ortiz et al. [93] applied an atlas-free fully automated method to segment brain MRIs into different types of tissues. During pre-processing, they removed the background noise in the image using Otsu’s method by minimizing the intra-class variance of the signal and noise voxels. Then, 24 important 1st order statistical features (intensity, mean and variance) and 2nd order statistical features (energy, entropy, contrast etc.) were extracted. Most discriminatory features were selected using a Genetic Algorithm. Next, a Self-Organizing Map (SOM) was trained using the selected features in an unsupervised way. A label representing a type of brain tissue was then assigned to each SOM unit. Their method performed better than Constrained Gaussian Mixture Model [46] in classifying WM and CSF, and gave promising results on high-resolution MRIs. Although this method is atlas-free, it does not work well on low-resolution preterm neonate MRIs. An example of a noisy low resolution (96×112) preterm neonate MRI used to test our method is shown in Fig. 3.3a, compared to a relatively noise-free image of much higher resolution (512×512) used by Ortiz et al. [93] shown in Fig. 3.3b.

Farzan [36] first segmented the brain tissues into GM, WM and CSF using Bayesian segmentation and then improved the results using domain knowledge obtained from experts in the form of heuristics. For applying the heuristics, the eight adjacent pixels of each pixel (its neighbours) were considered. A sample heuristic is: If “neighbours are WM,” then “new centre is WM.” The method assumes Normal Distribution of grey values in all tissues and uses



(a) An example of a low resolution 2D slice from our preterm neonate dataset. (b) An example of high resolution slice used by Ortiz et al. [93]

Figure 3.3: Comparison of image qualities to illustrate the challenge in segmenting low resolution preterm neonate brain MR images.

Expectation Maximization (EM) to maximize the likelihood probability of tissues. They compared their outputs against the expert-annotated versions in terms of sensitivity and specificity of each tissue type. Although the use of heuristics to classify tissues is the strength of this method, the target feature on a single 2D slice may not be obvious due to noise or artifacts. To address this issue, we analyze adjacent slices to validate and recover candidate features.

In [74] the authors applied skull-stripping as the initial step in the brain MR image segmentation process. The proposed hybrid skull-stripping algorithm, based on the Adaptive Balloon Snake (ABS) model has two steps: (1) Pixel clustering using Probabilistic Fuzzy C-Means (PFCM), which outputs a labelled image to identify the brain boundary; followed by (2) a contour initialized outside the surface of the brain. This contour is evolved guided by an ABS model. However, the ABS method employed here has the limitation that it ends up segmenting the contours into multiple objects. In comparison, our method effectively avoids the skull-stripping step and can discriminate irrelevant regions and false positives using their distance from the brain boundary.

In [58] the authors propose MSmetrix, which is an automatic MRI-based lesion segmentation method. This method is acquisition device independent, meaning that no parameter tuning is needed for the type of scanner deployed.

Differing from previous work [71], where multi-channel images were used simultaneously for lesion segmentation, their approach incorporates human expert input. By using T1-weighted and FLAIR images independently, they aim to fully exploit the characteristics of each sequence. Their unsupervised approach segments 3D T1-weighted and FLAIR MRIs into WM, GM and CSF following a probabilistic model, and treating WM lesions as outliers. The method assumes Gaussian distribution of the image intensities for each tissue class. However, this method uses MNI-atlas for skull stripping and GM, WM, CSF classification. Thus, it is not suitable for preterm neonates, where no reliable atlas is available as explained earlier.

There exist other very recent works on detection of brain lesions from MRIs [63], [106], but they are not specifically for preterm neonates. Some popular brain lesion detection methods are embedded in publicly available medical diagnostic software packages [113], [115], but they require multiple scans of the same subject using multiple modalities [113], [115]. Our method only needs a single sequence and takes T1 images as input. An example of using T1 images is the work of Griffis et al. [47]. It is a supervised method for detecting ischemic stroke lesions in T1-weighted MR scans using a naive Bayes classifier, which is trained on expert annotated scans. In contrast, our method is unsupervised and requires no training. Also, Griffis’s work [47] performs probabilistic segmentation of the scanned MR slices into four tissue classes (GM, GM, CSF and non-brain tissues). The output is subsequently normalized to the Montreal Neurological Institute (MNI) template space using the New Segmentation tool implemented in SPM12 [123]. Although their MNI atlas cannot be used to register preterm brains because of the rapid structural changes in preterm brains as explained earlier, we still tried out their segmentation step on our dataset to analyze its time performance. Their method took 1 minute and 17.5 seconds to complete, whereas our proposed method took 41.5 seconds. This comparison further validates the fast time performance of our proposed method. Our time gain is attributed by skipping the segmentation of WM, and instead, approximating the normal range of WM intensities by collecting samples from the WM region.

Relatively less research has been done on WMI detection in preterm neonates compared to WMI in adults and other tissue abnormalities. In our earlier algorithm [20] we use a stochastic process that estimates the likelihood of intensity variations in target pixels belonging to a WMI. First, it detects the boundaries between normal and injured regions of the white matter. Next, it measures pixel similarity to identify WMI regions. While the results showed effective WMI detection, the experiments were performed on relatively high resolution and noise-free slices, which may not often be the case for preterm neonate MR scans. In fact, as we will show later, when this method is tested on low-resolution noisy datasets, its accuracy is considerably lower. Also, the WMI prediction was done on individual slices, without considering adjacent slices. As we will show later, aggregating detection results of adjacent slices to predict WMI is crucial in low-resolution noisy scenarios, where the likelihood of detecting false positives and missing targets on a single slice is high.

3.2 Proposed Method

Our method considers both 2D and 3D spatial correlations. To give a high-level overview of our proposed method: we first detect potential WMIs in each 2D slice of the DICOM volume (coarse detection). We then analyze the WMI pixel correlation between each 2D slice and its adjacent slices in the DICOM volume. In the Fine Detection process, the Coarse Detection result obtained from a 2D slice is then validated with its adjacent slice information, in order to reduce false positives and recover true positives. An overview of the proposed method is shown in Fig. 3.4. Note that n adjacent slices are defined in the computational model, and $n = 1$ is used in the current implementation.

3.2.1 Coarse Detection phase

Our central idea behind White Matter Injury detection is to search for abrupt intensity peaks (hyper-intensities) in the white matter region of the brain. Since the intensities in the white matter are normally distributed [36], [58], the WMI represents an outlier with respect to the range of white matter intensities.

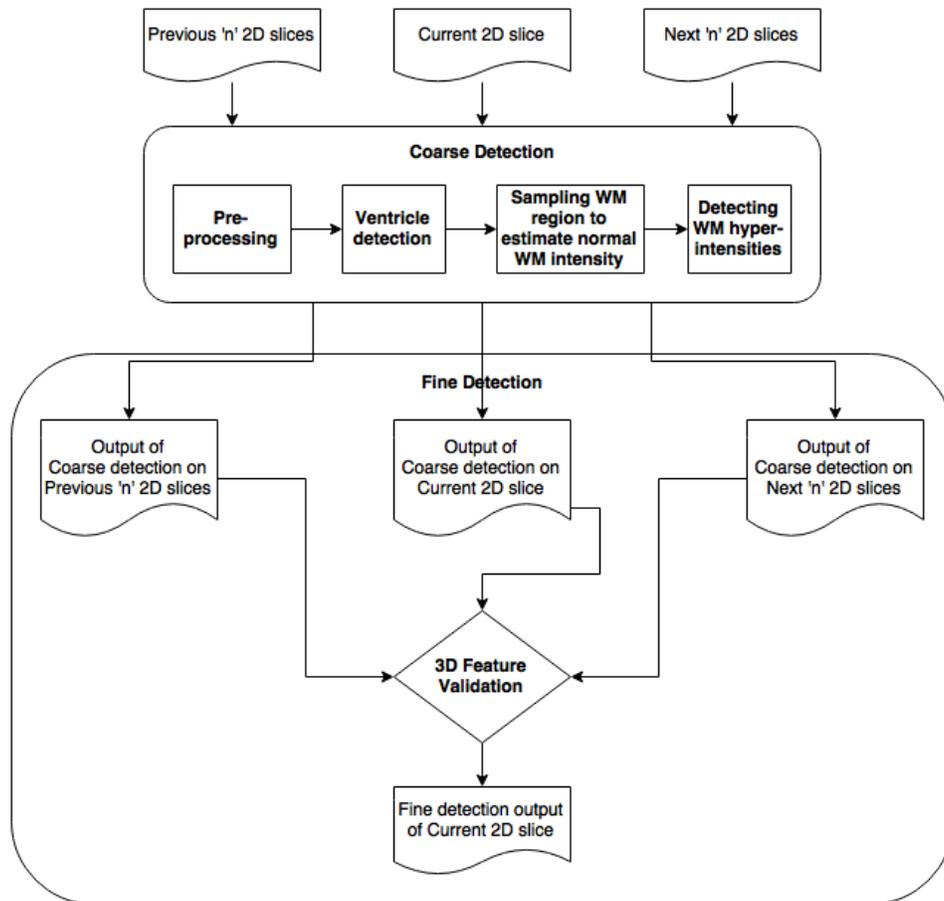


Figure 3.4: An overview of our proposed method.

As described later, we detect the ventricles as blobs in a 2D slice image. We then use a contour roughly between the ventricles and the brain boundary for estimating the range of white matter intensities. This is done by considering the intensities inside that contour which do not belong to the ventricles. The outlier intensities which are greater than the normal range would be potential candidates for WMI. In order to eliminate false positives, we further filter these candidates based on the size and distance criteria before a WM hyper-intensity is classified as WMI. All of these steps are performed on a single 2D slice (coarse detection).

Note that unlike traditional brain lesion detection methods, our method does not need to segment the WM into distinct patches. Instead, we sample the WM region to estimate the normal range of WM intensities. As shown later, this results in a significant reduction in execution time (without compromising accuracy) when compared to earlier segmentation-based WMI detection work [20]. However, if the ventricles are included in the estimation of WM normal intensities, the result will be unreliable. Thus, we identify the ventricles and eliminate them from our samples, using a confidence-based patch classification technique described later.

We next describe individual steps of the coarse detection phase in detail, followed by those of the fine detection phase. The reader can refer to Fig. 3.6 for outputs of individual steps of this entire processing pipeline.

Background segregation

Anisotropic Diffusion was used as a pre-processing step to de-noise those brain MRIs which were extremely noisy. The parameters used were $\frac{1}{7}$ for the integration constant, 3 for the gradient modulus threshold and the 2nd conduction coefficient function as proposed by Perona & Malik [97], because these parameter values, with 15 iterations, produced the best results.

Otsu’s method [94] was applied for separating the brain from the (noisy) background. Otsu’s method separates an input grey-scale image into foreground and background by determining a global threshold to minimize the intra-class variance of foreground and background pixels. It iterates through

all the possible thresholds in the image to find the threshold that gives the smallest within class variance.

Thus, by running Otsu’s method on the input image I , we obtained a binary threshold Th_O , such that intensities in I falling above Th_O constitute the foreground as determined by Otsu’s method. The *holes* in the foreground were filled using the fill operation based on morphological reconstruction [119], giving foreground mask M_f .

Similarly, the background mask M_b is obtained by taking the complement of M_f (Eq. 3.1). Using the background mask, we clean the background by altering the intensities of all background pixels to 255 (white). Thus, we remove the unwanted intensity variations (noise) in the background, which may interfere with subsequent processing. The rationale behind setting the background to white (and not black) will be explained later.

$$M_b = \overline{M_f} \tag{3.1}$$

Modelling ventricles as a collection of Maximally Stable Extremal Regions

We detect the ventricle as blobs inside the brain using the Maximally Stable Extremal Regions (MSER) algorithm [78]. MSER in an image is a connected region, which can be detected by an extremal property of the intensity function in the region and on its outer boundary. MSERs have properties that assist in their superior performance as a stable local detector: the set of MSERs is closed under continuous geometric transformations; MSERs are invariant to affine intensity changes and lastly, MSERs are detected at different scales.

Confidence based image patch classification for Ventricle detection

In order to prevent the MSER algorithm from including other parts of the brain as connected regions of the ventricles, we filter out these invalid parts by assigning a confidence value to each region based on its T1 intensity and its distance from the brain boundary. The applied criteria were motivated by the observation that ventricles are comprised of low-intensity (T1) image patches,

and are near the center of the brain. To detect the ventricles, we first calculate the following matrices:

D_p

The distance transform [104] of M_b (using the L1 norm or city-block distance measure) is normalized to the $[0, 1]$ range. We experimented using the L1 and L2 norms and got better results with L1. Distance transform of each point inside the brain gives its distance to the nearest point lying on the brain boundary (background mask). Thus, the points lying more towards the centre of the brain (where we expect to find the ventricles) tend to get higher values.

I_c

The normalized complement of I : Since I is a grey-scale image, this means $I_c = 255 - I$. I_c is then normalized to the $[0, 1]$ range. As the points inside the ventricles have low grey scale values (T1 images), the inverted image has high grey scale values. Recalling that in an earlier step, the background was marked white, allowing the ventricles to be easily extracted in the inverted image. After the ventricle detection step, the background is inverted to black with a zero value.

L_p

The $[0, 1]$ normalized *Hadamard* product of the matrices D_p and I_c , i.e., $L_p = D_p \bullet I_c$: This means that the distance transform of each point is multiplied with its (inverted) intensity. Following the definition of D_p and I_c , we can infer that the points lying inside the ventricle will have very high values for L_p , which can be used as a confidence measure for determining if a given point inside the brain belongs to the ventricles.

We apply the MSER algorithm on I , which detects blobs. MSER is one of the fastest region detectors because of its linear implementation [91]. It is also affine invariant, has good repeatability and performs well to classify patches with similar grey scale values on images containing homogeneous regions with distinctive boundaries. Let us assume that the MSER extracts R regions from

I , denoted as $r_1, r_2, r_3, \dots, r_R$. For each region r_i , we calculate the average confidence C_i , which is the average of L_p values for the pixels belonging to r_i . Thus, our problem is reduced to finding the maximal set of optimal regions $S_v = \{r_1, r_2, \dots, r_v\}$ forming the ventricles. As demonstrated later, sub-optimal ventricle detection results do not affect the accuracy of WMI detection.

Patch fitness evaluation for Ventricle detection

We model ventricle detection as an optimization problem and solve it using Genetic Algorithm (GA) [43]. When applying GA to our preterm brain images, for each region r_i (mentioned earlier), we have to make a *binary* choice of either to include it in a *candidate solution*, or leave it out. Thus, we represent a candidate solution as a *bit string*. We define a *fitness function* to obtain the candidate selection solution:

$$F_s = N_s * C_1 * C_2 * \dots * C_{N_s} \quad (3.2)$$

where N_s is the number of regions selected and C_j refers to the confidence of the selected region j . Thus, a GA returns the *optimal* choice of regions S_v , which are *most likely* to constitute the ventricles. As we will see later, this sub-optimal result of ventricle detection would not affect the overall effectiveness of the proposed method.

We also define a mask M_v for all pixels p_i belonging to the ventricles determined by the GA.

$$M_v = \{p_i \in r_j \forall r_j \in S_v\} \quad (3.3)$$

Detecting WM hyper-intensities

Next, we consider the white matter (WM) region around the ventricles. Our goal is to exclude the region where no WMI is present. Let D_v be the distance transform of M_v and we choose a contour that follows the relation:

$$|D_p - D_v| \leq 1 \quad (3.4)$$

This generates a contour, whose points are roughly halfway between the ventricles and the brain boundary. Thus, the mask M_w enclosed by the contour gives the candidate region, with the ventricles and the patches falsely detected as ventricles represented as holes. However, as we will show later, false ventricle detection does not affect the final WMI extraction. Also, ventricle patches included in M_w undetected by MSER (or GA) will not interfere in WMI detection, as their intensities are below the WM mean intensity.

Using the pixels described by M_w , we calculate their Median M_d and Median Absolute Deviation M_a . Any grey level g in the image I , satisfying Eq. 3.5, represents a potential WMI based on the Modified Z-score metric [56].

$$0.6745 \times \frac{g - M_d}{M_a} > 0 \quad (3.5)$$

The Modified Z-score uses the *median* instead of the *mean*, as the former is more robust to outliers. Also, as mentioned earlier, the parts of the ventricles undetected by MSER (or GA) lie in the range of values for g satisfying Eq. 3.6, and thus are not included as WMI.

$$0.6745 \times \frac{g - M_d}{M_a} < 0 \quad (3.6)$$

Filtering WM hyper-intensities

The next step is to define a mask M_c for the potential WMI candidates, such that pixels of image I whose grey level g satisfies Eq. 3.5 belong to M_c . We enumerate the 8-connected *objects* found in M_c using the method described in [50] (page 40-48). The general procedure is described below.

1. Run-length encode the input image.
2. Scan the runs, assign preliminary labels, and store label equivalences in a local equivalence table.
3. Resolve the equivalence classes.
4. Relabel the runs based on the resolved equivalence classes.

The above procedure returns a set of objects $S_w = \{O_1, O_2, \dots, O_w\}$. We then compute the sizes of the corresponding objects as $\{N_{O_1}, N_{O_2}, \dots, N_{O_w}\}$. We discard the 5% *largest* objects, e.g., segments of the skull boundary, as they are outliers. Our experimental observations verified that in T1 images, the skull shares similar intensity as the WM injuries; both lie above the normal range of WM intensities. Thus, the skull naturally forms the biggest objects in the set S_w (within the top 5%). In the next step, we perform a *binary* classification (big and small) of the remaining objects based on their sizes, using the K-means clustering algorithm. We initialize the starting means (or centroids) of the K-means algorithm with the sizes of the smallest and the largest objects (among the remaining 95%). Note that the WM injuries often fall in the small category, while the big category contains brain tissue boundaries. We impose these size constraint in our algorithm. Using the above process, we further enforce a distance constraint, based on our expert-annotated dataset, that WM injuries cannot lie close to the skull. More false positive WMI detections are eliminated subsequently in the fine detection process described next.

3.2.2 Fine Detection phase

The second phase of our algorithm combines the coarse detection results from adjacent DICOM slices. This is motivated by the understanding that white matter injury (WMI) position cannot change abruptly across slices. We consider slice number i in which a WMI is detected at a particular location (x, y) ; then, the same WMI spanning across slice $(i \pm n)$ will be roughly at the same location ($n = 1$ in the current implementation). We allow a *distance tolerance threshold* to account for a slight variation in position. By interpolating between slices, the algorithm can also identify noise and recover occluded WMI caused by weak intensity contrast. False positives can be reduced by considering adjacent slices. Since the coarse detection step is computationally efficient, the overall time performance is improved because fine detection is applied only on a small set of slices, which contain potential WMI candidates.

Let $l_i \in L_{-1}$, $l_j \in L_0$ and $l_k \in L_1$ be the centroids of the WMI regions detected in slice numbers $(n-1)$, n and $(n+1)$ respectively. Our fine detection

Table 3.1: MRI Acquisition Parameters

Slice Thickness (mm)	Field Strength (Tesla)	TR/TE (ms)	# of Signal Averaging	Flip angle (degrees)	Acquisition Matrix
1	1.5	36/9.2	1	30	192/96

constraint is defined in Eq. 3.7 and Eq. 3.8, where T is the set of positives predicted by the fine detection step and it consists of elements l_j which satisfies either Eq. 3.7 or Eq. 3.8. The notation $dist$ is the distance between two WMI centres and D_{th} denotes the distance tolerance threshold. We use a normalized Euclidean distance $(0, 1)$ with $D_{th} = 0.1$. Experiments showed that these parameter values produced the best results.

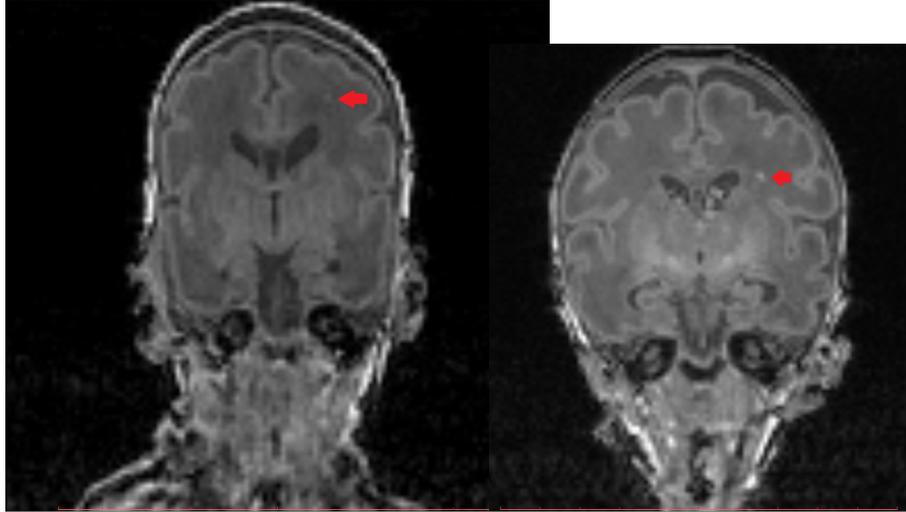
$$l_j \in T \Leftrightarrow \exists l_i \in L_{-1} \mid dist(l_j, l_i) \leq D_{th} \quad (3.7)$$

$$l_j \in T \Leftrightarrow \exists l_k \in L_1 \mid dist(l_j, l_k) \leq D_{th} \quad (3.8)$$

3.3 Results

We evaluate our method qualitatively and quantitatively using noisy and low-resolution (96×112) preterm neonate brain DICOM slices from three subjects provided by the SickKids Hospital in Toronto. Wherever applicable, WMIs were marked by expert radiologists on the slices as ground truth. Fig. 3.5 presents two representative Ground Truth WMIs marked on coronal cross-sections of two of those subjects. The MRI acquisition parameters are mentioned in Table 3.1.

It should be noted that WMI were not present in the slices from the first subject. However, those slices present more challenging scenarios for the ventricle detection step, due to the particular shape of the ventricles in them, as compared to slices from the other two subjects. Thus, in Fig. 3.7 we show results of ventricle detection on slices from the first subject and demonstrate the full WMI detection process on a representative slice from the third subject in Fig. 3.6. Also, throughout this work, we consider WMI detection on slices from the second and third subject only. For quantitative evaluation, we



(a) WMI marked on a 2D slice from DICOM stack of 2nd subject. (b) WMI marked on a 2D slice from DICOM stack of 3rd subject.

Figure 3.5: Visualization of representative Ground Truth WMIs on coronal cross-sections of subjects.

use the standard metrics: sensitivity and specificity, given in Eq. 3.9 and Eq. 3.10. We compare our results with a recent work on WMI detection [20]. It should, however, be noted that, the set of pixels in each slice used for computing sensitivity and specificity are those belonging to the brain and not the background.

Thus, true positives are the ground truth WMI correctly classified by a particular algorithm and similarly, true negatives are the correctly classified non-WMI pixels of the brain. False positives are the non-WMI pixels incorrectly classified as WMI, and similarly, false negatives refer to the missed detections.

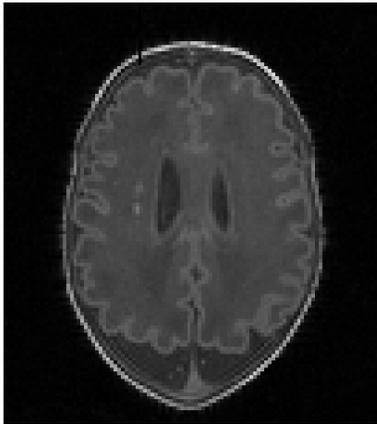
Also, in case of [20], significant variation in WMI detection performance in terms of sensitivity and specificity was noted by even slightly varying the upper threshold, \mathcal{T} for marking the potential WMI boundaries after calculating the transition matrix for each subject. Thus, for each subject, the value of \mathcal{T} has to be tuned separately to get the best performance for that subject, while using the method proposed in [20]. To cover these cases, we varied the value of \mathcal{T} for Subject-2 and Subject-3 to report the resulting variation in WMI detection performance for the method [20]. On the other hand, our

proposed method does not have such issues as it does not use any parameters in the main detection phase. The method [20] has several other parameters in addition to \mathcal{T} . However, the most significant variations in its output resulted particularly from even minute changes in \mathcal{T} , and hence we specifically highlight this parameter. Also, note that our work focuses on WMI detection in MRIs of preterm neonate brains, specifically on WM hyper-intensities in T1 images. Since there is insufficient relevant research result in the literature to compare, we chose a few WM lesion detection methods to show that applying those methods to our application does not produce better results.

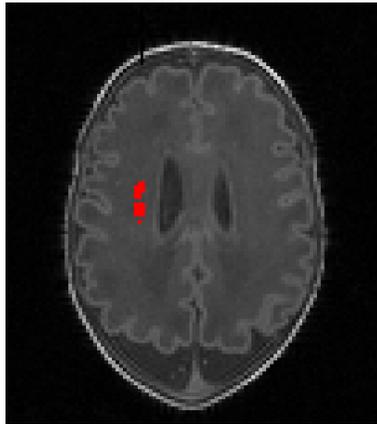
$$sensitivity = \frac{True\ Positives}{True\ Positives + False\ Negatives} \quad (3.9)$$

$$specificity = \frac{True\ Negatives}{True\ Negatives + False\ Positives} \quad (3.10)$$

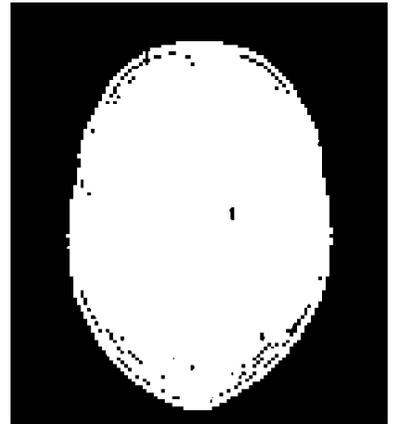
In Fig. 3.6 we illustrate the steps of our proposed method with a DICOM slice, and compare our WMI detection result with the expert-annotated ground truth. An input slice and ground truth are shown in Fig. 3.6a and Fig. 3.6b respectively. The output of Otsu’s method before and after the morphological hole filling operation is shown in Fig. 3.6c and Fig. 3.6d respectively. GA returns the set of blobs most likely to constitute the ventricles, as shown in Fig. 3.6e. However, it may be argued that for this particular input slice, the ventricles appear as relatively simple, regular shapes. Thus, we additionally show that the proposed ventricle detection approach works even for slices where the ventricles appear as more complex, irregular shapes, in Fig. 3.7. However, these slices do not contain WMI and are thus not used henceforth in showing WMI detection. The detected contour around the ventricles is shown in Fig. 3.6f. True ventricles and the patches falsely detected as ventricles are represented as holes in Fig. 3.6g. Brain hyperintensities based on the Modified Z-score metric are shown in Fig. 3.6h. The result of coarse detection, after imposing the size and distance constraints, is shown in Fig. 3.6i. Comparing with the ground truth shown in Fig. 3.6b, we see that there is both true positive and false positive WM injuries. Many of these false positives are eliminated by our fine detection step, as shown in Fig. 3.6j.



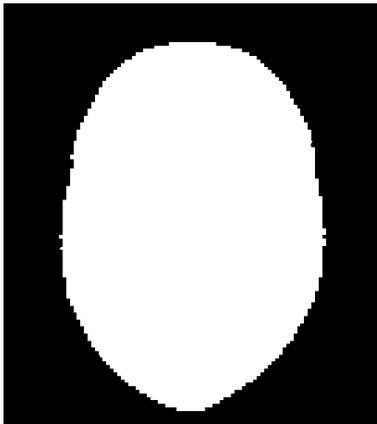
(a) Input Slice



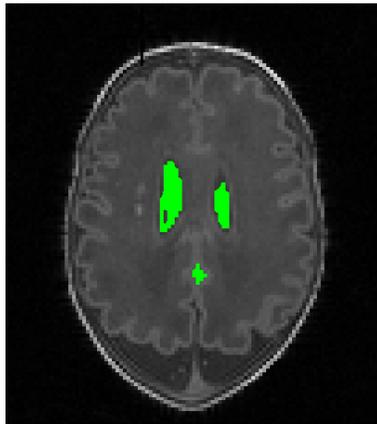
(b) Ground Truth with annotated WMI (Red)



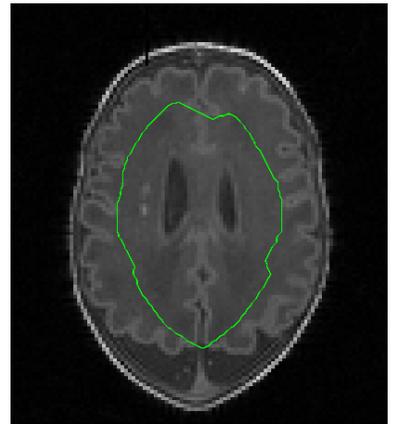
(c) Otsu's method



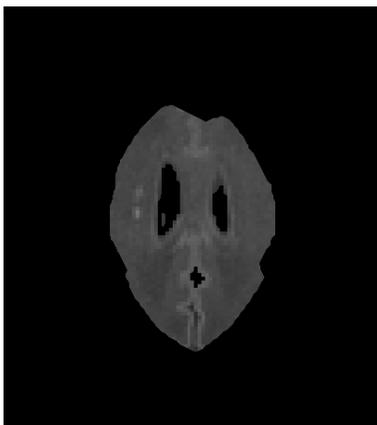
(d) Hole-filling



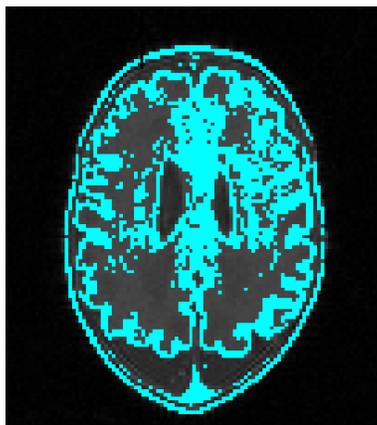
(e) Ventricle detection



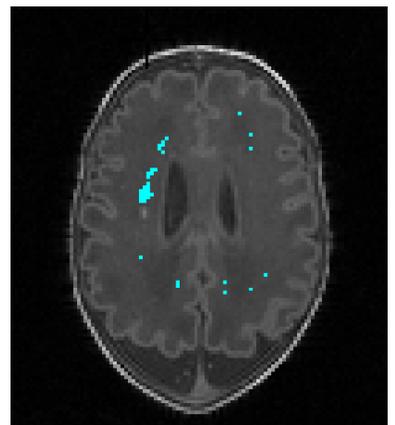
(f) Contour of Eq. 3.4



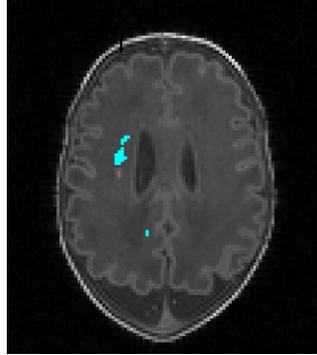
(g) Region mask M_w



(h) Hyperintensities detection



(i) Coarse detection includes false positives



(j) Fine detection removes many false positives

Figure 3.6: An illustration of the various steps in our proposed method.

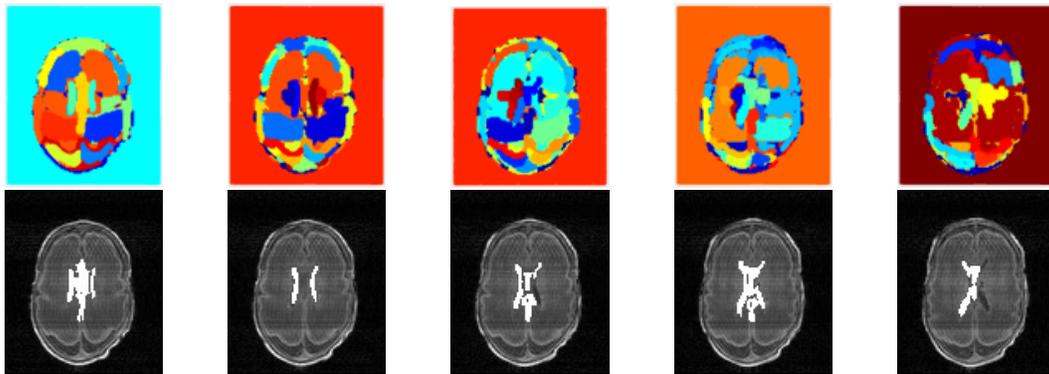


Figure 3.7: Result of blob detection (top row) and refinement (bottom row) to locate ventricles using the MSER algorithm on Slices 1 through 7. Each detected blob has been shown using a different colour for ease of visualization.

3.3.1 Quantitative Results

WMI Detection Accuracy

The sensitivity and specificity comparisons of proposed method with the method in [20] are presented in Table 3.2 (for each slice) and Table 3.3 (averaged over slices). Since we vary the value of \mathcal{T} [20], the scores reported for the method in [20] have been averaged over those obtained for individual values of \mathcal{T} for each slice. It can be seen that the average sensitivity and specificity are higher in case of the proposed method as compared to the one in [20], even though the proposed method does not segment the WM region.

Although the *average* sensitivity may appear as low for the proposed method, however, for some of the slices, the proposed method actually detects all the true positive WMIs, as shown later in the Qualitative Results section. The reason for the decrease in average sensitivity is due to missed detections in some other slices, and our analysis of the same is presented when we discuss the effects of our size and distance constraint features in the Discussion section. However, our proposed method is not dependent on any parameter like our earlier method [20], which is very sensitive to the value of \mathcal{T} which drives down its average sensitivity.

Since the sensitivity / specificity of proposed method are only slightly higher than that of [20], we perform the (non-parametric) Wilcoxon Rank-Sum test to investigate the statistical significance of these results. We perform the test using sensitivity, as it has a comparatively larger difference than specificity. We define ‘Sample 1’ as collection of sensitivity scores of proposed method and likewise, ‘Sample 2’ for the method [20]. We choose the rank-sum test as, although the two samples are independent, the sample size is small, which precludes the application of other (parametric) tests like Student’s t-test. For the same reason, we opt to perform a two-tailed test, and choose the significance level $\alpha = 0.01$ instead of 0.005. We put both samples together and organize it in ascending order, assign ranks to them, taking care of assigning the average rank to values with rank ties, as shown in the Table 3.4.

The sum of ranks for Sample 1 is $R_1 = 3 + 3 + 3 + 6 + 13 + 13 + 13 = 54$.

Table 3.2: Quantitative Comparison (for each slice) of Proposed Method using varying parameter values in [20]

Slice No	Sensitivity (Method [20])	Sensitivity (Proposed)	Specificity (Method [20])	Specificity (Proposed)
8	67.85	0	99.49	99.27
9	0	100	99.44	98.91
10	66.67	100	98.79	98.80
11	75	100	98.97	99.80
12	74.09	0	99.60	99.96
13	50.83	48.42	99.72	99.82
14	0	0	98.63	99.89

The sum of ranks for Sample 2 is $R_2 = 3 + 3 + 7 + 8 + 9 + 10 + 11 = 51$. Hence, the test statistic is $R = R_1 = 54$. The rank-sum test is performed as follows:

1. Null and Alternative Hypotheses: The following null (H_0) and alternative (H_a) hypotheses need to be tested:
 H_0 : Median (Difference) = 0
 H_a : Median (Difference) \neq 0
2. Rejection Region: The critical value for the significance level 0.01 and the two-tail test is $R_c = 12$, and the null hypothesis is rejected if $R \leq 12$.
3. Decision about the null hypothesis: Since in this case $R = 54 > 12$, there is not enough evidence to claim that the population median of differences is different than 0, at the 0.01 significance level.

We repeated the rank-sum test with $\alpha = 0.05$ for which $R_c = 21$, and the null hypothesis is rejected if $R \leq 21$. Since even in this case $R = 54 > 21$, there is not enough evidence to claim that the population median of differences is different than 0, at the 0.05 significance level. Thus, We see from the rank-sum tests and Table 3.3, that the chief advantage of the proposed method as compared to [20] lies in eliminating dependency on brain atlas and saving computational time by bypassing the time-consuming full WM segmentation.

In order to assess the effectiveness of our size and distance constraint features, we also compute the average sensitivity and specificity scores across all

Table 3.3: Accuracy & Execution Time comparison of Our Method with Method [20] by varying its parameter values

Average	Sensitivity (Method [20])	Sensitivity (Proposed)	Specificity (Method [20])	Specificity (Proposed)
	47.90	49.77	99.23	99.49
Time(ms)	Per Slice (Method [20])	Per Slice (Proposed)	Per Volume (Method [20])	Per Volume (Proposed)
	4150	500	796800	41500

Table 3.4: Wilcoxon Rank-Sum Test comparing Proposed Method’s sensitivity with Method [20]

Sample	Value	Rank	Rank (Adjusted for ties)
1	0	1	3
1	0	2	3
1	0	3	3
2	0	4	3
2	0	5	3
1	48.42	6	6
2	50.83	7	7
2	66.67	8	8
2	67.85	9	9
2	74.09	10	10
2	75	11	11
1	100	12	13
1	100	13	13
1	100	14	13

Table 3.5: Effectiveness of the Size and Distance Constraints on Proposed Method

Constraint	Sensitivity (average)	True Positives (average)	Specificity (average)	False Positives (average)
Both	49.77	36.71	99.49	501.86
None	100	110.71	56.70	32271
Size	49.77	36.71	97.82	1657.57
Distance	49.77	36.71	99.30	501.86

Table 3.6: Proposed Method’s Reduced performance with Minimum Lesion Size constraint [14]

Minimum Size	Sensitivity (average)	True Positives (average)	Specificity (average)	False Positives (average)
Min.Size N/A	49.77	36.71	99.49	501.86
Min.Size = 100	21.20	27.57	99.87	99.71
Min.Size = 150	6.92	26.28	99.94	49.71
Min.Size = 250	0	0	100	0

slices for the proposed method in each of the following scenarios:

1. with both constraints;
2. with no constraint;
3. with only size constraint; and,
4. with only distance constraint;

The above results are aggregated and presented in Table 3.5.

We conduct another experiment by restricting the minimum allowed lesion size, in terms of number of voxels, and the effect of this on WMI detection accuracy is presented in Table 3.6. The motivation behind this experiment is that sometimes random White Matter intensity variations may occur simply as a result of noise, and not the presence of actual White Matter Injury, as argued by authors in [14].

Execution Time Performance

For time performance, the average (serial) per-slice execution time of the proposed method is around 500 milliseconds on a Ubuntu 14.04 PC with 16 GB RAM and an Intel Core *i7-4790* 3.60 GHz CPU. Note that for the per-volume execution time to process a DICOM stack of 192 slices, the proposed method takes less than 500×192 milliseconds. This is because our method first performs coarse detection on all slices (which takes about 210 milliseconds per slice), benefiting the subsequent fine detection. The fine detection step focuses on a smaller set of WMI candidates and takes less than 6 milliseconds per slice. Thus, the total time taken to process the entire DICOM volume of 192 slices is 41460 milliseconds, or ≈ 41.5 seconds.

In comparison, the average per slice execution time of the segmentation-based preterm WMI detection method [20] is around 3.75 seconds (segmentation) + 0.4 seconds (detection) = 4.15 seconds (total). Its per volume execution time is 796.8 seconds = 13 minutes and 16.8 seconds. Also, note that the earlier method [20] performs only coarse detection without considering adjacent slices. This is because most of the execution time is taken up by the segmentation phase, whereas our proposed method bypasses segmentation, achieving better time performance without compromising accuracy.

Thus, from a comparison of the WMI detection accuracy and execution time of proposed method and our earlier work [20] shown in Table 3.3, we can conclude that our new method greatly reduces the execution time, but still manages to improve WMI detection accuracy compared to our old method.

Table 3.7 shows hemorrhage detection results on three representative DICOM slices of a preterm infant brain provided by SickKids Hospital in Toronto.

3.3.2 Qualitative Results

Figs. 3.8 through 3.10 present a side-by-side comparison of the output from [20] and the proposed method for three representative slices belonging to the 2nd and 3rd subjects. We observe that many false positives detected in [20] for certain values of \mathcal{T} are eliminated using our new method. Overall, Figs. 3.8

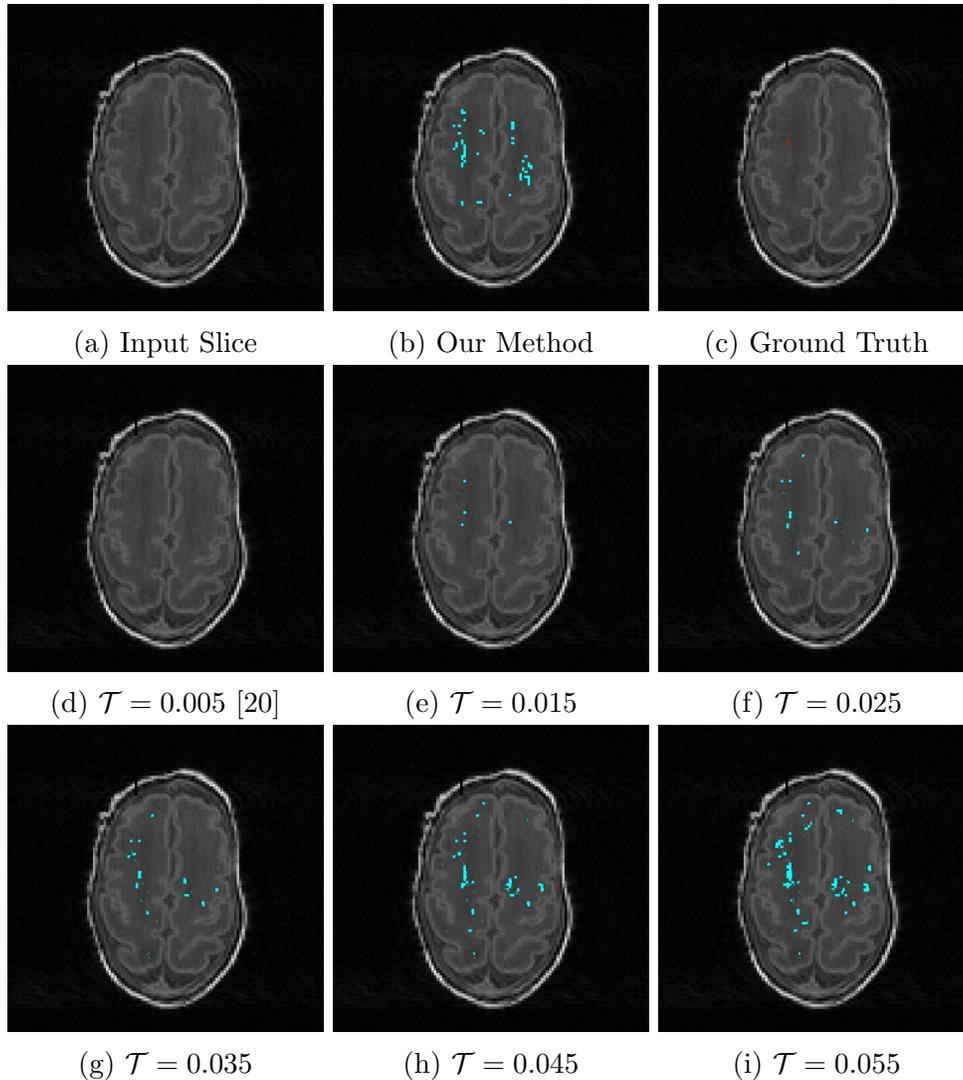


Figure 3.8: Comparison of Proposed Method and [20] using Slice 9.

through 3.10 show how our method performs better than the method in [20].

Fig. 3.11 shows hemorrhage detection results on three representative DICOM slices of a preterm infant brain provided by SickKids Hospital in Toronto. There are false positives mostly due to noise, but only one false negative (Slice 3, left hemorrhage) which also reflects in the quantitative results in Table 3.7.

A closer examination of the false positive detections reveals that they often have similar visual characteristics as the true positive detections. The best example of this is the false positive detection adjacent to the lower-right corner of the right ventricle in Slice 3 (Fig. 3.11).

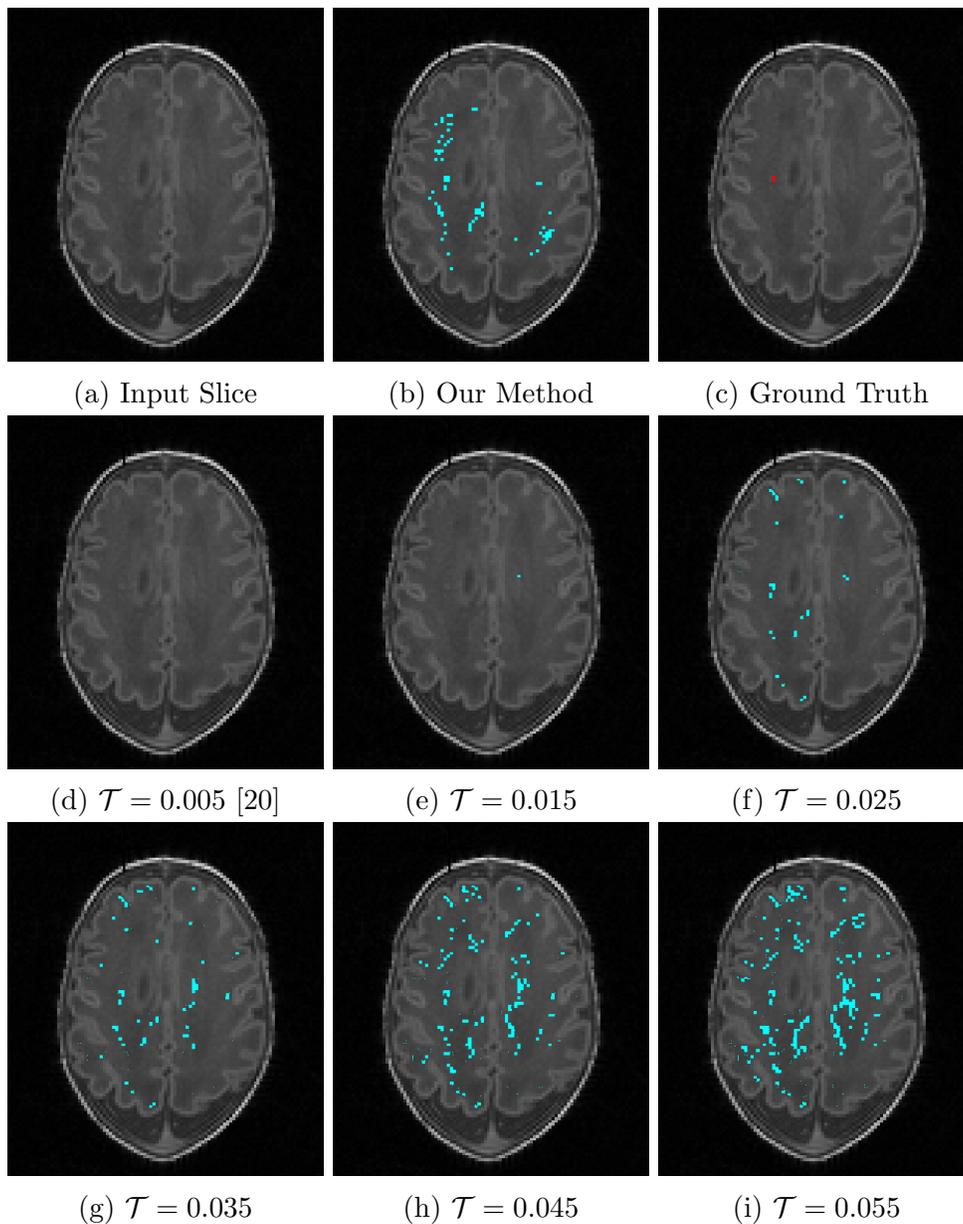


Figure 3.9: Comparison of Proposed Method and [20] using Slice 10.

Table 3.7: Periventricular hemorrhage detection quantitative results

Slice#	Sensitivity	Specificity
1	1.0000	0.9972
2	1.0000	0.9970
3	0.5056	0.9904

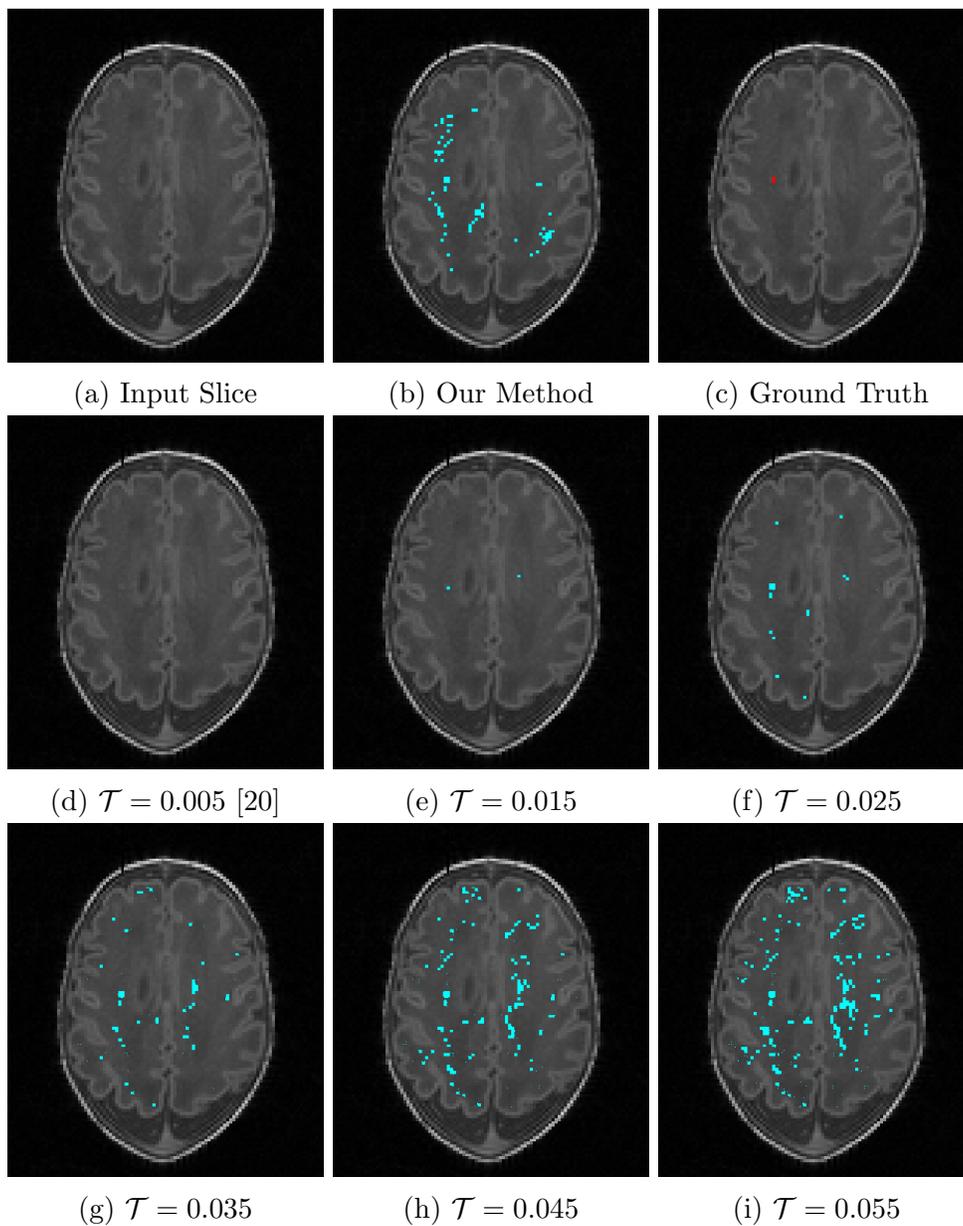
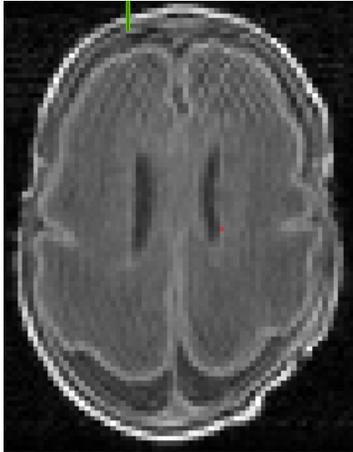
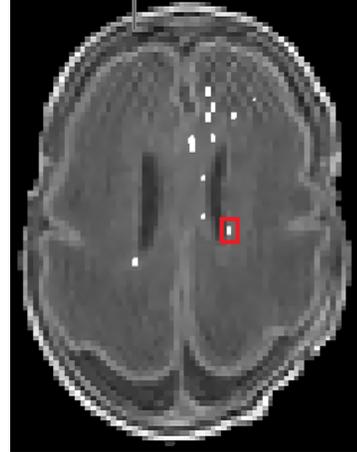


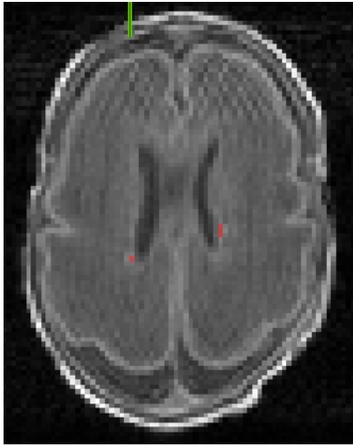
Figure 3.10: Comparison of Proposed Method and [20] using Slice 11.



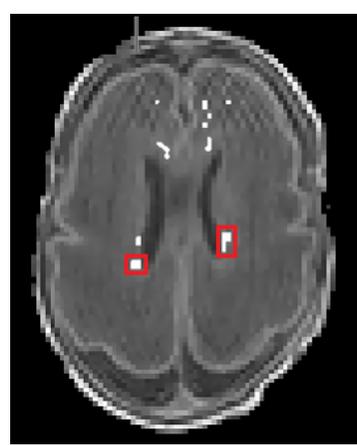
(a) Slice 1 with Ground Truth



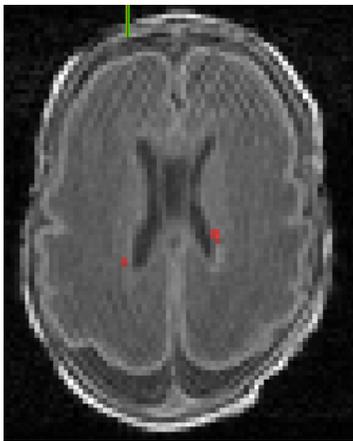
(d) Slice 1 Hemorrhage Detection



(b) Slice 2 with Ground Truth



(e) Slice 2 Hemorrhage Detection



(c) Slice 3 with Ground Truth



(f) Slice 3 Hemorrhage Detection

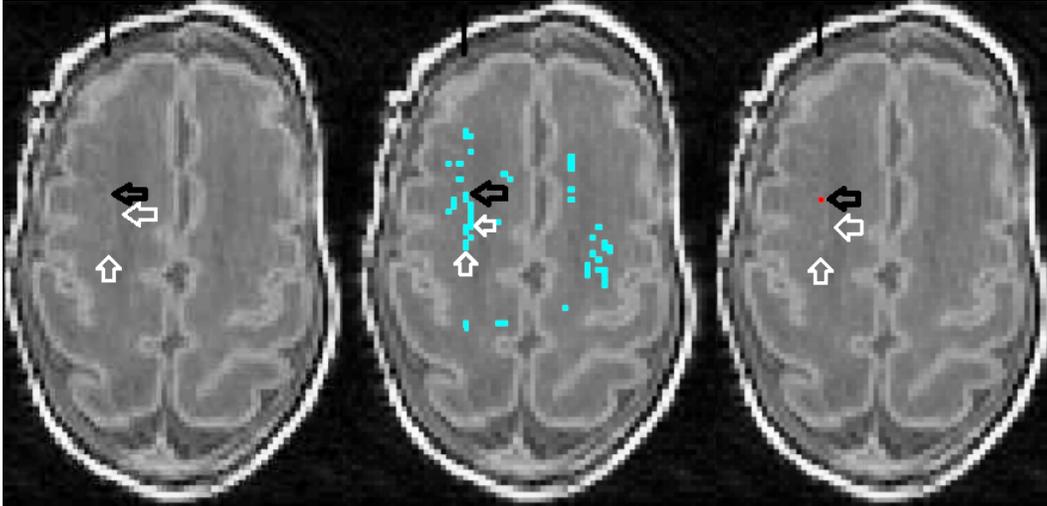
Figure 3.11: Periventricular hemorrhage detection using proposed method on DICOM slices obtained from SickKids Hospital, Toronto. Red squares mark True Positives.

3.4 Discussion

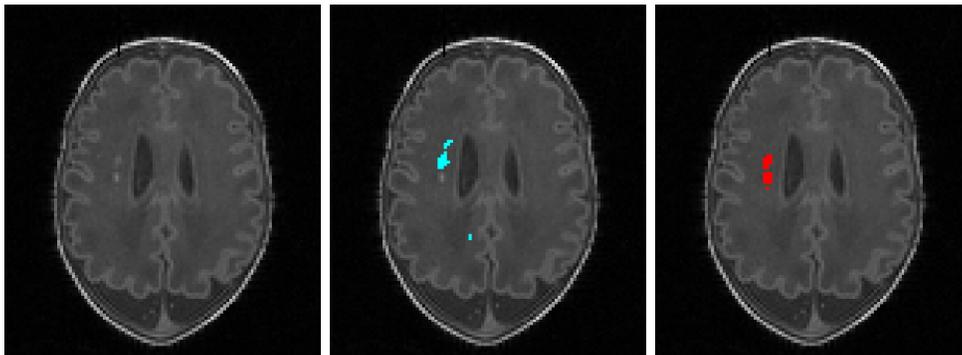
From the qualitative analysis, specifically Figs. 3.8 through 3.10, we observe the high dependence of the method in [20] on its parameter \mathcal{T} , whereas the method we propose in this work gives consistent results. There are instances where our proposed method detects the true positive WMI but deviates from the overall ground truth due to many false positive detections, e.g. Fig 3.8. However, a closer examination (Fig. 3.12a) reveals that the false positives often have similar visual characteristics as the true positive detections, and is thus difficult to identify.

Another strength of our method is that, even when the ventricles do not appear in a particular slice, the WMI detection performance is not compromised, as can be seen in Fig. 3.8. Most importantly, our method does not perform segmentation of the entire WM region like most other lesion detection methods in general. In fact, many such methods, assume that an already clearly segmented WM region will be supplied as input, so that the method only focuses on WMI detection disregarding potential segmentation errors. In any case, false positives which are obvious can be easily identified by the human expert when reviewing the computer assisted detection (first level filtering) results.

From the quantitative analysis, we can see that the enforcement of size and distance constraints clearly increases the specificity, and decreases the sensitivity. Specificity increases as false positives WMI detections are eliminated by the enforcement of the constraints. Sensitivity decreases because the constraints erroneously eliminate some true positive WMI detections. We believe that additional criteria based on prior clinical knowledge is needed in order to selectively retain the true positive WMI detections while discarding the false positives. This is especially true for situations in which the true positive WMIs are difficult to perceive visually. However, the size constraint feature of our proposed method does not impose any limitation on the *minimum* size of lesion to consider. In fact, in all of the three instances shown in Fig. 3.8, Fig. 3.9 and Fig. 3.10, the sensitivity is 100% (no misses), despite the ground truth lesions being very small. We additionally show the WMI detection performance of



(a) False positives often have similar visual characteristics as true positive WMI detections: Original Slice (left); WMI detected by proposed method (middle); Ground Truth (right). The white directional arrows point to the regions of false positive detections. The black directional arrows point to the location of true positives. The slices have been magnified and contrast of mid-tone region of the histogram of the entire figure has been enhanced for ease of visualization.



(b) Input Slice with Big Lesions (c) Proposed Method detects Big Lesions (d) Ground Truth showing Big Lesions

Figure 3.12: True and false positive WMI detection performance of proposed method relative to lesion size: top row shows small lesions, bottom row shows big lesions.

proposed method for large lesions in Figs. 3.12b through 3.12d. These figures had been used earlier to illustrate the various steps of our proposed method, in Fig. 3.6.

On the other hand, a specificity of 100 with no constraint implies that our WM hyper-intensity detection approach is robust enough to detect all true positive WMIs. Among the two constraints, the distance constraint alone performs as well as its combination with the size constraint. This is because, as we move from the periphery of the brain towards the vicinity of the ventricles, the size of the connected components of detected WM hyper-intensities keeps decreasing, as can be seen in Fig. 3.6h. The periphery has larger connected components, like skull and cortical ribbon, whereas those nearer to the ventricles are more likely to be true WMI. Thus, in this respect, the distance criteria subsumes the size criteria for WMI detection.

Furthermore, it should be borne in mind that even a small increase in the number of true positive WMI detection results in a large increase in the sensitivity score, and the opposite is true for false positives and specificity. The reason is that, in our preterm brain WMI detection based on the given ground truth, the total number of positives (WMIs) is very small compared to the total number of negatives (healthy brain tissue). Thus, the numerical values of sensitivity and specificity should be interpreted in light of the actual number of true positives and false positives while assessing the relative performance of the proposed method.

3.4.1 Slice Thickness

The inter-slice distance (thickness) in our test dataset is 1-mm and the magnetic strength of the MRI scans is 1.5 Tesla. It has been shown by Savlo et al. [112] that if we vary the slice thickness between 1-mm and 3-mm, the texture features used to detect lesions remain visible. While very thin slices would reduce the signal-to-noise ratio (SNR) resulting in unreliable texture patterns, very thick slices would compromise the texture detail.

3.4.2 Applicability of general assumptions regarding WM lesion characteristics

Multiple sclerosis lesion detection by thresholding FLAIR images, and subsequent refinement of the threshold mask in order to differentiate lesion regions from normal tissue, has been performed in [14]. The authors define a set of rules, like “Lesions are mostly surrounded by WM voxels” and “Lesions should not be present between the ventricles,” which are true in our test dataset as well. However, other rules like “Lesions (targets) should be of a minimum size” might not be true in our case. In their work, the authors set this minimum size as 10 voxels representing 30 mm^3 , which approximately represent a cube with 3 mm edges. They argue that due to inherent noise and intensity inhomogeneities in MR images, some voxels with random high intensities may persist even after the preprocessing stage. Thus, to remove these small outliers, they discard all lesion regions that do not have the defined minimum size. In terms of preterm WMI detection, the size of preterm neonate brains is very small, the MRIs themselves are of very low resolution and have a very low contrast-to-noise ratio. Thus, it is inaccurate to discard random high intensities assuming they are noise, because true WMIs can span just a small subset of voxels in each slice, as shown in our expert-annotated ground truth dataset. This observation is verified by our experimental results presented in Table 3.6, which shows that a minimum size cannot be imposed on preterm neonate WMI. It can be seen that by increasing the minimum size (in terms of number of voxels), there is an increasing number of missing detections, which lowers the sensitivity score. By the time we increase the minimum size to 250 voxels, all slices having correct targets detected in an unconstrained environment would have missed all correct detections. For this reason, the size criteria for eliminating false-positive in our proposed method does not impose a minimum acceptable WMI size. The specificity score increases as a result of increasing the minimum allowed lesion size, because fewer lesions are detected as such which also decreases the chance of false positive detections, as seen in Table 3.6. However, clearly, this cannot be regarded as an advantage at the

cost of severely increasing missed true detections.

3.5 Conclusion

We presented a robust and efficient method for the detection of white matter injury in preterm neonate brain MRI scans. We introduced a fast, automatic, unsupervised and atlas-free WMI detection approach, which avoids the WM segmentation step. We apply GA-based image patch classification to sample WM intensities and subsequently eliminate false positives using size-based and distance-based criteria. We also use adjacent slice validation. The proposed method is an effective technique for WMI and PVH detection in preterm neonates. Experimental results show that our method outperforms related work. However, there exist challenging MR scans, where the WMI cannot be identified purely relying on visual cues, leading to the failure of all methods. Our work also opens up new research possibilities to improve existing algorithms by incorporating demographic and other clinical information of patients to identify candidate targets. However, as of now, we do not have such data, and thus cannot work in that direction.

Chapter 4

InSAR signal recovery for WAM

4.1 Introduction

InSAR or Interferometric Synthetic Aperture Radar is an emerging, highly successful remote sensing technique for measuring several geophysical quantities like surface deformation [92]. It is based on generating an interferogram as the complex difference of two SAR acquisitions of the same scene from slightly different view angles, as shown in Fig. 4.1. The wrapped interferometric phase is then unwrapped to subsequently produce Digital Elevation Models (DEMs). However, several decorrelation factors create strong phase noise, affecting unwrapping and DEM accuracy [136]. Thus phase filtering is preferred, even when it results in some decrease in resolution and increase in spatial correlation [83] and we need filters adapted to enhance phase rather than amplitude [142]. Filtering the real and imaginary parts of the complex phase in its wrapped form [73], [99] can avoid blurring edges, whereas unwrapping before filtering increases computation and potentially decreases accuracy [92]. The boxcar filtering approach is still widely used today. This essentially involves computing moving average using a rectangular window. However, ground images acquired using InSAR constantly change (also called non-stationary) due to variations in topography and land displacements along the line of sight. This makes sample average methods like boxcar not the best solution for denoising InSAR images [114]. Also, strong smoothing effect of boxcar filtering results in spatial resolution loss, and noticeable errors in phase and coherence estimation near signal discontinuities. Thus, simple boxcar av-

eraging and non-adaptive filtering methods tend to distort the phase [83], [99]. Methods that adapt their parameters based on, e.g. local phase quality (coherence) yield better results, as coherence is related to phase noise deviation [73], [136]. Consequently, over the past few decades, various filtering methods have been developed to address the problem of non-stationary InSAR phase estimation. They are broadly classified as spatial methods, e.g., Lee [70], and frequency based methods, e.g., Goldstein [44]. Both filters, as well as [130] are adaptive to the local fringe direction. The Lee filter averages similar phase values in the locally adaptive estimation window, while the Goldstein filter discards all but the most dominant component of the local power spectrum. Both the original and the modified Goldstein filter of Baran et al. [7] preserves the signal in high coherence (low noise) areas, and thus, are adaptive to local noise as well. Enhancements of the Lee filter [17], [41], [133], [137] improve the adaptation to local fringe structure, whereas modifications to the Goldstein and Baran filters improve coherence estimation to avoid under-filtering the incoherent regions [61], [120]. Frequency based methods have been extended to the wavelet domain [75], including un-decimated wavelet transform [10] and wavelet packets [139], as filtering in wavelet domain preserves the spatial resolution, but struggles to filter dense fringes, whereas spatial methods in general sacrificed spatial resolution [138]. The additive noise model of interferometric phase [70] inspired early filtering methods which assumed a stationary and consistent phase over the filtering window, but real-world challenges of strong topographic change and restrictions imposed on window size motivated more recent non-linear models [54] and per-pixel filtering [128]. Local modeling based on polynomial approximation [12] and sparse coding [53] have been explored. Markov Random Field (MRF) based methods [31], [37] have also been attempted, though prior distribution modeling (required for MRF) is a challenging problem in itself. Recent attempts at Non-local filtering of various types of SAR images, which perform pixel estimation by matching patches over the whole image instead of just the pixel neighborhood, produced state-of-the-art results on amplitude, interferometric phase, polarimetric and multitemporal stack images [28], [117].

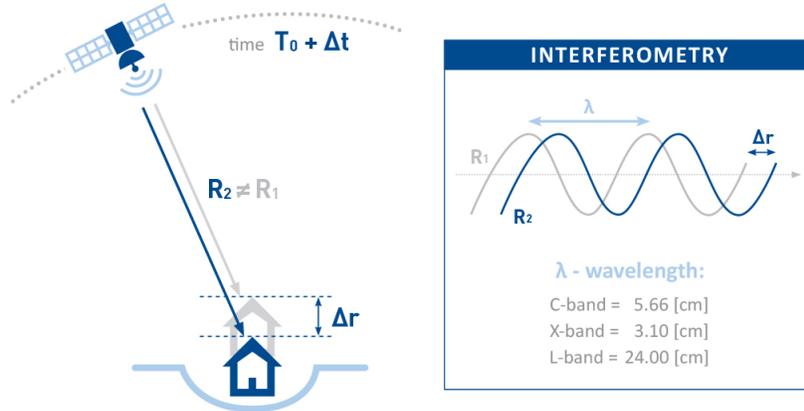


Figure 4.1: InSAR image capture by two passes of the satellite.

Recent advances in parallel computing architectures have motivated parallelism in the InSAR processing pipeline [96], which is critical to our proposed phase filtering method for InSAR-based Wide Area Monitoring (WAM) across geographical regions on petabytes of data. While optical images used in commercial or public applications are typically few megapixels in dimension, InSAR image dimensions can easily go upto the order of gigapixels (10^3 times larger), thereby making algorithmic scalability a prime concern for InSAR processing. Thus, we use a Convolutional Neural Network (CNN) architecture which seamlessly integrates with modern parallel architectures built on Graphics Processing Units (GPUs) and rival human performance on pattern recognition tasks. CNNs' use in InSAR phase processing in particular has been limited to volcano deformation monitoring [1] via transfer learning using a popular pre-trained optical image classification CNN [66], but not direct training on InSAR data. While Neural Network based despeckling of SAR images [129], [134], [135], [140] and geo-localization accuracy improvement of optical satellite images [82] have been attempted, the use of Convolutional Neural Network (CNN) based learning approaches to InSAR images has not been explored. Note that, methods like PtSel [102] which operate on a time series of numerous interferograms cannot start monitoring till all images of the stack have been acquired by numerous passes of the satellite. However, this could delay the start of monitoring by months, as each pass of the satellite happens only after about a week or more, depending on the particular satel-

lite cluster involved. This is impractical for many real-world applications like mines where miners are at risk from ground subsidence / landslides which can happen much earlier than the time required to acquire all the images required to start applying algorithms like PtSel. Thus, the current work focuses on, and compares against only those existing methods which work using a single interferogram.

In this work, we propose interferogram denoising and coherence prediction using CNN architectures. We show that even in the absence of any clean training images, unsupervised learning (using only noisy InSAR images) via CNN architectures and generative modelling can outperform traditional methods.

4.2 Proposed Methods

We first describe the initial method we developed (CNN-InSAR) which uses two separate CNN architectures for filtering the phase and predicting the coherence, given the noisy input image. We subsequently describe our improvement on this method by introducing generative modeling into the CNN to filter phase and predict coherence using a single architecture (GenInSAR).

4.2.1 CNN-InSAR: CNN-based phase filtering and coherence estimation

Dimensionality Reduction for InSAR Denoising via CNN

While most supervised neural network based learning architectures require pairs of noisy and clean training images to learn how to transform an unseen noisy image to clean image, autoencoders do not have this limitation. Autoencoders can perform unsupervised learning as they can learn from the noisy data itself to reconstruct the input noisy data. Since it performs dimensionality reduction within the network, it learns to reconstruct from a latent representation of its input, which is in a lower dimensional space (the ‘encoded’ layer). Consequently, it learns only features that are essential to reconstruct its input, getting rid of the noise in the process. However, if there is excessive dimensionality reduction, the image detail also gets lost along with the noise.

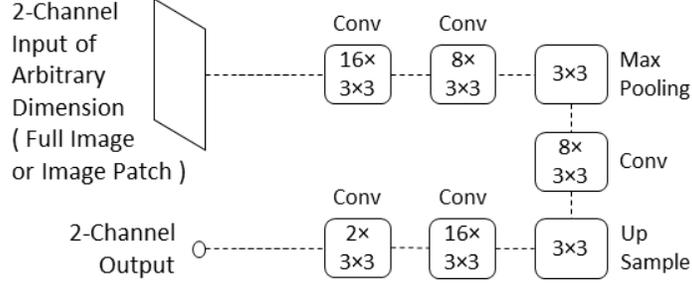


Figure 4.2: Autoencoder CNN for denoising InSAR images.

Thus, there is a trade-off.

The structure of our autoencoder is shown in Fig. 4.2. It should be noted that we do not constrain the size of the images used to train the network or run inference using the trained network (we can use whole image or patches). Also, the number of channels in the input is two, representing the real and imaginary part of the complex interferogram image. Thus, we exploit information from both channels. Nevertheless, we first process both the channels separately to saturate the outlier amplitudes, as some pixels in the input interferogram might have extremely high amplitudes. This may confuse the CNN during training / inference. Let the pixels in the interferogram be represented by $Z = [z_1, z_2, \dots, z_N]$ with amplitudes $A = [a_1, a_2, \dots, a_N]$. We can compute the amplitude of each pixel and threshold them to be $A' = \text{saturate}(A, M)$, where M is the mask denoting outlier amplitude values. The outliers are computed as in [25]. After saturation and normalization, real and imaginary channel values lie between -1 and +1. We add 1 to the two channels to use the Rectified Linear Unit (ReLU) activation for introducing nonlinearity in the CNN to learn complex features.

In Fig. 4.2, each CNN layer is represented by a box. Output feature map count is indicated by the integer at the top (2, 8, 16) whereas filter dimension is indicated at the bottom (3x3). Each 2D convolutional layer learns a number of filters. The maxpooling layer subsamples its input feature maps, while the upsampling layer brings them back to their original size. The output of the last convolutional layer has two feature maps representing the real and imaginary channel of the denoised interferogram. Mean squared error between the output

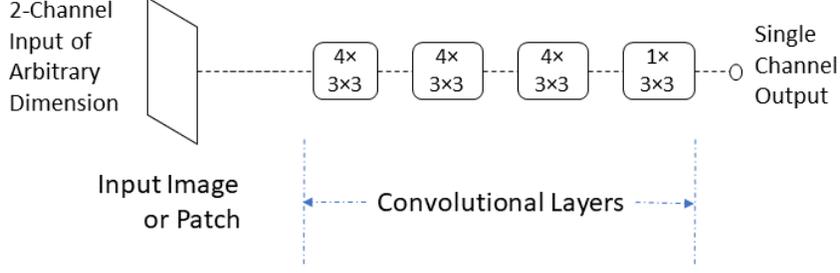


Figure 4.3: CNN for InSAR coherence estimation.

channels and their corresponding input channels is reduced using the popular Adam optimizer to train the network, by updating its filter weights and biases, using gradient backpropagation. The network is trained using 60×60 patches extracted from noisy, real-world InSAR images. Maxpool downsampling factor is 3. Xavier style weight initialization [42] is used for both networks.

CNN-based Coherence Estimation to reduce Artefacts

We propose the coherence estimation CNN shown in Fig. 4.3. This architecture also has two channel noisy interferogram input, and the preprocessing for the channels is same as the one described earlier. However, we do not have any downsampling or upsampling, and the output is a single channel, as we require prediction of just the amplitude of the coherence. While the design is similar to the earlier network, the activation for final convolution layer is sigmoid instead of ReLU. This is because, pixel coherence is a value between 0 and 1 (sigmoid output). We use kernel regularization in the 2^{nd} last convolution layer to limit standard deviation between learned filter weights for that layer. This helps make the coherence estimates sharper and less grainy. For training this network, we first use 11×11 patches to compute the raw coherence (Eq. 4.1) between the training noisy images and their filtered version output by the fully trained filtering network described earlier.

$$\hat{\gamma} = \frac{\sum_{n,m} u_1(n, m) \cdot u_2^*(n, m) \cdot e^{-j\phi(n,m)}}{\sqrt{\sum_{n,m} |u_1(n, m)|^2 \sum_{n,m} |u_2(n, m)|^2}} \quad (4.1)$$

where pixel n of interferogram u_1 and pixel m of interferogram u_2 have the topographic phase correction factor Φ , and the asterisk on top of u_2 denotes

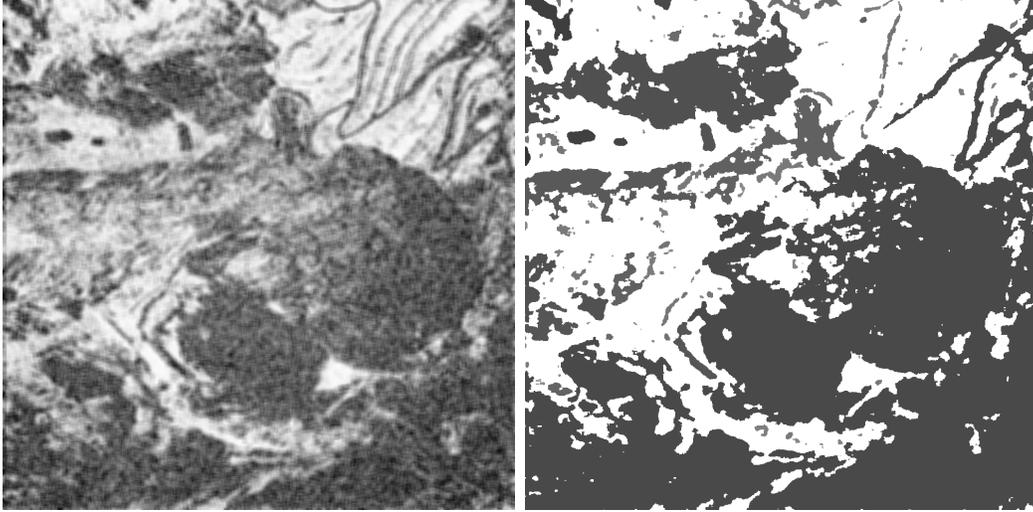


Figure 4.4: Preprocess raw 11×11 coherence to remove artefacts.

complex conjugate. A relatively larger window size is used to reduce bias in raw coherence computation, but this also makes the coherence speckled. So, instead of training directly on raw coherence, we preprocess the raw coherence to make it sharper. We first segment the raw coherence using Chan-Vese segmentation [16], which is useful for segmenting objects with ill-defined boundaries like our case. The segmentation isolates incoherent areas from coherent ones. We set coherence of all coherent areas to 1 (full coherence). For each incoherent area, we set its coherence to $\{ \text{Mean minus Standard Deviation} \}$ of coherence of all pixels comprising that area. We then extract corresponding 64×64 patches from training noisy image and its preprocessed coherence to train the coherence prediction CNN. The result of pre-processing is shown in Fig. 4.4 for a sample training image, where dark and light pixels show values closer to 0 (low coherence) and 1 (high coherence) respectively.

4.2.2 GenInSAR: CNN-based generative modeling for phase filtering and coherence estimation

Building on CNN-InSAR, we next propose GenInSAR, a novel InSAR phase filter inspired by Mixture Density Networks (MDN) [13]. A CNN’s convolutional layers operating on a phase patch predict the parameters of a bi-variate Gaussian distribution (real and imaginary channel) of the center pixel. The

predicted vector $\vec{\mu}$ represents the filtered pixel. Its coherence is a function of the predicted $\vec{\sigma}$. This approach improves phase filtering, and being a generative model, sampling from this distribution generates new interferograms which are slight variations on the filter output, and can be utilized to improve the InSAR pipeline, as discussed later on.

In contrast to CNN-InSAR, GenInSAR performs joint phase filtering and coherence estimation using only a single neural network. Our approach predicts the distribution of the center pixel given only its neighborhood (patch) and is thus embarrassingly parallel [52]. In contrast, non-local filters [29], [30] require computing patch similarity and suffer from terrace-like DEM artefacts, over-smoothing and rare patch effect [3]. For similar computational concerns, we do not adopt strategies that are iterative [83], [142] (could also result in loss of detailed features [83]), multi-stage [3], [142] and require optimization during inference, e.g. via sparse coding [92].

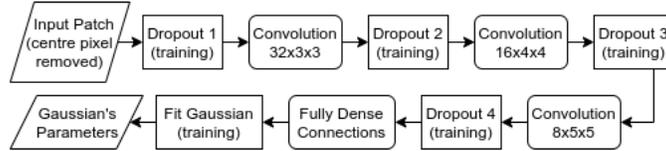


Figure 4.5: Architecture of GenInSAR.

The architecture of GenInSAR is shown in Fig. 4.5. The input to the architecture is a 11×11 phase patch centered around the pixel to be filtered. That pixel is removed from the input patch to avoid learning the identity mapping. We can understand this more clearly in terms of the training (fitting) and testing (prediction) steps. While training, patches extracted from a fixed set of phase images (training set) are input to the model. We set 20% dropout rate [124] for the first dropout layer and 50% for the remaining ones during training to prevent over-fitting. Dropout randomly sets a fraction of activations of its preceding layer to zero. Intuitively, this forces the network to learn simpler mappings for each training example, thus preventing over-fitting. Convolutional layers [68] of increasing filter size (3×3 , 4×4 , 5×5) and decreasing filter counts (32, 16, 8), each followed by an Exponential Linear Unit activation [23] (not shown) with $\alpha = 1.0$ promote fast convergence and non-linear

mappings. It also allows negative outputs ≥ -1.0 (lower limit of $\cos \theta$ and $\sin \theta$). Specifically, we use depth-wise separable 2D convolutions [111] with one filter per input channel (depth) $\hat{G}_{k,l,m} = \sum_{i,j} \hat{K}_{i,j,m} \cdot F_{k+i-1,l+j-1,m}$ for fast computation and convergence, where the m^{th} kernel \hat{K}_m is applied to m^{th} $k \times l$ input feature map F_m to obtain m^{th} output feature map \hat{G}_m , followed by a 1×1 convolution to combine the outputs. Finally, following MDN working principle, dense connections (weighted sums of all filter outputs) to the distribution fitting module outputs those Gaussian parameter values $(\vec{\mu}, \vec{\sigma})$ for the real and imaginary channel that make the input patch’s central pixel (training target) most likely.

Thus, our training is completely unsupervised, as we learn from the input data itself, without requiring its “clean” version as the training target. The central pixel t^q (surrounded by its neighborhood pixels, x^q) is treated as a sample drawn from the reference Gaussian distribution chosen to best encompass all n training set examples $\{x^q, t^q\}$, by minimizing the loss $E = -\ln \prod_{q=1}^n p(t^q|x^q) p(x^q)$ via gradient descent back-propagation using Adam optimizer [65]. The network is thus trained to parameterize a Gaussian density that best encompasses observed (noisy) data, by minimizing E . During testing, the central pixel is still removed from the input patch, but dropout and distribution fitting are not required: We predict the central pixel without any time-consuming optimization.

GenInSAR does not train to predict γ . It is directly computed from the predicted $\vec{\sigma}$. A nice property of γ is that it seems to be a better measure of filtering quality and filter output reliability, which partially depends on the spatial noise pattern (neighborhood), not just the noise underlying the center pixel. Considering two SAR acquisitions (u_1, u_2) with resulting interferometric unwrapped phase θ having probability density $p(\theta)$, variance σ_θ^2 , and real and imaginary components (R, I) with predicted variances (σ_R^2, σ_I^2) , we derive γ as the normalized index of mutual linear predictability of random variables u_1 and u_2 , thus quantifying noise in interferometric acquisitions:

$$\gamma = \frac{E\{u_1 u_2^*\}}{\sqrt{E\{|u_1|^2\}} \sqrt{E\{|u_2|^2\}}} = E\{u_1 u_2^*\}$$

denominator reduces to 1.

$$\begin{aligned}
\gamma &= \int_{-\infty}^{\infty} p(\theta) e^{i\theta} d\theta = \int_{-\infty}^{\infty} \frac{e^{-\frac{\theta^2}{2\sigma_\theta^2}}}{\sigma_\theta \sqrt{2\pi}} [\cos \theta + i \sin \theta] d\theta = e^{-\frac{\sigma_\theta^2}{2}} \\
\text{Again, } \sigma_R^2 + \sigma_I^2 &= E\{R^2\} - E\{R\}^2 + E\{I^2\} - E\{I\}^2 \\
&= \int_{-\infty}^{\infty} p(\theta) \cos^2 \theta d\theta + \int_{-\infty}^{\infty} p(\theta) \sin^2 \theta d\theta \\
&\quad - \left(\int_{-\infty}^{\infty} p(\theta) \cos \theta d\theta \right)^2 - \left(\int_{-\infty}^{\infty} p(\theta) \sin \theta d\theta \right)^2 \\
&= e^{-\sigma_\theta^2} \left(\cosh(\sigma_\theta^2) + \sinh(\sigma_\theta^2) \right) - e^{-\sigma_\theta^2} - 0 \\
&= 1 - e^{-\sigma_\theta^2} = 1 - \gamma^2 \implies \gamma = \sqrt{1 - (\sigma_R^2 + \sigma_I^2)} \tag{4.2}
\end{aligned}$$

We assume a Gaussian distribution for the unwrapped phase noise, to approximate the InSAR multiplicative speckle noise distribution. Generally, a Gaussian Mixture Model can approximate any distribution arbitrarily well by adding more terms, but we lower the free parameter count to achieve a lower bias. Moreover, since the number of effective samples is low, highly accurate characterization of the true underlying distribution is not necessary, since the mean for the two distributions will be equivalent and the only difference will be in the variance. Thus, the variance might be slightly underestimated, but this is a common problem with most coherence estimators.

GenInSAR’s chief advantage is that we can train it directly on real InSAR data. Thus, we have a filter that learns the real distribution of the data rather than relying on assumptions. This is very useful, as nowadays, countless images are being acquired daily by an ever-increasing number of SAR satellites around the world, creating a huge archive of real training data. Although most traditional filters rely on the spatial context to estimate the phase and coherence, GenInSAR’s coherence qualitatively appears to be more of an indication of the confidence of its estimated phase. This is a useful feature, because in real-world applications, GenInSAR might encounter a feature that it was not trained on, so there might be some error in its phase estimation. In that case, it would predict a low coherence, and that pixel will be down weighted or removed in subsequent stages of InSAR processing. Additionally, in areas

that are noisy but very smooth, the predicted coherence might be biased up slightly because GenInSAR is more confident of its estimated phase, as it can make better use of the contextual information.

4.3 Results and Discussion

We implemented both the networks comprising CNN-InSAR using Keras with Tensorflow back-end. We trained the filtering CNN and coherence estimation CNN by extracting 500 60×60 sized and 64×64 sized patches patches respectively from each of 135 1000×1000 training interferograms. The CNNs converged after 50 and 100 epochs respectively. During testing CNN-InSAR, we use the whole input noisy interferogram, instead of just patches, whereas for GenInSAR, the input is always the patch centered around a pixel.

We also implemented GenInSAR in Keras [21] with Tensorflow-GPU back-end, and compared its performance with CNN-InSAR and four existing methods: Boxcar, Goldstein, NLIInSAR, NLSAR, all implemented / executed in OpenCL 1.2 on 8 GB NVIDIA 1070 GPU. We present the qualitative and quantitative results of those experiments for real and simulated images respectively. For qualitative analysis, we observe whether a filter tends to over-filter less noisy regions, especially those having interferometric fringes, which might get distorted as a result (undesirable). For the coherence, we look at whether the filter can clearly distinguish between coherent and incoherent areas of the image. The metrics used for quantitative analysis are Root-Mean-Square-Error (RMSE) of the InSAR phase and coherence, Residue Reduction Percentage (RRP) [83], [136], [138], and Phase Cosine Error, $\epsilon_{\cos}^{\Delta\theta}$ (Eq. 4.3) where g_i and \bar{f}_i denote i^{th} ground truth and complex conjugate of filtered pixels of an n pixel interferogram. Residues are phase inconsistencies emphasized by computing curl of phase differences over the range of a reduced closed integral loop of four spatially adjacent pixels [27], [138], which are non-zero if residues are present. Most residues are caused by noise. However, few arise from signal structure, like steep change in topography or heavy deformations, and those residues should be preserved during filtering. Filtering should remove all other

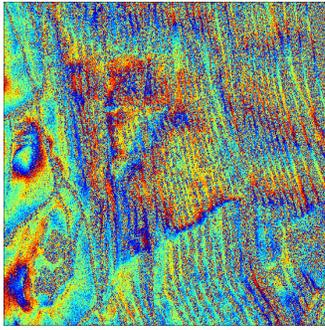
residues to facilitate phase unwrapping. Those that cannot be removed should have low values in the filter’s output coherence map; this prevents error propagation during phase difference integration by the unwrapper. Hence, filtering aims to reduce residues (high RRP) but preserve details (low Phase RMSE, $\epsilon_{\cos}^{\Delta\theta}$). These criteria drive our evaluations, as described next.

4.3.1 Experiments using satellite InSAR images

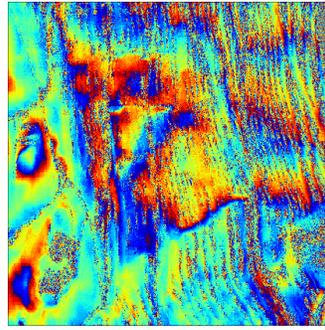
We trained GenInSAR for 100 epochs on 5 million 11×11 patches extracted from numerous interferograms of an airport and a mining site, having resolutions 5060×4040 and 1000×1000 respectively, in batches of 64 patches each. We tested the model on 1000×1000 interferograms of a different mining site, by extracting one 11×11 patch for each pixel (at center). The interferogram edges were replicated for obtaining patches corresponding to the edge pixels. Fig. 4.6 shows outputs for a test interferogram using proposed and existing methods. GenInSAR is a generative model: its filtered output per pixel (i, j) corresponds to the mean $\vec{\mu}_{i,j}$ (and $\vec{\sigma}_{i,j}$) of the distribution predicted for that pixel. To show this, we randomly sample five times from the normal distribution $N(\vec{\mu}_{i,j}, \vec{\sigma}_{i,j} * \alpha)$ setting $\alpha = 0.1$, and generate pixel $\vec{P}_{i,j}$ for five images in Fig. 4.7, which are slightly different outputs for the same input test interferogram of Fig. 4.6. Higher values of α generate more variations in the outputs, but also tend to make them noisy. This can be used in InSAR machine learning for data augmentation [116], or to test InSAR processing chains by running the same chain all the way through but with slightly different interferograms to measure variance of the outputs of the complete processing chain, and might be good for error analysis as well.

4.3.2 Experiments using simulated InSAR images

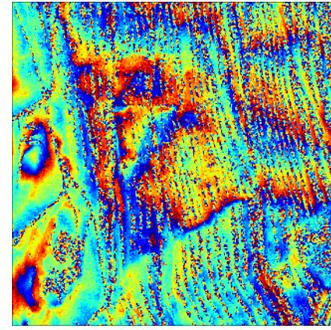
Satellite interferograms are always noisy, and thus cannot be used for quantitative analysis. So we used an InSAR simulator that can simulate ground truth interferograms with Gaussian bubbles, roads and buildings. We followed a similar training strategy as satellite InSAR images for training GenInSAR with simulated InSAR images, by adding Gaussian noise to simulated ground



(a) Input Phase



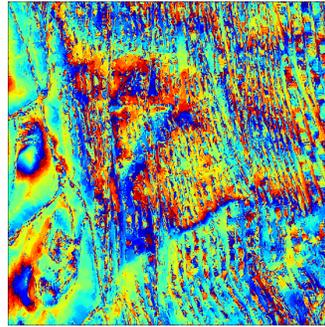
(b) Goldstein Phase



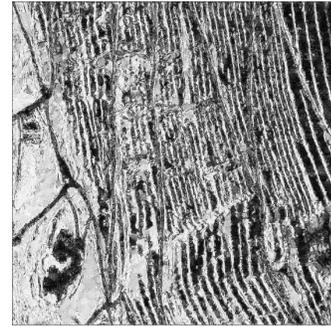
(c) Boxcar Phase



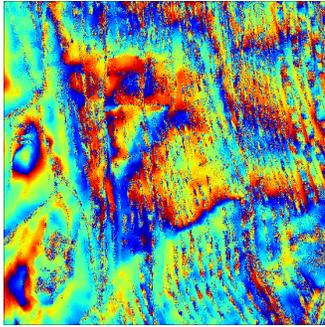
(d) Boxcar Coherence



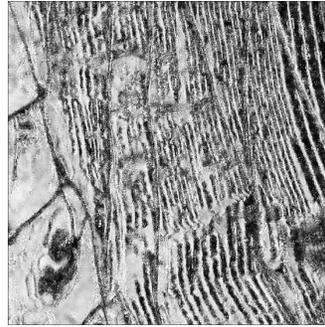
(e) NLIInSAR Phase



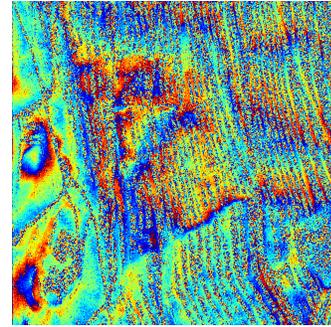
(f) NLIInSAR Coherence



(g) NLSAR Phase



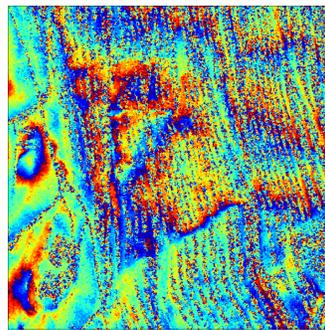
(h) NLSAR Coherence



(i) CNN-InSAR Phase



(j) CNN-InSAR Coherence



(k) GenInSAR Phase



(l) GenInSAR Coherence

Figure 4.6: Filtered phase and coherence outputs for satellite InSAR images processed by GenInSAR, CNN-InSAR, and four existing methods. Visualizations for phase are coloured between $-\pi$ (blue) to $+\pi$ (red), and coherence between 0 (black: low) to 1 (white: high) respectively.

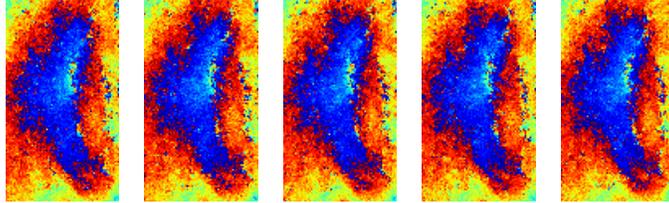


Figure 4.7: Cropped interferometric phase images generated by GenInSAR for the same noisy input. Visualizations coloured from $-\pi$ (blue) to $+\pi$ (red).

truth images, and inputting patches extracted from those noisy versions. For CNN-InSAR, we generated two sets of results: one using the model as-is and another by retraining it with simulated noisy images as mentioned above. For evaluating the proposed and five existing methods mentioned earlier including CNN-InSAR (as-is and retrained), we used a set of 60 1000×1000 noisy simulated images. Fig. 4.8 shows the performance of all methods for a sample simulated test image. The corresponding clean (ground truth) versions of those images facilitated quantitative evaluation in terms of the three metrics mentioned earlier. Table 4.1 shows overall superior quantitative performance of GenInSAR against others and almost linear speedup with increasing number of GPUs, as it filters each pixel independently, based on its neighborhood.

The results show that coherence produced via both the CNN-based methods clearly marks where the signal could be reliably recovered, and creates lesser phase artefacts in incoherent regions compared to the non-local methods. CNN-InSAR creates far less spatial variance in coherence estimates in incoherent areas by using a large window size (11×11) to compute raw coherence, preprocessing it before training, and regularizing kernel weights of the 2^{nd} last convolution layer. As shown earlier, the raw coherence is generated by multiplying the input interferogram with complex conjugate of its filtered version to prevent the signals from biasing down the coherence estimates. However, where denoising fails to remove noise, this cancels out noise in areas

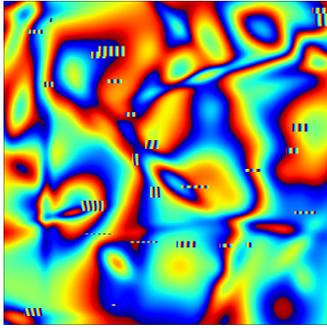
where noise is present in both input and filtered version, and thus drives up coherence of affected regions. GenInSAR circumvents this problem with its unique approach to coherence estimation based on the estimated parameters of the Gaussian generating the filtered (central) pixel. Goldstein and NLSAR tend to warp fringes and Goldstein under-filters near incoherent areas.

GenInSAR almost totally reduces residues and produces far less over-smoothing/artefacts around branch cuts compared to Boxcar because its greatest strength is unsupervised learning of true spatial smoothing only from noisy training data. It could potentially detect real residues better if trained more on such types of features, and an efficient implementation like those of four existing methods [4] could reduce its run time. In general, NLInSAR handles residues well and avoids artefacts by selecting neighbors with similar phase, but produces streaking correlated with amplitude bands. NLSAR (conservatively) interpolates well only over heavy noise. A final future work for GenInSAR is improving the coherence function to more elegantly handle practical scenarios where the input data does not lie on the unit circle as currently, $(\sigma_R^2 + \sigma_I^2)$ is clipped to $[0, 1]$, although most values lie in that range.

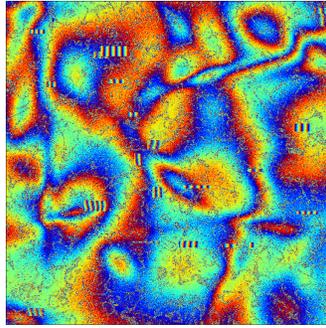
$$\epsilon_{cos}^{\Delta\theta} = \frac{1}{n} \sum_{i=1}^n \frac{1}{2} (1 - \cos(\arg(g_i \bar{f}_i))) \quad (4.3)$$

4.3.3 Key observations from few other experiments

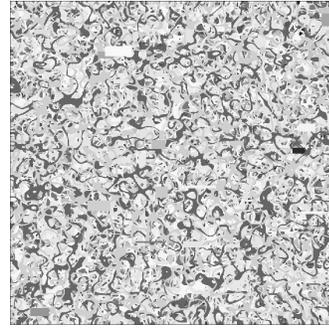
It may be interesting to note that, when we started to investigate the problem of InSAR phase filtering and coherence estimation for WAM, we tried sparse coding approaches based on learning a dictionary of features, using both optical natural images (landscapes, portraits, etc.) and InSAR images. We observed that, although the approach seems to work for optical images, it cannot be directly carried over to InSAR filtering for WAM, because of two factors. Firstly, the prediction step involves optimization, which greatly slows down the prediction performance. Especially, the approach that seemed most promising involved multi-scale prediction using dictionaries of different patch sizes. There, both the training and the prediction time increases drastically as



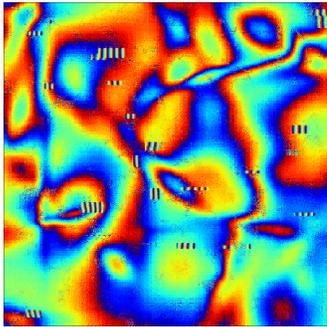
(a) Ground Truth Phase



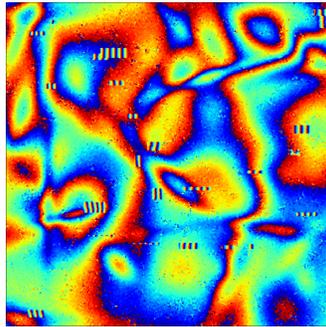
(b) Input Phase



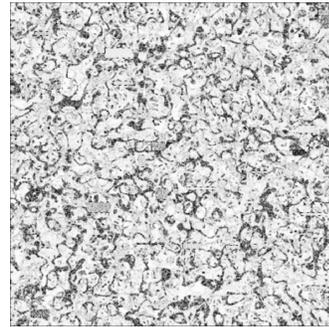
(c) Ground Truth Coherence



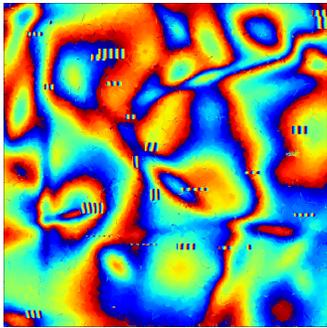
(d) Goldstein Phase



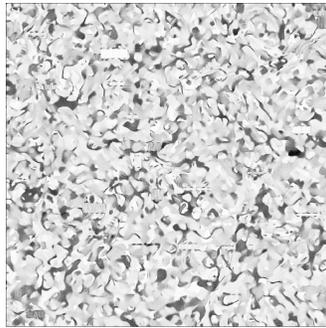
(e) Boxcar Phase



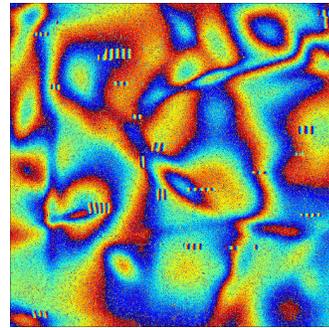
(f) Boxcar Coherence



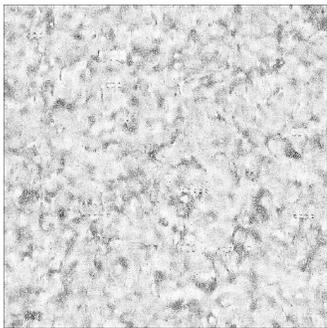
(g) NLInSAR Phase



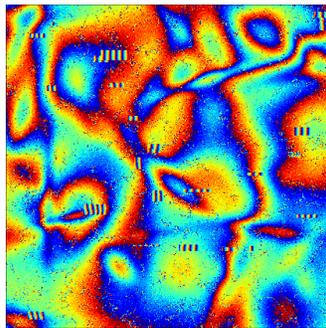
(h) NLInSAR Coherence



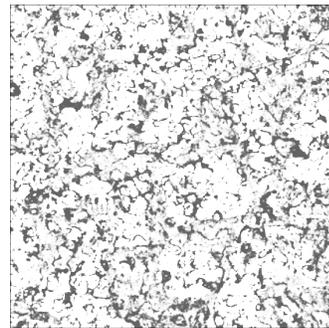
(i) NLSAR Phase



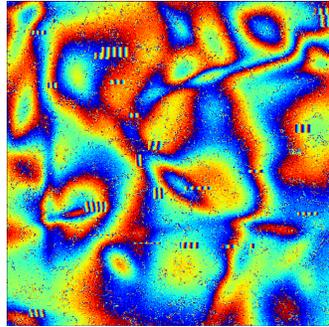
(j) NLSAR Coherence



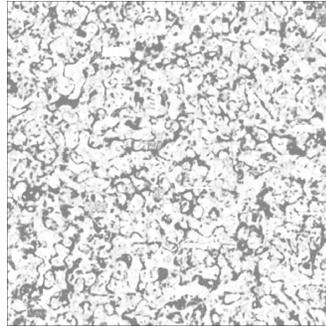
(k) CNN-InSAR (as-is) Phase



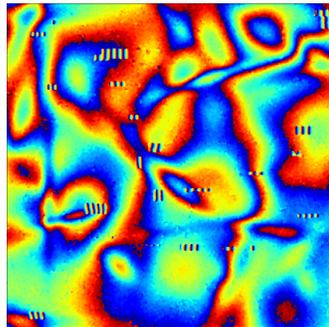
(l) CNN-InSAR (as-is) Coherence



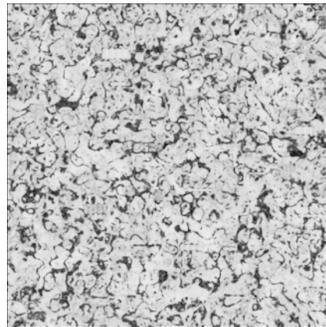
(m) CNN-InSAR (re-trained) Phase



(n) CNN-InSAR (re-trained) Coherence



(o) GenInSAR Phase



(p) GenInSAR Coherence

Figure 4.8: Filtered phase and coherence outputs for simulated InSAR images processed by GenInSAR, CNN-InSAR, and four existing methods. Visualizations for phase are coloured between $-\pi$ (blue) to $+\pi$ (red), and coherence between 0 (black: low) to 1 (white: high) respectively.

Table 4.1: Quantitative evaluation of GenInSAR, CNN-InSAR, and four existing methods, and scalability of GenInSAR over increasing GPU counts

Method Name	Phase RMSE	Coherence RMSE	Residue Red. %	Cosine Error ($\epsilon_{cos}^{\Delta\theta}$)	Time (sec)		
CNN-InSAR (as-is)	1.270 ± 0.191	0.257 ± 0.013	92.74 ± 3.30	0.060 ± 0.036	1.42		
CNN-InSAR (retrained)	1.392 ± 0.192	0.200 ± 0.025	86.91 ± 4.71	0.073 ± 0.040	1.42		
NLSAR	1.537 ± 0.073	0.301 ± 0.055	35.85 ± 17.71	0.132 ± 0.012	11.49		
NLInSAR	0.850 ± 0.122	0.159 ± 0.018	97.59 ± 2.08	0.014 ± 0.009	20.44		
Goldstein	1.260 ± 0.229	N/A	88.51 ± 11.96	0.048 ± 0.040	2.17		
Boxcar	1.025 ± 0.173	0.143 ± 0.018	97.64 ± 1.94	0.025 ± 0.021	1.32		
Proposed (GenInSAR)	0.805 ± 0.128	0.144 ± 0.025	99.66 ± 0.20	0.010 ± 0.008	20.61		
GPU Count	64	32	16	8	4	2	1
Time (sec)	0.38	0.65	1.35	2.44	5.11	9.38	20.61

we include more scales, and especially larger patch sizes. Secondly, it is difficult to have an intuitive or elegant coherence formulation using this approach, and it involves too many assumptions or heuristics, which often do not work.

4.4 Conclusion

We propose CNN-based filtering and coherence estimation for InSAR images and develop it further into a generative modelling based approach. It creates lesser phase artefacts in incoherent regions and outputs coherence clearly showing where the signal could be reliably recovered, which are desirable properties of an InSAR filter in the context of the entire InSAR processing pipeline. Our experiments demonstrate the capability of CNN-based learning for InSAR filtering and coherence estimation. Future directions of research can focus on improving GenInSAR’s coherence function to more elegantly handle those practical scenarios where the input data does not lie on the unit circle.

Chapter 5

GPU-friendly debanding for mobile HDR

5.1 Introduction

HDR imaging with 12+ bits per color channel is becoming commonplace [84]. Traditional 8-bit LDR imaging has much smaller dynamic range (peak brightness 100 nits) with narrow color gamut compared to HDR (peak brightness 1000+ nits) with wide color gamut. A visualization of how HDR compares with SDR is shown in Fig. 5.1. However, current videos are mostly distributed at 8-bit depth. Modern cameras can capture 12-/16-bit videos, but for compression or transmission, they are quantized to 8 bits. Moreover, videos are tuned for 100 nits LDR displays, which use gamma encoding¹ and Rec.709² color space. In order to watch LDR videos on HDR displays, inverse Tone-Mapping Operator (iTMO) [34], [100] must be applied. If the HDR display uses different Electro-Optical Transfer Function (EOTF) (e.g., Perceptual Quantizer [84]) and color space (e.g., DCI-P3 [107]), the iTMO can also include an EOTF [84] conversion and color space conversion. iTMO has been explored in [6], [19], [67], [101], and is not the focus of this work. iTMO-generated HDR videos often suffer from false contours called banding/ringing artifacts, arising due to the Mach band effect [39], [45]. Banding occurs especially when the iTMO is a one-to-one mapping, because 8-bit LDR video has maximum 256 codewords

¹<https://www.itu.int/rec/R-REC-BT.1886/en>

²<https://www.itu.int/rec/R-REC-BT.709/>



Figure 5.1: A visualization of how HDR compares against SDR

(only 220 codewords in Rec. 601³ and Rec. 709⁴). Hence, the output HDR video also has 256 codewords. According to [84], to show a banding-free image on a 1000+ nits display, 12-bit (i.e., 4096) codewords, is necessary. The dithering approach is an attempt to mask banding by placing a combination of pixels with different colors, within a neighborhood, so that the combination perceptually distracts the banding artefacts [90].

Adaptive dithering methods use neighborhood information in addition to pixel values, whereas non-adaptive methods use only pixel values. But these traditional dithering methods, which are performed before or during quantization, need to access the original high bit-depth and un-quantized pixel values [8], [38], [40], [51], [59], [62], [89], [90], [103], [125], [131], [132]. However, many applications require image compression or quantization, where only quantized data is available at the dithering stage. Later on, we prove mathematically

³<https://www.itu.int/rec/R-REC-BT.601/>

⁴<https://www.itu.int/rec/R-REC-BT.709/>



(a)



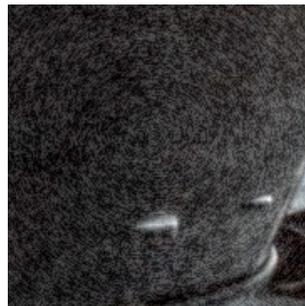
(b)



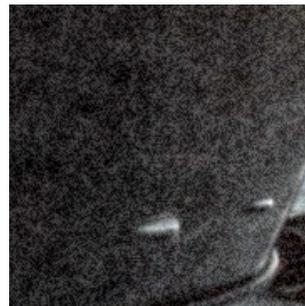
(c)



(d)



(e)



(f)

Figure 5.2: (a, b): Quantized Image and its Cropped region with banding; (c, d): Output from Proposed Method and the same Cropped; (e): Circular Noise, compared with (f): Curved Noise, which shows better blending result. Our method de-bands quantized images effectively. Note that (a-f) are tone-mapped in order to visualize banding/noise in printed form or LDR displays. This is a very challenging image in our dataset.

that while adding zero-mean noise can remove banding from un-quantized data, the same method cannot remove the banding from quantized signal. So, we need to use other noise patterns, which this work focuses on.

We propose a method to perform dithering on the quantized pixel values. One use case is where video decoder can decode 10 bit bit-stream but only output 8 bits signal. Another use case is: where an 8-bit bit-stream is received. The former case arises when the bit depth interface between two connected hardware chips are different. The latter happens for an 8-bit AVC⁵/HEVC⁶ bit stream. We especially consider mobile computing environment where computation capacity is limited and neighbor information is difficult to utilize. GPU usage is an example where having neighboring information is expensive in terms of memory I/O and timing. The inadequate neighbor information limits the usage of filtering and false contour prediction based on neighborhood information [9], [26], [55], [69], [126], [127]. To address these issues, our method is guided by single-pixel operation based dithering, without the need to detect texture-less regions. We assume that the said mobile device is HDR-capable (quite common nowadays). So, the input image would undergo inverse tone mapping to output an HDR image of significantly higher bit-depth.

It is interesting to note that finite or infinite impulse response filters that work on individual pixels across consecutive sets of frames along the temporal dimension might be explored as future work for temporal debanding, but our current focus is spatial debanding on individual frames, or in the more general case, simply HDR images. Although such filters are cheap in terms of memory and I/O, they increase dependency between GPU worker threads, especially in our use case (debanding on individual frames). So, we do not explore such filters for algorithm development or comparison in our current work.

To demonstrate the difference in number of computational operations between a pixel-based (dithering) and patch-based (filtering) approach, let us assume a Full-HD video frame size 1920×1080 with 3 channels. Even a sim-

⁵ https://www.itu.int/rec/dologin_pub.asp?lang=e&id=T-REC-H.264-201003-S!!PDF-E&type=items

⁶ https://www.itu.int/rec/dologin_pub.asp?lang=e&id=T-REC-H.265-201612-I!!PDF-E&type=items

ple convolution operation using a typically sized 5×5 filter kernel involves 25 multiplication operations followed by 1 addition operation per pixel, i.e. $1920 \times 1080 \times 3 \times 26 = 161,740,800$ operations per frame, whereas a pixel-based approach would require just over $1920 \times 1080 \times 3 = 6,220,800$ operations of adding noise to each pixel. Considering at least 24 frames per second for smooth video playback, this would translate to an additional 3,881,779,200 operations per second (3.88 Giga Floating Point Operations Per Second or 3.88 GFLOPS) for filtering compared to just 149,299,200 operations per second (0.15 GFLOPS) for dithering. Thus, even popular smartphones like Samsung Galaxy S, LG Optimus 2X, Motorola Droid Razr, Galaxy Nexus etc. cannot playback such videos, because their GPUs support only upto 3.2 GFLOPS.

Fig. 5.2(a-d) show how de-banding is performed by proposed method and Fig. 5.2(e, f) compares the circular and curved noise patterns. Note that images in Fig. 5.2 have been tone-mapped using [76] since our original HDR output images can only be effectively perceived on HDR displays. Fig. 5.2(a, b) show how banding is easily visible in texture-less regions (sky) but less visible in textured regions (airplane). Fig. 5.2d shows how noise injection depends on the degree of banding at input intensity. Our goal is to inject minimal noise so that the banding is masked but the image is visually pleasant.

For consistency, we henceforth refer LDR as SDR (Standard Dynamic range) and HDR as EDR (Enhanced Dynamic range) in this work. We consider the images to be in the $Y-C_b-C_r$ color space, where Y is luminance (luma) and C_b and C_r are the two chrominance (chroma) channels. We also consider the inverse tone mapping function as a one-to-one mapping between the SDR and EDR intensities in the Y channel. Thus, it can be computed once for each frame and stored as a look-up table, indexed by the SDR code-words. We call this look-up table as BLUT (Backward Look-Up Table) as it is used to get back the EDR image from the SDR. We assume that the original SDR uses 10 bits in each channel, and thus has 1024 code-words with integer values in the range $[0, 1023]$, but due to its quantization to 8 bits per channel, the input to proposed method has only 256 code-words. We simulate quantization by right shifting the original SDR input by 2 bits and then left-shifting

it by 2-bits. Thus, the input is still 10 bits, but has very sparsely distributed code-words (banding). The EDR uses 16 bits per channel, but its code-words are normalized to the $[0, 1)$ range. We assume that chrominance channels are inverse tone mapped using methods other than look-up tables, so we dither them without assuming any look-up tables. We assume that a range of luma intensities near the top and bottom of the code-word range are clipped, as shown in Fig. 5.3a. This is a common use case where the range of intensities is restricted to SMPTE range⁷. We will henceforth refer to the highest SDR intensity in the lower flat region as Y_0 and the lowest SDR intensity in the upper flat region as Y_1 respectively.

5.1.1 Motivation

To motivate our method, we prove mathematically that, while adding zero-mean noise can easily remove banding from un-quantized data, the same method cannot remove the banding from quantized signal.

For the proofs, we first define the following notations:

s_{ji}^y : i^{th} pixel of j^{th} frame for SDR Y channel from the video decoder (e.g. 10 bit)

n_{ji}^y : i^{th} pixel of j^{th} frame added noise for SDR Y channel

$f_j^y(\cdot)$: j^{th} frame inverse tone mapping function.

Δ : quantization step size

To prove that dithering before quantization can easily break banding, we have to prove that $E[f_j^y(s_{ji}^y)] = f_j^y(s_{ji}^y)$

Let us denote the rounded value after quantization as

$$q_{ji}^y = \left\lfloor \frac{s_{ji}^y + 0.5\Delta}{\Delta} \right\rfloor$$

Therefore, the truncated value (residual) is

$$\alpha_{ji}^y = \frac{s_{ji}^y + 0.5\Delta}{\Delta} - \left\lfloor \frac{s_{ji}^y + 0.5\Delta}{\Delta} \right\rfloor$$

Let us further define $\beta_{ji}^y = 1 - \alpha_{ji}^y$

⁷ https://kws.smpete.org/kws/public/projects/project/details?project_id=161

The noise CDF is $P(\cdot)$. The CDF for noise smaller than β_{ji}^y is $P(\beta_{ji}^y)$. Having noise smaller than β_{ji}^y will be quantized to q_{ji}^y . The CDF for noise larger than β_{ji}^y is $1 - P(\beta_{ji}^y)$. Having noise larger than β_{ji}^y will be quantized to $q_{ji}^y + 1$.

$$\begin{aligned}
E[s_{ji}^y] &= E\left[\Delta \cdot \left\lfloor \frac{s_{ji}^y + n_{ji}^y + 0.5\Delta}{\Delta} \right\rfloor\right] \\
&= \Delta \cdot \int_{-\infty}^{\infty} \left\lfloor \frac{s_{ji}^y + n_{ji}^y + 0.5\Delta}{\Delta} \right\rfloor \cdot dn_{ji}^y \\
&= \Delta \cdot \left(\int_{-\infty}^{\beta_{ji}^y} q_{ji}^y \cdot dn_{ji}^y + \int_{\beta_{ji}^y}^{\infty} (q_{ji}^y + 1) \cdot dn_{ji}^y \right) \\
&= \Delta \cdot (q_{ji}^y \cdot P(\beta_{ji}^y) + (q_{ji}^y + 1) \cdot (1 - P(\beta_{ji}^y))) \\
&= \Delta \cdot (q_{ji}^y + 1 - P(\beta_{ji}^y))
\end{aligned}$$

If $P(\cdot)$ is uniform between $[-0.5, 0.5]$, then we can get back to original signal, i.e., $E[s_{ji}^y] = s_{ji}^y$

Then,

$$E[f_j^y(s_{ji}^y)] = f_j^y(s_{ji}^y)$$

We will now extend the above proof to deduce that dithering using local zero-mean noise like simple Gaussian noise cannot break the banding. In the computing environment we are considering, the most effective noise injection can be during inverse tone mapping. Thus, the input, X to the inverse tone mapping function is $X = \Delta \cdot \left\lfloor \frac{s_{ji}^y + 0.5\Delta}{\Delta} \right\rfloor$

The output from the inverse tone mapping function is $f_j^y(X)$

So, the expected value of the inverse tone mapping function is $E[f_j^y(X)] = f_j^y\left(\Delta \cdot \left\lfloor \frac{s_{ji}^y + 0.5\Delta}{\Delta} \right\rfloor\right)$ since the signal and noise are independent, and noise is zero-mean. Adding noise will not affect the quantized results. Thus, we know having a local zero-mean noise cannot mask the banding artifact. Even if we add Multimodal Random Number Dithering noise, which is still a local zero-mean noise, the banding artifact cannot be removed.

5.2 Proposed Method

Our proposed method considers different regions of the BLUT which are mutually exclusive. An example of BLUT is shown in Fig. 5.3a, where the horizontal axis has SDR code-words and the vertical axis has the normalized EDR code-words.

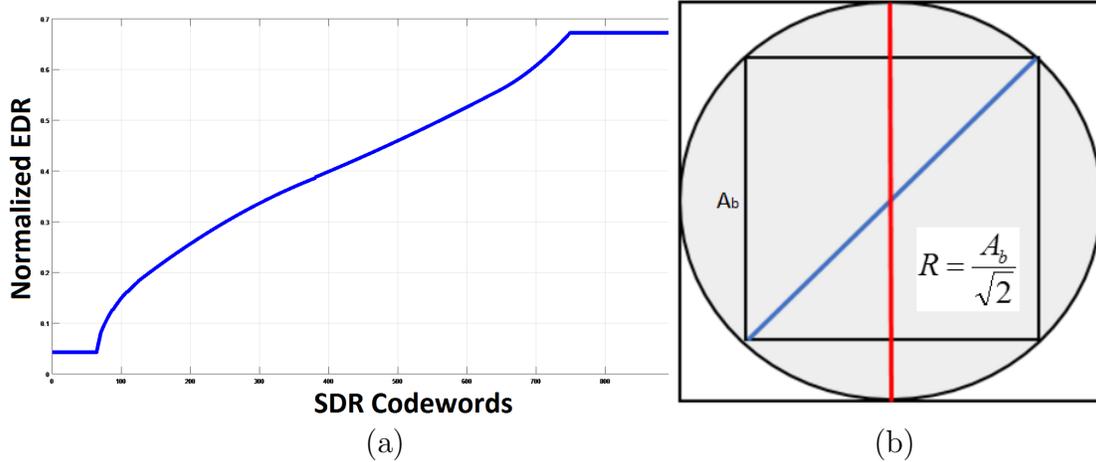


Figure 5.3: (a) BLUT for Y channel from a sunset scene b) Extraction of square block of circular pattern noise

We refer to SDR intensities lesser than Y_0 as ‘lower flat region’ and greater than or equal to Y_1 as ‘upper flat region’. Experiments showed that the smallest SDR intensity, whose corresponding normalized EDR intensity greater than the ‘0.625’ can be treated as the starting intensity for highlight region. For example, for 4000 nits display, this EDR intensity would be 3000 nits. Let the corresponding SDR intensity be denoted by ‘ Y_h ’. Let all individual SDR intensities be denoted by t and the set of all SDR intensities be denoted by $T = \{t_0, t_1, t_2, \dots, t_{1023}\}$. Let the ‘lower flat region’, $BLUT_{Down} = \{t < Y_0\}$. Let the ‘low-lights & mid-tones region’, $BLUT_{Mid} = \{Y_0 \leq t < Y_h\}$. Let the ‘highlights region before reaching the upper flat region’, $BLUT_{High} = \{Y_h \leq t < Y_1\}$. Let the ‘upper flat region’, $BLUT_{Up} = \{t \geq Y_1\}$. Thus, we partition the BLUT into 4 separate regions: $T = \{BLUT_{Down} \cup BLUT_{Mid} \cup BLUT_{High} \cup BLUT_{Up}\}$

5.2.1 Markov Gaussian Noise

The most straightforward dithering method is adding simple Gaussian noise to images. Our experiments showed that, increasing the mean and standard deviation of the Gaussian distribution perceptually masks banding artefacts, but makes the image progressively noisier. If we add *global non-zero mean* noise to the image, the image may appear to be darker or brighter depending

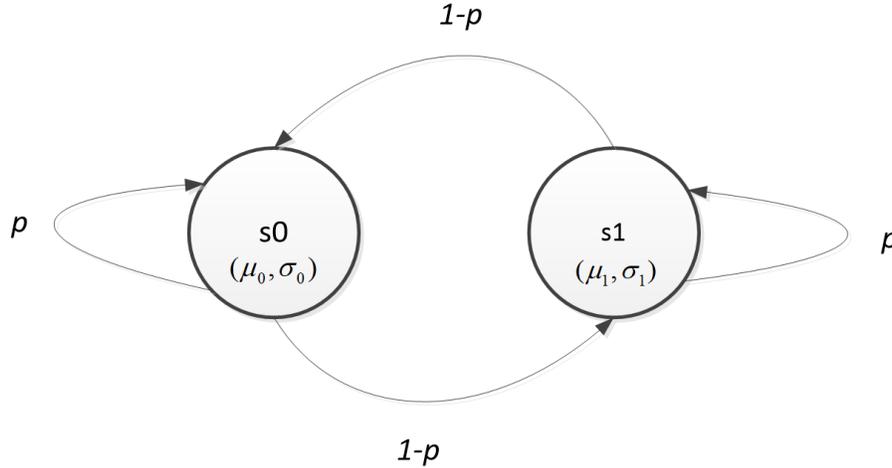


Figure 5.4: Two-state Markov-Gaussian Noise generator

on whether the mean is negative or positive, respectively. On the other hand, if we add *local zero-mean* noise, it cannot break the banding artefacts. So, in the proposed method we create noise patterns which have a global zero mean property, although locally the noise pattern has non-zero means to break banding. Using Markov chain to build state transition, where each state has non-zero mean noise can help to achieve this local non-zero mean goal.

As shown in Fig. 5.4, while applying the Markov Gaussian model of noise generation to each pixel, we need to know that pixel's previous state. Based on that, with an intra-state probability p , we can stay in the same state, or with an inter-state probability $1-p$ to move to another state. In each state, we generate different Gaussian noise with different parameters (non-zero-mean).

Consider a two-state Markov chain, where each state has intra-state transition probability p and inter-state $1-p$. In each state, s , we can generate a Gaussian noise with mean μ_s , variance σ_s , i.e. (μ_0, σ_0) for state 0 and (μ_1, σ_1) for state 1 . We use the value $\mu_0 = 2, \sigma_0 = 1, \mu_1 = -2, \sigma_1 = 1$ in our implementation, as these values gave the best results based on our experiments.

As we can see by comparing the patterns in Fig. 5.5, higher intra-state transition probability can generate longer texture which, according to our experimental observations, has better ability to destroy the banding artefact, but at the same time makes the image noisier.

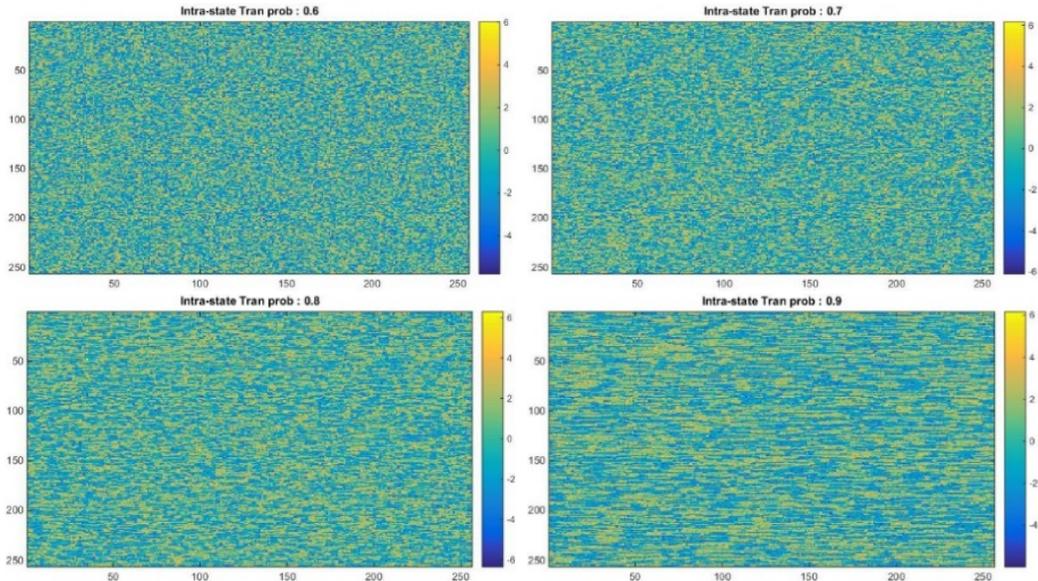


Figure 5.5: Two-state Markov Gaussian Noise with intra-state transition probabilities 0.6 (top-left), 0.7 (top-right), 0.8 (bottom-left), and 0.9 (bottom-right)

5.2.2 Enhancement of Markov Gaussian Noise

Orientation of Noise Stripes

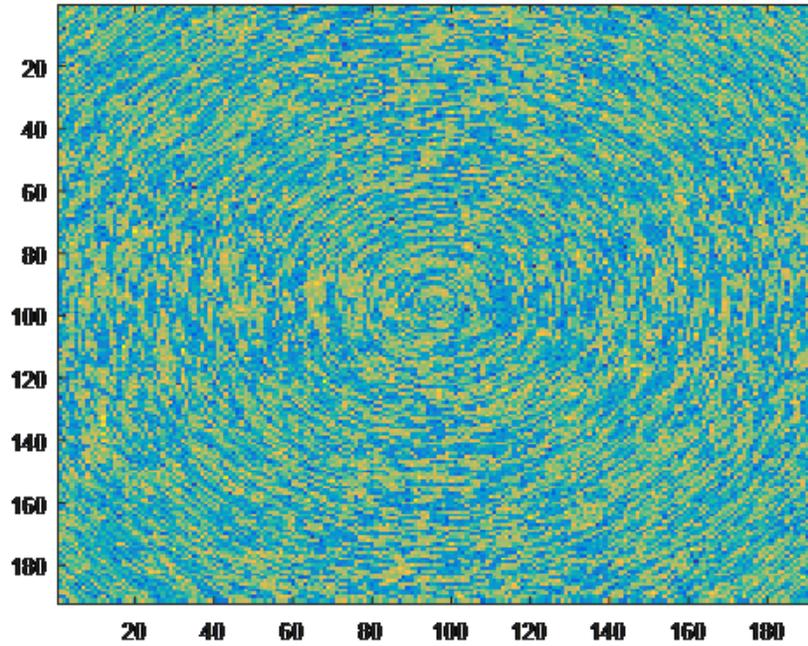
A limitation of the 2-state Markov Gaussian noise is that it can generate noise patterns (stripes) in only two fixed directions: horizontal and vertical. But, it was observed experimentally that this looks unnatural, and the effectiveness of de-banding also depends on the angle at which the stripes meet the simulated contours. So, we proceed to make the noise pattern curved instead of straight, in order to ensure that it meets the contours at all possible orientations (angles) and also blends with the image content, thus looking more natural. To generate curved patterns using the 2-state Markov Gaussian noise generator, we first create concentric circles of decreasing radius R , $(R-1)$, $(R-2)$... $(R-N_m)$, where N_m is an integer lesser than R . In our implementation, we use $N_m = (R-1)$. Since the value of each pixel on the circumference of each circle will map to a corresponding noise value output by the 2-state Markov Gaussian noise generator, we determine in Eq. 5.1 the length, ‘ L ’ of the sequence which the noise generator should generate to construct a $A_b \times A_b$ square noise block.

$$L = 2\pi \sum_{v=0}^{N_m} \left(\left(\frac{A_b}{\sqrt{2}} \right) - v \right) \quad (5.1)$$

Typical value of $R = 200/\sqrt{2} \sim 142$ (we use square blocks of each side length, $A_b = 200$). This metric is illustrated in Fig. 5.3b. Radius of the biggest circle is R . Its diameter is $2R$, which is also the length of each side of the larger square. The length of each side of the smaller square is A_b , which is ultimately extracted as a square block of circular noise pattern. To be more specific, to create the circle of radius R , per the properties of a circle inscribed inside a square, we need to inscribe it in a square matrix whose length of each side = $2R$. The magnitude of the greatest radius, R is also related to the desired size of the noise block. Specifically, it can be shown from the geometric properties of a square inscribed in a circle that, if the length of each side of the square (block) is ' A_b ' then the radius of the biggest circle is ($R=A_b/\sqrt{2}$) and its circumference, $C = 2\pi * (A_b/\sqrt{2})$. Following this approach, for each circle, starting with the point on its circumference subtending the least angle θ (counter-clockwise) w.r.t the horizontal axis, we copy the output of the noise generator to that pixel. Similarly, for the next pixel on the circumference in increasing order of θ , we copy the next output of noise generator, and so on. Fig. 5.6a shows an example of circular noise pattern. Although Fig. 5.6 uses one particular value of transition probability (0.815), the proposed method actually adaptively determines this value based on BLUT slope. Thus, this value is dictated by the BLUT of the current image.

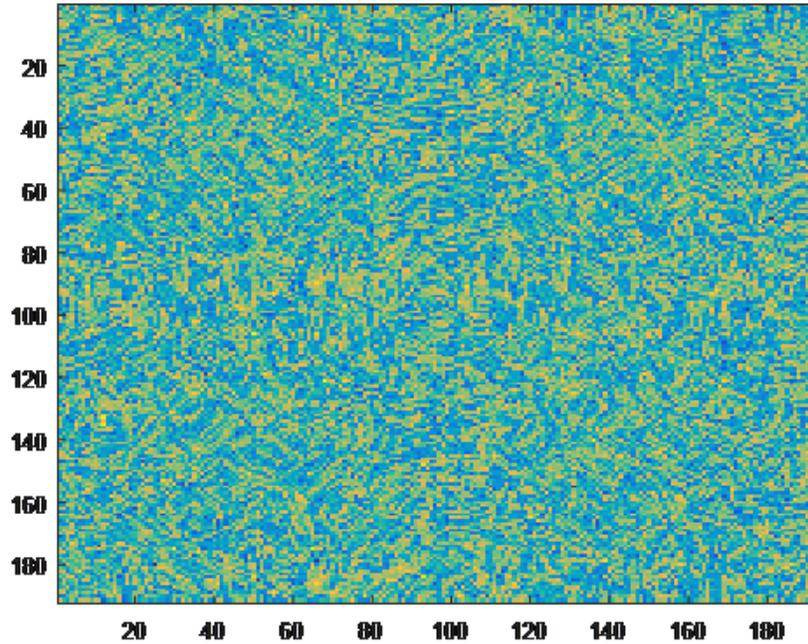
To generate a curved pattern from a circular pattern, we first need to partition the square block of circular noise into four equal-sized *quadrants* or sub-blocks. We treat each quadrant as an independent image and choose a set of points as the sites for performing Voronoi tessellation for generating the irregular sized patches (Voronoi cells). A Voronoi diagram (Fig. 5.7) is a partitioning of a plane into regions based on distance to points in a specific subset of the plane. That set of points (called seeds, sites, or generators) is specified beforehand, and for each seed there is a corresponding region consisting of all points closer to that seed than to any other. These regions are called Voronoi

200 x 200 block circular noise with Transition Probability = 0.815



(a)

200 x 200 block Curved noise with Transition Probability = 0.815



(b)

Figure 5.6: Convert (a) Circular to (b) Curved Markov-Gaussian Noise; Block size = 200×200 ; Trans. Prob. = 0.815

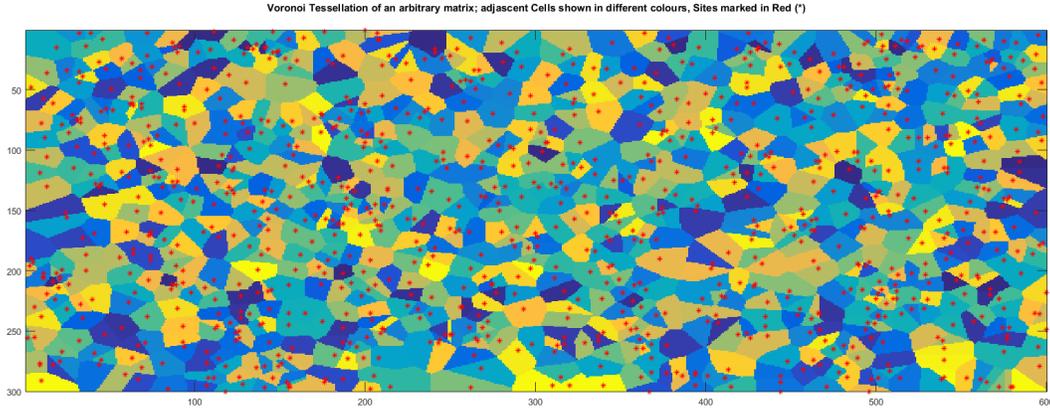


Figure 5.7: Voronoi Tessellation of arbitrary matrix; adjacent cells: shown in different colors; sites (*): red

cells. Of course, in the finite sense, the plane represents a matrix, and the sites represent discrete locations inside the matrix, as shown in the diagram below. Here the sites are marked in red asterisks and the adjacent Voronoi cells are shown in as different colors as possible just for ease of visualization.

In our proposed method, while generating the curved noise pattern from the circular pattern, we cannot simply divide each quadrant into identical sized square- or rectangular-shaped patches to replace them with each other, as this would result in blocky artifacts in the noise matrix, as observed in our experiments. To overcome this issue, we need to use Voronoi Tessellation to create irregular-shaped blocks (Voronoi cells) and replace them with each other. The *number of cells* is determined by the number of sites, and for our case, we made it proportional to the size of the image. The location of the sites must be random for each frame, so that the noise pattern does not remain fixed for all frames. Experiments showed that a fixed noise pattern with changing frame contents (video) looks unnatural. The *shape of the cells* is determined by the distance metric used. Experimental results showed that the *city-block* distance metric gave the best results. The process of tessellation is *deterministic*, i.e., for the same number and location of sites and same distance metric, the output of tessellation will not change. This is why we apply identical tessellation to each quadrant of the circular noise block, so that the resultant patches (cells) have one-to-one correspondence with each other across quadrants. Specifically,

we select the same set of points for each quadrant. Since Voronoi tessellation is a deterministic process, we now have one-to-one correspondence between the cells in all quadrants with respect to their shapes and sizes. Now, for each cell of each quadrant, we replace its contents with those of any randomly selected cell from its corresponding set of four co-located cells across all quadrants. We then integrate the quadrants to get the curved pattern block shown in Fig 5.6b. We generate such blocks and concatenate them to get the noise matrix which is added to the quantized image. Experiments showed that a block size of 200×200 gave good results. If we would have simply concatenated blocks of circular (not curved) pattern to create the noise matrix, then the dithered image would have had an impression of the patterns of circles and looked ugly, as seen from our experiments. Our experiments also showed that even if we concatenate different sized blocks of circular patterns of different transition probabilities, or even irregular shaped blocks (using tessellation), we can still sometimes see impressions of long circular arcs on the dithered image, if we observe closely on the EDR display. This was the basic motivation behind making the noise stripes curved, like short arcs of a circle. Typical number of sites (N_s) is 300. Generating a block of circular noise offline takes 11.79 seconds, and converting that block to curved noise real-time, with randomly initialized site locations for each image, only takes 0.11 seconds. Randomness ensures that the noise pattern changes naturally across video frames. The serial execution time per frame is 4.36 seconds on Matlab running on a Windows 10, AMD 4.2 MHz, 16 GB RAM PC.

Noise Variance and Transition Probability

The transition probability of the 2-state Markov Chain may become a bottleneck in determining the effectiveness of de-banding; a higher transition probability will do a better de-banding, but make the image noisier. Similarly, we can increase the variance of the noise to get rid of more banding, but making the image noisier. We alter the noise variance as

$$D = Q + s \cdot N_p \tag{5.2}$$

where ‘ s ’ is the noise variance and N_p is the matrix containing the generated noise pattern. Q is the image after quantization and has banding artefacts. We add noise to Q to produce the dithered image D , to reduce banding. In the following, we label the type of texture generated from the transition probability function as *pattern* of the noise. To reduce computation, we store in memory noise patterns having transition probabilities

$$P_T = \{0.545 + k * 0.045\} \text{ where } k = [0, 1, 2 \dots 9].$$

An effective way to compute the transitional probability for any given intensity is to use the slope of the Backward Look-Up Table (BLUT) at that intensity as an indication of the degree of banding likely to be present for that intensity. The noise pattern for all pixels having that intensity will be selected according to this slope, fetched from memory, and added to those pixels. The fine granularity of probabilities in P_T ensures that the aggregated noise pattern applied finally to the entire quantized image does not have abrupt variations in spatially adjacent regions. The rationale behind this assumption is that, wherever the BLUT has a steep slope, adjacent SDR intensities are mapped to EDR intensities which are far apart from each other, and thus, more prone to introduce banding artefacts in the backward reshaped (EDR) image. Thus, wherever the banding is more noticeable, we need to use more structured noise for de-banding, i.e., higher intra-state transition probability noise.

Computing the BLUT slope at individual intensities is neither efficient nor effective; hence we calculate the slope over a range of intensities. Experiments showed that the best results were obtained when this range is eight consecutive SDR intensities, as well as when different operations (noise variance, noise patterns) are applied in each BLUT range. We elaborate this BLUT slope computation as follows:

Let us denote EDR intensity corresponding to SDR intensity t as $E_t = BLUT(t)$. To calculate the slope, Z_t at SDR intensity t , we first select every 8th SDR intensity starting from 0. Let this set of intensities be denoted by $W = \{ t_0, t_7, t_{15}, \dots, t_{1023} \}$. Next, we get the EDR intensities corresponding to these SDR intensities as $E_W = \{ E_{t_0}, E_{t_7}, E_{t_{15}}, \dots, E_{t_{1023}} \}$. The slope, Z_t at SDR intensity t is computed following the below steps:

1. For each SDR intensity $t \in T$, find the biggest $t_p \in W$ and smallest $t_n \in W$ such that $\{t_p \leq t \leq t_n\}$
2. $Z_t = |E_{t_n} - E_{t_p}| //$ absolute difference

Even though $slope = \frac{rise}{run}$, we do not divide Z_t (rise) by $t_n - t_p$ (run), because the difference of t_n and t_p will always be 8.

Based on slope of BLUT, we define the transition probability determination factor G_t at SDR intensity t as follows:

$$G_t = \log_2(1 + t_{1023} * Z_t) \quad (5.3)$$

Since the Z_t value is a fraction, we multiply it with the highest SDR intensity and add 1 to the result to ensure that the input to the logarithm is non-negative. The logarithm itself ensures that G_t does not increase abruptly when the BLUT rises steeply, otherwise there may be big variations in noise pattern between adjacent image regions. In Eq. 5.3, if $Z_t = 0$ then $G_t = 0$, its minimum possible value. Thus, flat regions of BLUT still have noise, but with minimum transition probability. We cannot put zero noise in those regions, as that may create stark contrast with the surrounding regions having noise, and thus make the image look unnatural. Finally, the transition probability index p^{opt} to be used for intensity t is determined as shown in Eq. 5.4.

$$\begin{aligned} p^{opt} &= \arg \max \{G_t > p \cdot P_L\} \forall t \in BLUT_{Mid} \\ &= \arg \max \{G_t > p \cdot P_H\} \forall t \in BLUT_{High} \end{aligned} \quad (5.4)$$

where $P_L = 1$ and $P_H = 3$ ensure that we favor lower transition probabilities for the highlight region. This is because we want to make the highlight region less noisy, as experiments proved that noise in the highlights is more easily visible. Also, we use the horizontal pattern (Fig. 5.5) in $BLUT_{High}$ and $BLUT_{Up}$ due to its lesser visibility (than curved pattern) in high intensities.

We use a fixed minimum value of variance for the flat regions $S_{min}^{Luma}, S_{min}^{Chroma}$ and constant variance S_C^{Luma}, S_C^{Chroma} for $BLUT_{Mid}$. In $BLUT_{High}$, variance falls from S_C to S_{min} per Eq. 5.5.

Table 5.1: BLUT-Aware Variance and Noise Pattern

Range	Luma Noise Injection	Chroma Noise Injection
BLUT _{Down}	Variance = $S_{\min}^{Luma} = 2$ Curved Noise Pattern Trans. Prob. = $P_T[0]$	Variance = $S_{\min}^{Chroma} = 1$ Curved Noise Pattern Trans. Prob. = $P_T[5]$
BLUT _{Mid}	Variance = $S_C^{Luma} = 5$ Curved Noise Pattern Transition Probability depends on BLUT slope	Variance = $S_C^{Chroma} = 2.5$ Curved Noise Pattern Trans. Prob. = $P_T[5]$
BLUT _{High}	Variance decreases gradually from 5 to 2 Horizontal Noise Pattern Transition Probability depends on BLUT slope	Variance is modulated by luma, and variance decreases gradually from 2.5 to 1 Horizontal Noise Pattern Trans. Prob. = $P_T[5]$
BLUT _{Up}	Variance = $S_{\min}^{Luma} = 2$ Horizontal Noise Pattern Trans. Prob. = $P_T[0]$	Variance = $S_{\min}^{Chroma} = 1$ Horizontal Noise Pattern Trans. Prob. = $P_T[5]$

$$\begin{aligned}
 s_t^{Luma} &= S_C^{Luma} \left(\frac{S_{\min}^{Luma}}{S_C^{Luma}} \right)^{\frac{(t-Y_h)}{(Y_1-Y_h)}} \\
 s_t^{Chroma} &= S_C^{Chroma} \left(\frac{S_{\min}^{Chroma}}{S_C^{Chroma}} \right)^{\frac{(t-Y_h)}{(Y_1-Y_h)}}
 \end{aligned} \tag{5.5}$$

where s_t is noise variance at intensity t for $t \in BLUT_{High}$. Both luma and chroma noise variance are decreased from C to min over the same range $BLUT_{High}$ for any frame. So, chroma noise variance is modulated by luma. The constants S_C and S_{min} were obtained experimentally. Table 5.1 summarizes these points. It may be noted that we use lower variance in chroma channel to inject less noise, as we observed much less banding in chroma. Thus, from the above discussion and Table 5.1, we can infer that both the noise variance and transition probability are bounded, and neither of them are linear over the entire BLUT range.

5.3 Results and Discussion

5.3.1 Dataset

In order to understand the impact of banding, and the factors affecting the design and testing of our proposed method, we observed on a 4000 nits EDR display, Y , C_b , and C_r channels of all frames extracted from five sequences created at Dolby Vision research lab. We classified the banding into nine categories, based on the following criteria: (a) Degree of banding (weak, moderate, strong), and (b) Luma intensity at which banding is present (dark regions, mid-tones, highlights). Accordingly, we selected 20 scenes from those five sequences. Each scene is composed of six consecutive frames. Thus, our dataset has 120 frames from different categories.

5.3.2 Quality of Dithered Images

We verified by running the proposed method on our dataset, that the noise pattern and variance change smoothly across successive frames, which is important in order to deliver videos in good visual quality, and that there are no detrimental side-effects of proposed dithering method on image quality. It should be noted that comparing the output of proposed method to other de-banding methods which use pixel neighborhood information or sparse / dense filters instead of dithering would be unfair because of their inapplicability to our use cases (GPUs, mobile devices) as briefly mentioned earlier, because simple pixel-based operations and low resource requirements are essential for our use cases. Also, comparisons to other inverse tone mapping methods would be pointless as inverse tone mapping is not the aim of our proposed method.

5.3.3 Subjective Testing

Subjective testing was conducted on 12 subjects with computing science and engineering backgrounds. Pairs of images were shown on the 4000 nits EDR display and subjects were asked to compare the images and choose their preferred image using the Google Form on their smart phones, with the lights off in the room. The questionnaire is shown in Fig. 5.8.

Image Comparison Test 1

* Required

Pair-1: Which image do you prefer ? *

Image A

Image B

NO preference

Your preferred image is

Much better than the other one

Slightly better than the other one

Your preferred image has

Lesser banding than the other one

Lesser noise than the other one

 Page 1 of 5

Figure 5.8: Google Forms Questionnaire for Subjective Tests

Table 5.2: Image pair displayed randomly to subjects

Test-ID	ImageA	ImageB
Test-1	Simple Gaussian Noise	Proposed Adaptive Method Noise
Test-2	Proposed Adaptive Method Noise	Non-adaptive Method Noise (curved)
Test-3	Low-Pass Filtered Gaussian Noise	Proposed Adaptive Method Noise

The questionnaire asks the subject which of the two images is preferred, or has no preference. In the former case, the subject is asked whether the preferred image is slightly better or much better than the other, and (optionally) whether the preferred image has lesser banding and/or lesser noise than the other one, so that later, we could account for human errors during the rating by the subjects. Initially, the subjects were trained by showing sample images with different degrees of banding in different intensity ranges, to make them familiar with the criteria for their opinion. For each of the three actual tests, five image pairs were shown in different order in each test, per Table 5.2, though these details were not revealed to subjects during subjective testing to avoid biasedness. It should be noted that, although we refer to our proposed method as ‘adaptive’, it simply means BLUT-slope modulated dithering, and not usage of neighborhood information at run-time like traditional adaptive dithering methods mentioned earlier. The non-adaptive version of our proposed method uses the curved noise for all intensities of luma and chroma, with fixed noise variance and fixed transition probability.

The subjective opinion scores were assigned these values:

1. If a method is preferred by the subject as “much better than the other method” then it gets a score of +2
2. If a method is preferred by the subject as “slightly better than the other method” then it gets a score of +1
3. In case of “NO preference”, both methods get a score of 0

The test results in terms of DMOS are shown in Table 5.3. Positive DMOS values and confidence interval plots at significance level $\alpha = 0.05$ shown in

Table 5.3: DMOS scores from Subjective Test results

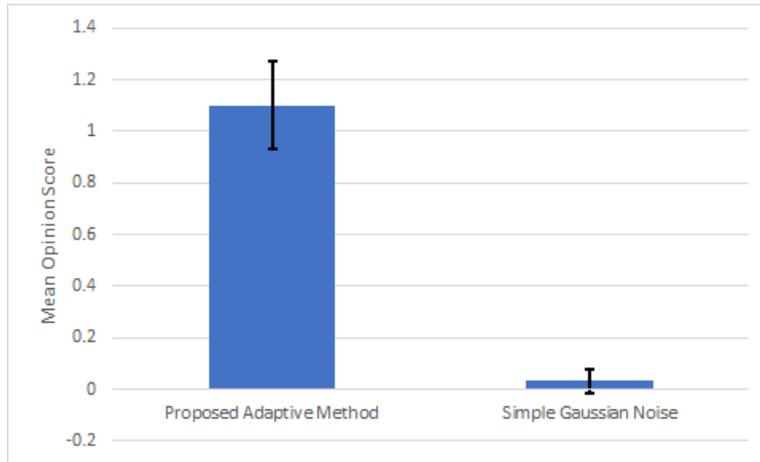
Test-ID	Differential Mean Opinion Scores (DMOS)
Test-1	1.066666667
Test-2	0.716666667
Test-3	1.033333333

Fig. 5.9 indicate that proposed adaptive method outperformed other methods, namely, dithering using simple Gaussian, non-adaptive method (curved noise) and low-pass filtered Gaussian noise.

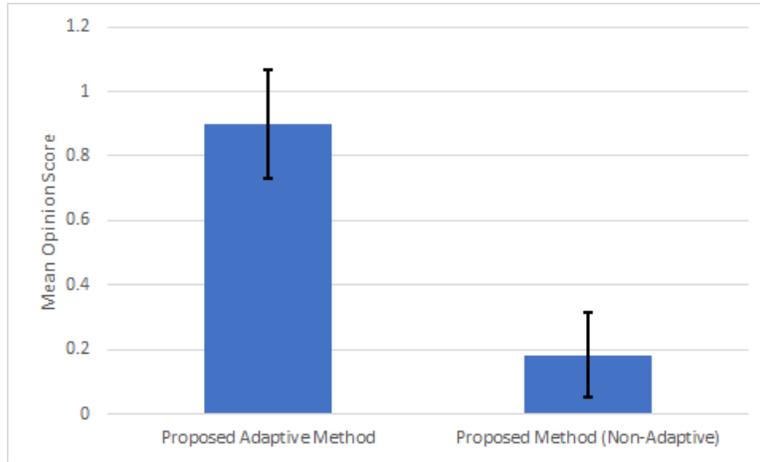
5.4 Conclusion

In this work, we proposed a method for pixel-based, adaptive dithering, modulated by slope of the inverse tone mapping function. Our goal is to work in a resource constrained and time efficient environment, so we did not use neighborhood information during run-time. Our method achieves significantly better de-banding compared to simple Gaussian noise with and without low-pass filter, as well as the non-adaptive version of proposed method. Better results can potentially be obtained by dense- and sparse-filter based methods, or dithering methods which utilize spatial location information of pixels at run time. But these methods are more computationally resource intensive.

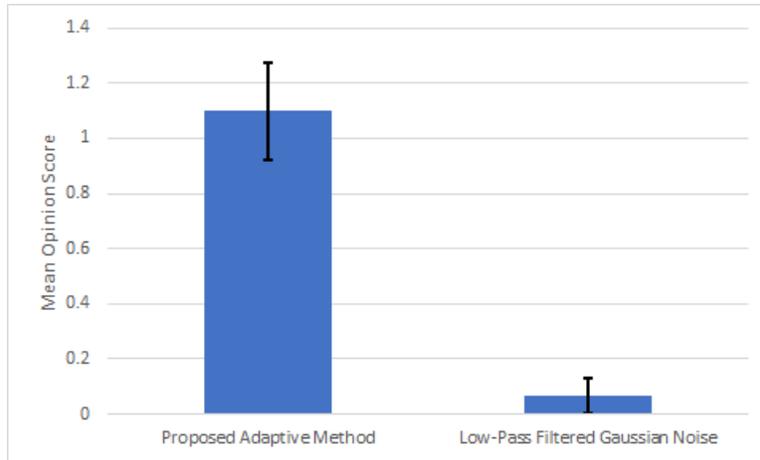
Our research also opens up new directions in investigating whether noise pattern should be different in image regions of higher luma intensity and color saturation, as opposed to regions of neutral color. This may involve fine-tuning variance of BLUT-modulated noise, based on color saturation. Lastly, since we separate the (offline) noise pattern generation and (online) noise injection modules in our proposed method design, we thus pave the way for future researchers to use our noise patterns in conjunction with their own noise injection schemes, thereby potentially opening up new avenues of research.



(a) Test-1 confidence interval plot



(b) Test-2 confidence interval plot



(c) Test-3 confidence interval plot

Figure 5.9: Confidence interval plots at significance level $\alpha = 0.05$ showing superior performance of proposed adaptive method.

Chapter 6

Conclusion and Future Work

In this thesis, we explored image abnormality detection/restoration scenarios involving different sensor modalities, where existing methods were insufficient to address the computational and signal processing challenges involved due to scalability requirements or availability of limited contextual information. Here we suggest potential future directions which interested researchers can continue to explore in these fields.

In WMI detection from preterm brain MRIs, we addressed the lack of reliable atlases using heuristics and more generic assumptions about human brain structure, and developed a segmentation-free method which also drastically reduces computation. The future directions of research would involve exploring more clinical investigations into the correlations between WMI and patient information like demographic factors, and how those could be incorporated to improve the WMI detection process.

In InSAR phase filtering and coherence estimation for WAM, we showed the promise of CNN-based unsupervised learning solutions to address the lack of clean training data, which naturally lend to parallelism using GPUs, thus addressing the data volume concerns of WAM. We also showed how CNN-based generative modelling can not only improve over existing methods, but also has auxiliary benefits like helping in data augmentation for machine learning InSAR applications. The future direction involves improving the coherence estimation of our proposed GenInSAR method to better handle those practical scenarios where the the input data does not lie on the unit circle.

As an aside from the purely CNN-based InSAR filtering described in this thesis, we briefly explored the possibility of using GPU implementations of traditional denoising algorithms in conjunction with CNN-based methods for tuning the former’s parameters based on the input image [87]. This approach combines the explain-ability of traditional computer vision based denoising methods with the data-driven processing of CNNs. It might be interesting to develop and extend this type of approach to InSAR images and compare performance. Similarly, it might be interesting to develop a coherence classification method based on GenInSAR and compare its performance with that of the CNN-InSAR-based one [88] we briefly worked upon as auxiliary research.

In dithering-based debanding suitable for mobile GPUs, we addressed the lack of neighborhood information, lack of access to the original (un-quantized) image and computational constraints of a mobile environment by designing a pixel-based solution to debanding based on novel noise pattern generation and efficient noise injection based on the intensity profile of the image and the slope of the iTMO. Future directions of research could involve investigations into whether noise pattern should be different in image regions of higher luma intensity and color saturation, as opposed to regions of neutral color. This may involve fine-tuning variance of BLUT-modulated noise, based on color saturation. Lastly, we separate (offline) noise pattern generation from (online) noise injection based on simple pixel operations in our proposed method design. This not only makes the proposed solution efficient and feasible to implement on a mobile environment, but also paves the way for future researchers to use our noise patterns in conjunction with their own noise injection schemes, thereby potentially opening up new avenues of research.

References

- [1] N. Anantrasirichai, J. Biggs, F. Albino, and D. Bull, “A deep learning approach to detecting volcano deformation from satellite imagery using synthetic datasets,” *Remote Sensing of Environment*, vol. 230, p. 111 179, 2019, ISSN: 0034-4257.
- [2] C. Asao, Y. Korogi, Y. Kondo, T. Yasunaga, and M. Takahashi, “Neonatal periventricular-intraventricular hemorrhage: Subacute and chronic MR findings,” *Acta Radiologica*, vol. 42, no. 4, pp. 370–375, Jul. 2001.
- [3] G. Baier, C. Rossi, M. Lachaise, X. X. Zhu, and R. Bamler, “A nonlocal insar filter for high-resolution dem generation from tandem-x interferograms,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 11, pp. 6469–6483, Nov. 2018, ISSN: 1558-0644.
- [4] G. Baier. (). Github - gbaier/despeckcl: A c++ library for insar denoising images using opencl, [Online]. Available: <https://github.com/gbaier/despeckCL> (visited on 12/17/2019).
- [5] P. Ballabh, “Intraventricular hemorrhage in premature infants: Mechanism of disease,” *Pediatric Research*, vol. 67, no. 1, pp. 1–8, Jan. 2010.
- [6] F. Banterle, P. Ledda, K. Debattista, and A. Chalmers, “Inverse tone mapping,” in *Proceedings of the 4th International Conference on Computer Graphics and Interactive Techniques in Australasia and South-east Asia*, ser. GRAPHITE '06, Kuala Lumpur, Malaysia: ACM, 2006, pp. 349–356, ISBN: 1-59593-564-9.
- [7] I. Baran, M. P. Stewart, B. M. Kampes, Z. Perski, and P. Lilly, “A modification to the goldstein radar interferogram filter,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 9, pp. 2114–2118, Sep. 2003, ISSN: 0196-2892.
- [8] B. Bayer, “An optimum method for two-level rendition of continuous-tone pictures,” *Ineternl. Conf. on Comm.*, vol. 50, pp. 69–77, 1976.
- [9] S. Bhagavathy, J. Llach, and J. f. Zhai, “Multi-scale probabilistic dithering for suppressing banding artifacts in digital images,” in *2007 IEEE International Conference on Image Processing*, vol. 4, Sep. 2007, pp. IV - 397-IV –400.

- [10] Y. Bian and B. Mercer, "Interferometric sar phase filtering in the wavelet domain using simultaneous detection and estimation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 4, pp. 1396–1416, Apr. 2011, ISSN: 0196-2892.
- [11] M. Bilello, M. Arkuszewski, P. Nucifora, I. Nasrallah, E. R. Melhem, L. Cirillo, and J. Krejza, "Multiple sclerosis: Identification of temporal changes in brain lesions with computer-assisted detection software," *The Neuroradiology Journal*, vol. 26, no. 2, pp. 143–150, Apr. 2013.
- [12] J. Bioucas-Dias, V. Katkovnik, J. Astola, and K. Egiazarian, "Absolute phase estimation: Adaptive local denoising and global unwrapping," *Appl. Opt.*, vol. 47, no. 29, pp. 5358–5369, Oct. 2008.
- [13] C. M. Bishop, "Mixture density networks," Aston University, Birmingham, United Kingdom, Tech. Rep. NCRG/94/004, Feb. 1994.
- [14] M. Cabezas, A. Oliver, E. Roura, J. Freixenet, J. C. Vilanova, L. Ramió-Torrentà, À. Rovira, and X. Lladó, "Automatic multiple sclerosis lesion detection in brain MRI by FLAIR thresholding," *Computer Methods and Programs in Biomedicine*, vol. 115, no. 3, pp. 147–161, Jul. 2014.
- [15] R. A. Castellino, "Computer aided detection (CAD): An overview," *Cancer Imaging*, vol. 5, no. 1, pp. 17–19, 2005.
- [16] T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Transactions on Image Processing*, vol. 10, no. 2, pp. 266–277, Feb. 2001, ISSN: 1057-7149.
- [17] C. Chao, K. Chen, and J. Lee, "Refined filtering of interferometric phase from insar data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 12, pp. 5315–5323, Dec. 2013, ISSN: 0196-2892.
- [18] V. Chau, K. J. Poskitt, D. E. McFadden, T. Bowen-Roberts, A. Synnes, R. Brant, M. A. Sargent, W. Soulikias, and S. P. Miller, "Effect of chorioamnionitis on brain development and injury in premature newborns," *Annals of Neurology*, vol. 66, no. 2, pp. 155–164, Aug. 2009.
- [19] Q. Chen, G.-M. Su, and P. Yin, "Near constant-time optimal piecewise LDR to HDR inverse tone mapping," in *Digital Photography XI*, N. Sampat, R. Tezaur, and D. Wüller, Eds., SPIE, Feb. 2015.
- [20] I. Cheng, S. P. Miller, E. G. Duerden, K. Sun, V. Chau, E. Adams, K. J. Poskitt, H. M. Branson, and A. Basu, "Stochastic process for white matter injury detection in preterm neonates," *NeuroImage: Clinical*, vol. 7, pp. 622–630, 2015.
- [21] F. Chollet *et al.*, *Keras*, <https://keras.io>, 2015.
- [22] P. Clas, S. Groeschel, and M. Wilke, "A semi-automatic algorithm for determining the demyelination load in metachromatic leukodystrophy," *Academic Radiology*, vol. 19, no. 1, pp. 26–34, 2012, ISSN: 1076-6332.

- [23] D. Clevert, T. Unterthiner, and S. Hochreiter, “Fast and accurate deep network learning by exponential linear units (elus),” in *4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings*, 2016.
- [24] J. Conklin, F. L. Silver, D. J. Mikulis, and D. M. Mandell, “Are acute infarcts the cause of leukoaraiosis? brain mapping for 16 consecutive weeks,” *Annals of Neurology*, vol. 76, no. 6, pp. 899–904, Oct. 2014.
- [25] T. Crosby, “How to detect and handle outliers,” *Technometrics*, vol. 36, no. 3, pp. 315–316, 1994.
- [26] S. J. Daly and X. Feng, “Decontouring: Prevention and removal of false contour artifacts,” in *Human Vision and Electronic Imaging IX*, B. E. Rogowitz and T. N. Pappas, Eds., SPIE, Jun. 2004.
- [27] C. Dănișor and A. Pepe, “Comparative study of SAR interferometric phase filtering algorithms,” in *Advanced Topics in Optoelectronics, Microelectronics, and Nanotechnologies IX*, M. Vladescu, R. D. Tamas, and I. Cristea, Eds., International Society for Optics and Photonics, vol. 10977, SPIE, 2018, pp. 611–620.
- [28] C. Deledalle, L. Denis, G. Poggi, F. Tupin, and L. Verdoliva, “Exploiting patch similarity for sar image processing: The nonlocal paradigm,” *IEEE Signal Processing Magazine*, vol. 31, no. 4, pp. 69–78, Jul. 2014, ISSN: 1053-5888.
- [29] C. Deledalle, L. Denis, and F. Tupin, “NI-insar: Nonlocal interferogram estimation,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 4, pp. 1441–1452, Apr. 2011, ISSN: 0196-2892.
- [30] C. Deledalle, L. Denis, F. Tupin, A. Reigber, and M. Jäger, “NI-sar: A unified nonlocal framework for resolution-preserving (pol)(in)sar denoising,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 4, pp. 2021–2038, Apr. 2015, ISSN: 0196-2892.
- [31] L. Denis, F. Tupin, J. Darbon, and M. Sigelle, “Joint regularization of phase and amplitude of insar data: Application to 3-d reconstruction,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 11, pp. 3774–3785, Nov. 2009, ISSN: 0196-2892.
- [32] C. N. Devi, A. Chandrasekharan, V. Sundararaman, and Z. C. Alex, “Neonatal brain MRI segmentation: A review,” *Computers in Biology and Medicine*, vol. 64, pp. 163–178, Sep. 2015.
- [33] M. Donoser and H. Bischof, “3d segmentation by maximally stable volumes (msvs),” in *18th International Conference on Pattern Recognition (ICPR’06)*, vol. 1, 2006, pp. 63–66.
- [34] F. Dufaux, P. L. Callet, R. K. Mantiuk, and M. Mrak, *High Dynamic Range Video. From Acquisition to Display and Applications*. Elsevier, 2016.

- [35] Y. Fan, T. Jiang, and D. Evans, "Volumetric segmentation of brain images using parallel genetic algorithms," *IEEE Transactions on Medical Imaging*, vol. 21, no. 8, pp. 904–909, Aug. 2002.
- [36] A. Farzan, "Heuristically improved bayesian segmentation of brain MR images," *Science World Journal*, vol. 9, no. 3, 2014, ISSN: 1597-6343.
- [37] G. Ferraiuolo and G. Poggi, "A bayesian filtering technique for sar interferometric phase fields," *IEEE Transactions on Image Processing*, vol. 13, no. 10, pp. 1368–1378, Oct. 2004, ISSN: 1057-7149.
- [38] R. Floyd and L. Steinberg, "An adaptive algorithm for spatial gray scale," *SID International Symposium Digest of Technical Papers*, vol. 4, pp. 36–37, 1975.
- [39] H. J. Foley and M. W. Matlin, *Sensation and Perception (5th Edition)*. Pearson, 2009.
- [40] J. D. Foley, A. van Dam, S. K. Feiner, and J. F. Hughes, *Computer Graphics: Principles and Practice (2Nd Ed.)* Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 1990, ISBN: 0-201-12110-7.
- [41] S. Fu, X. Long, X. Yang, and Q. Yu, "Directionally adaptive filter for synthetic aperture radar interferometric phase images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 1, pp. 552–559, Jan. 2013, ISSN: 0196-2892.
- [42] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, Y. W. Teh and M. Titterington, Eds., ser. Proceedings of Machine Learning Research, vol. 9, Chia Laguna Resort, Sardinia, Italy: PMLR, May 2010, pp. 249–256.
- [43] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*, 1st. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 1989, ISBN: 0201157675.
- [44] R. M. Goldstein and C. L. Werner, "Radar interferogram filtering for geophysical applications," *Geophysical Research Letters*, vol. 25, no. 21, pp. 4035–4038, 1998.
- [45] R. C. Gonzalez and R. E. Woods, *Digital Image Processing (3rd Edition)*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 2006, ISBN: 013168728X.
- [46] H. Greenspan, A. Ruf, and J. Goldberger, "Constrained gaussian mixture model framework for automatic segmentation of mr brain images," *IEEE Transactions on Medical Imaging*, vol. 25, no. 9, pp. 1233–1245, Sep. 2006, ISSN: 0278-0062.

- [47] J. C. Griffis, J. B. Allendorfer, and J. P. Szaffarski, “Voxel-based gaussian naïve bayes classification of ischemic stroke lesions in individual t1-weighted MRI scans,” *Journal of Neuroscience Methods*, vol. 257, pp. 97–108, Jan. 2016.
- [48] T. Guo, E. G. Duerden, E. Adams, V. Chau, H. M. Branson, M. M. Chakravarty, K. J. Poskitt, A. Synnes, R. E. Grunau, and S. P. Miller, “Quantitative assessment of white matter injury in preterm neonates,” *Neurology*, vol. 88, no. 7, pp. 614–622, Jan. 2017.
- [49] B. de Haan, P. Clas, H. Juenger, M. Wilke, and H.-O. Karnath, “Fast semi-automated lesion demarcation in stroke,” *NeuroImage: Clinical*, vol. 9, pp. 69–74, 2015, ISSN: 2213-1582.
- [50] R. M. Haralick and L. G. Shapiro, *Computer and Robot Vision, Volume I*. Addison-Wesley, 1992.
- [51] P. Heckbert, “Color image quantization for frame buffer display,” *SIGGRAPH Comput. Graph.*, vol. 16, no. 3, pp. 297–307, Jul. 1982, ISSN: 0097-8930.
- [52] M. Herlihy and N. Shavit, *The Art of Multiprocessor Programming, Revised Reprint*. Morgan Kaufmann Publishers, 2012.
- [53] H. Hongxing, J. M. Bioucas-Dias, and V. Katkovnik, “Interferometric phase image estimation via sparse coding in the complex domain,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 5, pp. 2587–2602, May 2015, ISSN: 0196-2892.
- [54] H. Huang and Q. Wang, “A method of filtering and unwrapping sar interferometric phase based on nonlinear phase model,” *Progress In Electromagnetics Research*, vol. 144, pp. 67–78, 2014.
- [55] Q. Huang, H. Y. Kim, W.-J. Tsai, S. Y. Jeong, J. S. Choi, and J. Kuo, “Understanding and removal of false contour in hevc compressed images,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 2, pp. 378–391, Feb. 2018, ISSN: 1051-8215.
- [56] B. Iglewicz and D. Hoaglin, *How to Detect and Handle Outliers*, ser. ASQC basic references in quality control. ASQC Quality Press, 1993, ISBN: 9780873892476.
- [57] K. K. Iyer, J. A. Roberts, L. Hellström-Westas, S. Wikström, I. H. Pupp, D. Ley, M. Breakspear, and S. Vanhatalo, “Early detection of preterm intraventricular hemorrhage from clinical electroencephalography,” *Critical Care Medicine*, vol. 43, no. 10, pp. 2219–2227, Oct. 2015.
- [58] S. Jain, D. M. Sima, A. Ribbens, M. Cambron, A. Maertens, W. V. Hecke, J. D. Mey, F. Barkhof, M. D. Steenwijk, M. Daams, F. Maes, S. V. Huffel, H. Vrenken, and D. Smeets, “Automatic segmentation and volumetry of multiple sclerosis brain lesions from MR images,” *NeuroImage: Clinical*, vol. 8, pp. 367–375, 2015.

- [59] J. Jarvis, C. Judice, and W. Ninke, “A survey of techniques for the display of continuous tone pictures on bilevel displays,” *Computer Graphics and Image Processing*, vol. 5, no. 1, pp. 13–40, Mar. 1976.
- [60] Z. Ji, Q. Sun, Y. Xia, Q. Chen, D. Xia, and D. Feng, “Generalized rough fuzzy c-means algorithm for brain MR image segmentation,” *Computer Methods and Programs in Biomedicine*, vol. 108, no. 2, pp. 644–655, Nov. 2012.
- [61] M. Jiang, X. Ding, Z. Li, X. Tian, W. Zhu, C. Wang, and B. Xu, “The improvement for baran phase filter derived from unbiased insar coherence,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 7, pp. 3002–3010, Jul. 2014, ISSN: 1939-1404.
- [62] C. N. Judice, J. F. Jarvis, and W. H. Ninke, “Using ordered dither to display continuous tone pictures on an ac plasma panel,” *Proc. of SID*, vol. 15, no. 4, pp. 161–169, 1974.
- [63] I. E. Kaya, A. Ç. Pehlivanlı, E. G. Sekizkardeş, and T. Ibrikci, “PCA based clustering for brain tumor segmentation of t1w MRI images,” *Computer Methods and Programs in Biomedicine*, vol. 140, pp. 19–28, Mar. 2017.
- [64] K. Keraudren, M. Kuklisova-Murgasova, V. Kyriakopoulou, C. Malamateniou, M. Rutherford, B. Kainz, J. Hajnal, and D. Rueckert, “Automated fetal brain segmentation from 2d MRI slices for motion correction,” *NeuroImage*, vol. 101, pp. 633–643, Nov. 2014.
- [65] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015.
- [66] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds., Curran Associates, Inc., 2012, pp. 1097–1105.
- [67] P. Kuo, C. Tang, and S. Chien, “Content-adaptive inverse tone mapping,” in *2012 Visual Communications and Image Processing*, Nov. 2012, pp. 1–6.
- [68] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998, ISSN: 1558-2256.
- [69] J. W. Lee, B. R. Lim, R.-H. Park, J.-S. Kim, and W. Ahn, “Two-stage false contour detection using directional contrast and its application to adaptive false contour reduction,” *IEEE Transactions on Consumer Electronics*, vol. 52, no. 1, pp. 179–188, Feb. 2006, ISSN: 0098-3063.

- [70] J.-S. Lee, K. P. Papathanassiou, T. L. Ainsworth, M. R. Grunes, and A. Reigber, "A new technique for noise filtering of sar interferometric phase images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 36, no. 5, pp. 1456–1465, Sep. 1998, ISSN: 0196-2892.
- [71] K. V. Leemput, F. Maes, D. Vandermeulen, A. Colchester, and P. Suetens, "Automated segmentation of multiple sclerosis lesions by model outlier detection," *IEEE Transactions on Medical Imaging*, vol. 20, no. 8, pp. 677–688, Aug. 2001, ISSN: 0278-0062.
- [72] H. Li, A. Yezzi, and L. D. Cohen, "Computer vision for biomedical image applications: First international workshop, cvbia 2005, beijing, china, october 21, 2005. proceedings," in Y. Liu, T. Jiang, and C. Zhang, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, ch. Fast 3D Brain Segmentation Using Dual-Front Active Contours with Optional User-Interaction, pp. 335–345, ISBN: 978-3-540-32125-5.
- [73] X. Lin and D. Niu, "Experiments of interferometric phase filtering through weighted nuclear norm minimization," in *Proceedings of the 2nd International Conference on Big Data Technologies*, ser. ICBDT2019, Jinan, China: ACM, 2019, pp. 278–282, ISBN: 978-1-4503-7192-6.
- [74] H.-T. Liu, T. Sheu, and H.-H. Chang, "Automatic segmentation of brain MR images using an adaptive balloon snake model with fuzzy classification," *Medical & Biological Engineering & Computing*, vol. 51, no. 10, pp. 1091–1104, 2013, ISSN: 0140-0118.
- [75] C. Lopez-Martinez and X. Fabregas, "Modeling and reduction of sar interferometric phase noise in the wavelet domain," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 12, pp. 2553–2566, Dec. 2002, ISSN: 0196-2892.
- [76] R. Mantiuk, K. Myszkowski, and H.-P. Seidel, "A perceptual framework for contrast processing of high dynamic range images," *ACM Trans. Appl. Percept.*, vol. 3, no. 3, pp. 286–308, Jul. 2006, ISSN: 1544-3558.
- [77] S. T. M. Marba, J. P. S. Caldas, L. E. F. Vinagre, and M. A. Pessoto, "Incidence of periventricular/intraventricular hemorrhage in very low birth weight infants: A 15-year cohort study," pt, *Jornal de Pediatria*, vol. 87, pp. 505–511, Dec. 2011, ISSN: 0021-7557.
- [78] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide baseline stereo from maximally stable extremal regions," in *Proc. BMVC*, 2002, pp. 36.1–36.10, ISBN: 1-901725-19-7.
- [79] H. Matre, "A review of image quality assessment methods with application to computational photography," in *MIPPR 2015: Multispectral Image Acquisition, Processing, and Analysis*, Z. Cao, J. K. Udupa, and H. Matre, Eds., SPIE, Dec. 2015.

- [80] U. Maulik, “Medical image segmentation using genetic algorithms,” *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 2, pp. 166–173, Mar. 2009, ISSN: 1089-7771.
- [81] A. Mekhmoukh and K. Mokrani, “Improved fuzzy c-means based particle swarm optimization (PSO) initialization and outlier rejection with level set methods for MR brain image segmentation,” *Computer Methods and Programs in Biomedicine*, vol. 122, no. 2, pp. 266–281, Nov. 2015.
- [82] N. Merkle, W. Luo, S. Auer, R. Müller, and R. Urtasun, “Exploiting deep matching and sar data for the geo-localization accuracy improvement of optical satellite images,” *Remote Sensing*, vol. 9, no. 6, 2017, ISSN: 2072-4292.
- [83] A. Mestre-Quereda, J. M. Lopez-Sanchez, J. Selva, and P. J. Gonzalez, “An improved phase filter for differential sar interferometry based on an iterative method,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 8, pp. 4477–4491, Aug. 2018, ISSN: 1558-0644.
- [84] S. Miller, M. Nezamabadi, and S. Daly, “Perceptual signal coding for more efficient usage of bit codes,” in *The 2012 Annual Technical Conference Exhibition*, Oct. 2012, pp. 1–9.
- [85] S. P. Miller, D. M. Ferriero, C. Leonard, R. Piecuch, D. V. Glidden, J. C. Partridge, M. Perez, P. Mukherjee, D. B. Vigneron, and A. J. Barkovich, “Early brain injury in premature newborns detected with magnetic resonance imaging is associated with adverse early neurodevelopmental outcome,” *The Journal of Pediatrics*, vol. 147, no. 5, pp. 609–616, Nov. 2005.
- [86] J. C. Moreno, V. S. Prasath, H. Proença, and K. Palaniappan, “Fast and globally convex multiphase active contours for brain MRI segmentation,” *Computer Vision and Image Understanding*, vol. 125, pp. 237–250, Aug. 2014.
- [87] S. Mukherjee, N. K. Kottayil, X. Sun, and I. Cheng, “Cnn-based real-time parameter tuning for optimizing denoising filter performance,” in *Image Analysis and Recognition*, F. Karray, A. Campilho, and A. Yu, Eds., Cham: Springer International Publishing, 2019, pp. 112–125, ISBN: 978-3-030-27202-9.
- [88] S. Mukherjee, A. Zimmer, X. Sun, P. Ghuman, and I. Cheng, “Cnn-based insar coherence classification,” in *2018 IEEE SENSORS*, Oct. 2018, pp. 1–4.
- [89] A. N. Netravali and B. G. Haskell, *Digital Pictures*. Springer US, 1988.
- [90] T. Q. Nguyen, J. Kay, and J. Pasquale, “Fast source-based dithering for networked digital video,” in *High-Speed Networking and Multimedia Computing*, A. A. Rodriguez, M.-S. Chen, and J. Maitan, Eds., SPIE, Apr. 1994.

- [91] D. Nistér and H. Stewénus, “Linear time maximally stable extremal regions,” English, in *Computer Vision – ECCV 2008*, ser. Lecture Notes in Computer Science, D. Forsyth, P. Torr, and A. Zisserman, Eds., vol. 5303, Springer Berlin Heidelberg, 2008, pp. 183–196, ISBN: 978-3-540-88685-3.
- [92] C. Ojha, A. Fusco, and I. M. Pinto, “Interferometric sar phase denoising using proximity-based k-svd technique,” *Sensors*, vol. 19, no. 12, 2019, ISSN: 1424-8220.
- [93] A. Ortiz, J. Gorriz, J. Ramirez, and D. Salas-Gonzalez, “Improving MR brain image segmentation using self-organising maps and entropy-gradient clustering,” *Information Sciences*, vol. 262, pp. 117–136, 2014, ISSN: 0020-0255.
- [94] N. Ostu, “A threshold selection method from gray-level histograms,” *Systems, Man and Cybernetics, IEEE Transactions on*, vol. 9, no. 1, pp. 62–66, Jan. 1979, ISSN: 0018-9472.
- [95] X. Ou, C. Glasier, R. Ramakrishnaiah, S. Mulkey, Z. Ding, T. Angtuaco, A. Andres, and J. Kaiser, “Impaired white matter development in extremely low-birth-weight infants with previous brain hemorrhage,” *American Journal of Neuroradiology*, vol. 35, no. 10, pp. 1983–1989, May 2014.
- [96] A. Pepe, Y. Yang, M. Manzo, and R. Lanari, “Improved emcf-sbas processing chain based on advanced techniques for the noise-filtering and selection of small baseline multi-look dinsar interferograms,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 8, pp. 4394–4417, Aug. 2015, ISSN: 1558-0644.
- [97] P. Perona and J. Malik, “Scale-space and edge detection using anisotropic diffusion,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 7, pp. 629–639, Jul. 1990, ISSN: 0162-8828.
- [98] X. Qian, J. Wang, S. Guo, and Q. Li, “An active contour model for medical image segmentation with application to brain CT image,” *Med. Phys.*, vol. 40, no. 2, p. 021 911, Feb. 2013.
- [99] X. Qing, J. Guowang, Z. Caiying, W. Zhengde, H. Yu, and Y. Peizhang, “The filtering and phase unwrapping of interferogram,” in *Proceedings of International Society for Photogrammetry and Remote Sensing (ISPRS), Volume XXXV, Technical Commission V1/Working Group 4*, Istanbul, Turkey, 2004.
- [100] E. Reinhard, W. Heidrich, P. Debevec, S. Pattanaik, G. Ward, and K. Myszkowski, *High Dynamic Range Imaging (2nd Edition)*. Elsevier, 2010.

- [101] A. G. Rempel, M. Trentacoste, H. Seetzen, H. D. Young, W. Heidrich, L. Whitehead, and G. Ward, “Ldr2hdr: On-the-fly reverse tone mapping of legacy video and photographs,” in *In SIGGRAPH '07: ACM SIGGRAPH 2007 papers*, ACM Press, 2007.
- [102] T. Reza, A. Zimmer, J. M. D. Blasco, P. Ghuman, T. K. Aasawat, and M. Ripeanu, “Accelerating persistent scatterer pixel selection for insar processing,” *IEEE Transactions on Parallel and Distributed Systems*, vol. 29, no. 1, pp. 16–30, 2018.
- [103] L. Roberts, “Picture coding using pseudo-random noise,” *IRE Transactions on Information Theory*, vol. 8, no. 2, pp. 145–154, Feb. 1962, ISSN: 0096-1000.
- [104] A. Rosenfeld and J. L. Pfaltz, “Sequential operations in digital picture processing,” *J. ACM*, vol. 13, no. 4, pp. 471–494, Oct. 1966, ISSN: 0004-5411.
- [105] E. Roura, A. Oliver, M. Cabezas, J. C. Vilanova, À. Rovira, L. Ramió-Torrentà, and X. Lladó, “MARGA: Multispectral adaptive region growing algorithm for brain extraction on axial MRI,” *Computer Methods and Programs in Biomedicine*, vol. 113, no. 2, pp. 655–673, Feb. 2014.
- [106] S. Roy, D. Bhattacharyya, S. K. Bandyopadhyay, and T.-H. Kim, “An effective method for computerized prediction and segmentation of multiple sclerosis lesions in brain MRI,” *Computer Methods and Programs in Biomedicine*, vol. 140, pp. 307–320, Mar. 2017.
- [107] “Rp 431-2:2011 - smpte recommended practice - d-cinema quality — reference projector and environment,” *RP 431-2:2011*, pp. 1–14, Apr. 2011.
- [108] J. Sachdeva, V. Kumar, I. Gupta, N. Khandelwal, and C. K. Ahuja, “A novel content-based active contour model for brain tumor segmentation,” *Magnetic Resonance Imaging*, vol. 30, no. 5, pp. 694–715, Jun. 2012.
- [109] S. D. Salman, Q. A. Habash, and Z. T. Ahmed, “3d brain segmentation using active contour with multi labeling method,” in *Engineering Sciences (FNCEs), 2012 First National Conference for*, Nov. 2012, pp. 1–4.
- [110] G. L. Y. San, M. L. Lee, and W. Hsu, “Constrained-mser detection of retinal pathology,” in *Pattern Recognition (ICPR), 2012 21st International Conference on*, Nov. 2012, pp. 2059–2062.
- [111] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. Chen, “Mobilenetv2: Inverted residuals and linear bottlenecks,” in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Jun. 2018, pp. 4510–4520.

- [112] S. J. Savio, L. C. Harrison, T. Luukkaala, T. Heinonen, P. Dastidar, S. Soimakallio, and H. J. Eskola, "Effect of slice thickness on brain magnetic resonance image texture analysis," *BioMedical Engineering OnLine*, vol. 9, no. 1, pp. 1–14, 2010, ISSN: 1475-925X.
- [113] P. Schmidt, C. Gaser, M. Arsic, D. Buck, A. Förschler, A. Berthele, M. Hoshi, R. Ilg, V. J. Schmid, C. Zimmer, B. Hemmer, and M. Mühlau, "An automated tool for detection of FLAIR-hyperintense white-matter lesions in multiple sclerosis," *NeuroImage*, vol. 59, no. 4, pp. 3774–3783, Feb. 2012.
- [114] M. S. Seymour and I. G. Cumming, "Maximum likelihood estimation for sar interferometry," in *Proceedings of IGARSS '94 - 1994 IEEE International Geoscience and Remote Sensing Symposium*, vol. 4, Aug. 1994, 2272–2275 vol.4.
- [115] N. Shiee, P.-L. Bazin, A. Ozturk, D. S. Reich, P. A. Calabresi, and D. L. Pham, "A topology-preserving approach to the segmentation of brain images with multiple sclerosis lesions," *NeuroImage*, vol. 49, no. 2, pp. 1524–1535, 2010, ISSN: 1053-8119.
- [116] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, no. 1, Jul. 2019.
- [117] F. Sica, D. Cozzolino, X. X. Zhu, L. Verdoliva, and G. Poggi, "Insar-bm3d: A nonlocal filter for sar interferometric phase restoration," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 6, pp. 3456–3467, Jun. 2018, ISSN: 0196-2892.
- [118] N. P. Simon, *Periventricular/intraventricular hemorrhage (pvh/ivh) in the premature infant*, <http://www.pediatrics.emory.edu/divisions/neonatology/dpc/pvhivh.html>, [Online; accessed 02-April-2018].
- [119] P. Soille, *Morphological Image Analysis*. Springer Science + Business Media, 2004.
- [120] R. Song, H. Guo, G. Liu, Z. Perski, and J. Fan, "Improved goldstein sar interferogram filter based on empirical mode decomposition," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 2, pp. 399–403, Feb. 2014, ISSN: 1545-598X.
- [121] Y. Song, W. Cai, H. Huang, Y. Wang, D. Feng, and M. Chen, "Region-based progressive localization of cell nuclei in microscopic images with data adaptive modeling," *BMC Bioinformatics*, vol. 14, no. 1, p. 173, 2013.
- [122] M. Sonka, S. K. Tadikonda, and S. M. Collins, "Knowledge-based interpretation of mr brain images," *IEEE Transactions on Medical Imaging*, vol. 15, no. 4, pp. 443–452, Aug. 1996, ISSN: 0278-0062.

- [123] *Spm12 - statistical parametric mapping*, <http://www.fil.ion.ucl.ac.uk/spm/software/spm12/>, Accessed: 2016-06-09.
- [124] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: A simple way to prevent neural networks from overfitting,” *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, Jan. 2014, ISSN: 1532-4435.
- [125] J. Stoffel and J. Moreland, “A survey of electronic techniques for pictorial image reproduction,” *IEEE Transactions on Communications*, vol. 29, no. 12, pp. 1898–1925, Dec. 1981, ISSN: 0090-6778.
- [126] G.-M. Su, T. Chen, P. Yin, and S. Qu, *Guided post-prediction filtering in layered vdr coding (patent us8897581b2)*, Nov. 2014.
- [127] G.-M. Su, S. Qu, and S. Daly, *Adaptive false contouring prevention in layered coding of images with extended dynamic range (patent us8873877b2)*, Oct. 2014.
- [128] Z. Suo, J. Zhang, M. Li, Q. Zhang, and C. Fang, “Improved insar phase noise filter in frequency domain,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 2, pp. 1185–1195, Feb. 2016, ISSN: 1558-0644.
- [129] X. Tang, L. Zhang, and X. Ding, “Sar image despeckling with a multilayer perceptron neural network,” *International Journal of Digital Earth*, vol. 12, no. 3, pp. 354–374, 2019.
- [130] E. Trouve, J.-M. Nicolas, and H. Maitre, “Improving phase unwrapping techniques by the use of local frequency estimates,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 36, no. 6, pp. 1963–1972, Nov. 1998, ISSN: 0196-2892.
- [131] R. Ulichney, *Digital Halftoning*. MIT Press, 1987.
- [132] —, “Video rendering,” *Digital Tech. J.*, vol. 5, no. 2, pp. 9–18, Mar. 1993.
- [133] G. Vasile, E. Trouve, J.-S. Lee, and V. Buzuloiu, “Intensity-driven adaptive-neighborhood technique for polarimetric and interferometric sar parameters estimation,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 6, pp. 1609–1621, Jun. 2006, ISSN: 0196-2892.
- [134] P. Wang, H. Zhang, and V. M. Patel, “Generative adversarial network-based restoration of speckled sar images,” in *2017 IEEE 7th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, Dec. 2017, pp. 1–5.
- [135] P. Wang, H. Zhang, and V. M. Patel, “Sar image despeckling using a convolutional neural network,” *IEEE Signal Processing Letters*, vol. 24, no. 12, pp. 1763–1767, Dec. 2017, ISSN: 1070-9908.

- [136] Y. Wang, H. Huang, Z. Dong, and M. Wu, "Modified patch-based locally optimal wiener method for interferometric sar phase filtering," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 114, pp. 10–23, 2016, ISSN: 0924-2716.
- [137] N. Wu, D.-Z. Feng, and J. Li, "A locally adaptive filter of interferometric phase images," *IEEE Geoscience and Remote Sensing Letters*, vol. 3, no. 1, pp. 73–77, Jan. 2006, ISSN: 1545-598X.
- [138] Y. ong Gao, S. Zhang, K. Zhang, and S. Li, "Frequency domain filtering SAR interferometric phase noise using the amended matrix pencil model," *Computer Modeling in Engineering & Sciences*, vol. 119, no. 2, pp. 349–363, 2019.
- [139] X. Zha, R. Fu, Z. Dai, and B. Liu, "Noise reduction in interferograms using the wavelet packet transform and wiener filtering," *IEEE Geoscience and Remote Sensing Letters*, vol. 5, no. 3, pp. 404–408, Jul. 2008, ISSN: 1545-598X.
- [140] Q. Zhang, Q. Yuan, J. Li, Z. Yang, and X. Ma, "Learning a dilated residual network for sar image despeckling," *Remote Sensing*, vol. 10, no. 2, 2018, ISSN: 2072-4292.
- [141] H. Zhu, J. Sheng, F. Zhang, J. Zhou, and J. Wang, "Improved maximally stable extremal regions based method for the segmentation of ultrasonic liver images," *Multimed Tools Appl*, Jul. 2015.
- [142] A. Zimmer and P. Ghuman, "Cuda optimization of non-local means extended to wrapped gaussian distributions for interferometric phase denoising," *Procedia Computer Science*, vol. 80, pp. 166–177, 2016, International Conference on Computational Science 2016, ICCS 2016, 6-8 June 2016, San Diego, California, USA, ISSN: 1877-0509.