

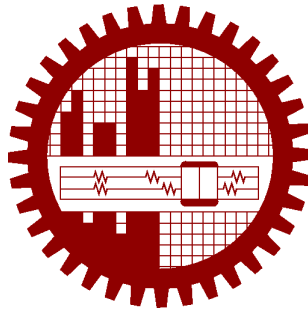
# **AUTOMATIC CELL NUCLEI SEGMENTATION IN HISTOPATHOLOGY IMAGES USING BOUNDARY PRESERVING GUIDED ATTENTION BASED DEEP NEURAL NETWORK**

by

Tamjid Imtiaz

Submitted to

Department of Electrical and Electronic Engineering  
in fulfillment of the requirements for the degree of  
Master of Science in Electrical and Electronic Engineering



Department of Electrical and Electronic Engineering  
Bangladesh University of Engineering and Technology

Dhaka, Bangladesh

August, 2022

This thesis titled, “Automatic Cell Nuclei Segmentation in Histopathology Images Using Boundary Preserving Guided Attention Based Deep Neural Network”, submitted by Tamjid Imtiaz, Roll No.: 0419062285, Session: April 2019, to the Department of EEE, Bangladesh University of Engineering and Technology, has been accepted as satisfactory in partial fulfillment of the requirements for the degree of Master of Science in Electrical & Electronic Engineering on 7<sup>th</sup> August, 2022.

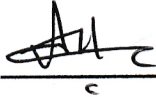
## BOARD OF EXAMINERS



---

Dr. Shaikh Anowarul Fattah  
Professor  
Department of Electrical and Electronic Engineering,  
Bangladesh University of Engineering and Technology, Dhaka

Chairman  
(Supervisor)



---

Dr. Md. Aynal Haque  
Head and Professor  
Department of Electrical and Electronic Engineering,  
Bangladesh University of Engineering and Technology, Dhaka


Member  
(Ex-officio)



---

Dr. Mohammad Ariful Haque  
Professor  
Department of Electrical and Electronic Engineering,  
Bangladesh University of Engineering and Technology, Dhaka

Member



---

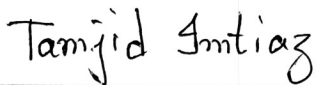
Dr. Mohiuddin Ahmad  
Professor  
Department of Electrical and Electronic Engineering,  
Khulna University of Engineering & Technology, Khulna

Member  
(External)

## Candidate's Declaration

This is to certify that the work presented in this thesis entitled, "AUTOMATIC CELL NUCLEI SEGMENTATION IN HISTOPATHOLOGY IMAGES USING BOUNDARY PRESERVING GUIDED ATTENTION BASED DEEP NEURAL NETWORK", is the outcome of the research carried out by Tamjid Imtiaz under the supervision of Dr. Shaikh Anowarul Fattah, Professor, Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology (BUET), Dhaka-1000, Bangladesh. It is also declared that neither this thesis nor any part thereof has been submitted anywhere else for the award of any degree, diploma, or other qualifications.

Signature of the Candidate



---

Tamjid Imtiaz

0419062285

*Dedicated to my parents*



# Contents

<b>Certification</b>	<b>ii</b>
<b>Candidate’s Declaration</b>	<b>iii</b>
<b>Dedication</b>	<b>iv</b>
<b>List of Figures</b>	<b>vii</b>
<b>List of Tables</b>	<b>x</b>
<b>Acknowledgement</b>	<b>xii</b>
<b>Abstract</b>	<b>xiii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Nuclei Segmentation: Challenges and Significance . . . . .	1
1.2 Live-cell Imaging Microscopy Techniques . . . . .	4
1.2.1 Fluorescent Microscopy . . . . .	5
1.2.2 Widefield Microscopy . . . . .	5
1.3 Literature Survey . . . . .	6
1.4 Motivation of This Research . . . . .	9
1.5 Objective of this thesis . . . . .	9
1.6 Oraganization of the Thesis . . . . .	10
<b>2 Proposed Cell Nuclei Segmentation Framework Using Boundary Aware     Feature Map Guided Attention Network</b>	<b>12</b>
2.1 Proposed Method . . . . .	13
2.1.1 Preprocessing . . . . .	13
2.1.2 Proposed Boundary Aware Feature Map Guided Attention Net- work . . . . .	15
2.2 Results and Discussions . . . . .	22
2.2.1 Dataset Description . . . . .	22
2.2.2 Experimental Setup . . . . .	22
2.2.3 Analysis of the Segmentation Performance . . . . .	24
2.3 Conclusion . . . . .	28

<b>3</b>	<b>Proposed Cell Nuclei Segmentation Scheme Based on Contourlet Driven Attention Mechanism</b>	<b>29</b>
3.1	Proposed Method . . . . .	30
3.1.1	Preprocessing . . . . .	30
3.1.2	Proposed Contourlet Driven Attention Network . . . . .	32
3.2	Results and Discussion . . . . .	37
3.2.1	Datasets . . . . .	38
3.2.2	Training and Implementation Details . . . . .	38
3.2.3	Ablation Study of the proposed nuclei segmentation Network . . . . .	39
3.2.4	Comparison to the State-of-the-art Methods . . . . .	40
3.2.5	Qualitative evaluation . . . . .	41
3.3	Conclusion . . . . .	44
<b>4</b>	<b>Proposed Cell Nuclei Segmentation Scheme Based on Boundary Aware Wavelet Guided Network</b>	<b>45</b>
4.1	Proposed Cell Nuclei Segmentation Scheme . . . . .	45
4.1.1	Preprocessing . . . . .	46
4.1.2	Proposed Boundary Aware Wavelet Guided Network . . . . .	46
4.1.3	Hybrid Loss Function . . . . .	52
4.2	Results and Discussions . . . . .	53
4.2.1	Dataset Description . . . . .	53
4.2.2	Experimental Setup . . . . .	54
4.2.3	Analysis of the Segmentation Performance . . . . .	56
4.3	Conclusion . . . . .	61
<b>5</b>	<b>Conclusions</b>	<b>62</b>
5.1	Contribution of This Thesis . . . . .	62
5.2	Future Prospects of Our Work . . . . .	63
	<b>References</b>	<b>64</b>

# List of Figures

1.1	Structure of the cell. . . . .	2
1.2	Different types of nuclei. (a) Lymphocyte nuclei (b) Epithelial Nuclei (c) Epithelial Nuclei (Cancer). (d) Epithelial Nuclei (Mitosis). . . . .	3
1.3	Nuclear structure in normal and cancer cells. . . . .	3
1.4	Microscopy images extracted from Fluorescence (First row) and Bright-field (second row) microscopy techniques. . . . .	5
2.1	Histopathology mage sample acquired from (a) Brightfield (b) Fluorescence microscopy techniques. . . . .	13
2.2	Visualization of microscopy images at different stages of preprocessing step. . . . .	14
2.3	Graphical overview of the proposed Boundary Aware Signal Generation Network. Here, the edge regions of the nucleus present in a histopathology images are utilized as the mask. . . . .	16
2.4	Morphological dilation based edge extraction. . . . .	17
2.5	Visualization of feature maps at different stages of encoder and decoder. (a-b) feature maps extracted from encoder stage, (c-d) feature maps extracted from decoder stage, (e) a closer look of the feature maps extracted from decoder stage. . . . .	18
2.6	Graphical overview of the proposed boundary aware feature map guided attention network. . . . .	19
2.7	A visualization of the attention mechanism used in the segmentation network to combine the boundary aware and spatial domain information. . . . .	21
2.8	A visualization of the samples collected from (a) Data Science Bowl 2018, (b) MoNuSeg and (c) TNBC datasets. . . . .	23
2.9	Visualization of the nuclei segmentation performance of some of the state-of-the-art networks in three datasets. . . . .	27
3.1	Visualization of microscopy images at different stages of preprocessing step. . . . .	30
3.2	Graphical overview of the proposed contourlet driven attention network (ConDANet). . . . .	31
3.3	Visualization of contourlet driven attention mechanism. . . . .	33

3.4	Contourlet filter bank. . . . .	34
3.5	Directional filter bank. Frequency partitioning where level (a) $l = 2$ and there are $2^2 = 4$ real wedge-shaped frequency bands, and (b) $l = 3$ and there are $2^3 = 8$ real wedge-shaped frequency bands. . . . .	35
3.6	Example of the contourlet transform on the histopathology image. For clear visualization, each image is only decomposed into one pyramidal level, which are then decomposed into four directional subbands. Small coefficients are shown in black while large coefficients are shown in white. 36	36
3.7	Laplacian pyramid. (a) One level of decomposition. The outputs are a coarse approximation $a[n]$ and a difference $b[n]$ between the original signal and the prediction. (b) The new reconstruction scheme for the Laplacian pyramid . . . . .	36
3.8	Effect of Contourlet driven attention module (CDAM). (a) Original image (b) Ground truth (b) feature map without CDAM (d) feature map with CDAM . . . . .	36
3.9	Visualization of the nuclei segmentation performance of some of the state-of-the-art networks in three datasets. The inputs of the first two rows correspond to DSB-2018 dataset, the next two rows correspond to MoNuSeg dataset, and the final two rows correspond to TNBC dataset. Here, the 'green' color is used to represent false positive cases and the 'red' color is used to represent false negative cases. . . . .	43
4.1	A simplified representation of the proposed architecture. Here, $E_i$ , $D_i$ , and $G_i$ correspond to the encoder, decoder and guide unit at $i^{th}$ stage of the network, respectively. . . . .	47
4.2	Graphical overview of the proposed wavelet attention U-net. Here, three separate loss functions are utilized to optimize the overall network and three individual segmented results are considered to get the final segmentation result. Boundary Aware Information Fusion' unit (yellow block) is shown in detail in Fig.4.3 . . . . .	48
4.3	(a) Proposed aggregation unit (AU): It consists of two major parts-wavelet guided attention unit (WGAU) and boundary aware unit (BAU). (b) Boundary Aware information fusion: It fuses multi-level boundary information $S_i$ ( $i= 1, 2, 3$ ) generated from BAU. . . . .	52
4.4	A visualization of the attention mechanism used in WGAU to combine the wavelet and spatial domain information. . . . .	52
4.5	A visualization of the (a) original segmented output and (b) the corresponding boundary constraint map. . . . .	53

4.6 Visualization of the nuclei segmentation performance of some of the state-of-the-art networks in three datasets. The inputs of the first two rows correspond to DSB-2018 dataset, the next two rows correspond to MoNuSeg dataset, and the final two rows correspond to TNBC dataset. . 58

# List of Tables

2.1	Ablation study of the proposed Network on three datasets. . . . .	25
2.2	Performance comparison of different architectures in the DSB-2018 dataset (8:1:1 split setup) . . . . .	25
2.3	Performance comparison of different architectures in the MoNuSeg dataset (8:1:1 split setup) . . . . .	26
2.4	Performance comparison of different architectures in the TNBC dataset (8:1:1 split setup) . . . . .	26
2.5	Computational Efficiency Analysis of Numerous Architectures along with the Performance of the proposed method. . . . .	28
3.1	Ablation study of the proposed network on three datasets. . . . .	41
3.2	Performance comparison of different architectures in the DSB-2018 dataset (8:1:1 split setup). . . . .	41
3.3	Performance comparison of different architectures in the MoNuSeg dataset (8:1:1 split setup). . . . .	41
3.4	Performance comparison of different architectures in the TNBC dataset (8:1:1 split setup). . . . .	42
3.5	Computational Efficiency Analysis of Numerous Architectures along with the Performance of the proposed method. . . . .	44
4.1	Ablation study of the proposed network on three datasets. . . . .	54
4.2	Performance comparison of different architectures in the TNBC dataset (8:1:1 split setup). . . . .	55
4.3	Performance comparison of different architectures in the MoNuSeg dataset (8:1:1 split setup). . . . .	56
4.4	Performance comparison of different architectures in the DSB-2018 dataset . . . . .	56
4.5	Performance comparison of different architectures in the DSB-2018 dataset. . . . .	56
4.6	Computational Efficiency Analysis of Numerous Architectures along with the Performance of the proposed method. . . . .	59
4.7	Performance of the proposed method on DSB dataset (10 fold cross validation) . . . . .	59

4.8	Comparison of the three proposed architectures on three datasets (DSB-2018, MoNuSeg, and TNBC) . . . . .	60
-----	--	----

# Acknowledgement

First of all, I would like to thank my loving parents, without whose prayers, constant support and sacrifice, I could never reach this stage of my life.

After them, I express my sincere gratitude to my supervisor Professor Shaikh Anowarul Fattah for his kind motivation, guidance and support during my master's research. I have been doing my research under him since my undergraduate level. For me, he is such a person without whom I have never found the interest in doing scientific research. He has been a guide for me by teaching me the way of doing research and the art of critical thinking. From him I learned that perseverance and constant hard work are the most important elements for succeeding in any challenges. I want to thank him for providing me his valuable time for discussion which significantly contributes to the advancement of this research and improves the writing of this dissertation.

Moreover, I would also like to thank the rest of the members of my thesis committee: Professor Dr. Md. Aynal Haque, Professor Dr. Mohammad Ariful Haque, and Professor Dr. Mohiuddin Ahmad, for their insightful comments. I want to thank the Head of the Department of Electrical and Electronic Engineering for allowing me to use the lab facilities. Last but not the least, I would also like to thank my friend, A F M Saif and Shahriar Rifat, for their consistent inspiration and for believing in me. I enjoyed the many insightful discussions I had with Saif and Rifat.



# Abstract

Precise cell nucleus segmentation is very critical in many biologically related analyses and disease diagnoses. Two of the major challenges in this task are the precise delineation of the small-shaped nucleus and the characterization of the edge region. In this thesis, deep neural network based nucleus segmentation methods are proposed where a unique idea of generating various types of boundary aware guiding signal is introduced to guide the spatial information of the segmentation architectures through attention mechanism.

In the first method, the attention module of the segmentation network is guided by a separate boundary extractor shallow encoder-decoder network which minimizes a separate objective function from a synthetically generated nucleus edge mask. The edge aware information found from different decoder stages of this shallow network acts as the guiding signal. Although the network offers comparatively better segmentation performance with respect to the other network, the complexity of the network increases because of the utilization of separate shallow network.

Considering the inspiring characteristics of various transformation techniques, a contourlet driven attention network, namely ConDANet, is developed which utilizes contourlet transformed signal as the guiding signal. The contourlet transform based controlling signal generation scheme exploits the advantage of the multi-scale time frequency localization and provides a high degree of directionality. Additionally, the wavelet pooling strategy is incorporated to the network which preserves the textural content of the nucleus.

Furthermore, a boundary aware wavelet guided network (BAWGNet) is proposed which utilizes the wavelet transform based guiding signal generation along with three separate loss function for optimization purpose. A boundary aware unit is designed that captures the nuclei's boundary information by employing a unique boundary aware loss function, ensuring accurate prediction of the nuclei pixels in the edge region.

The proposed method is employed for analyzing three publicly available histopathology datasets to manifest its effectiveness. Using the proposed framework, significant performance improvement is found over the other state-of-the-art techniques while evaluating on these datasets.

# Chapter 1

## Introduction

In this chapter, the biological background required to understand this work is presented. Next, the biological significance of the cell nuclei segmentation and the necessity of an automated computer-aided technique to address this issue is delineated. Later on, the past and present scenarios of state-of-the-art research in cell nuclei segmentation from histopathology images are described. Following that, the scope of potential future work in this area and the objective of this thesis are highlighted. Finally, the organization of the thesis is outlined at the end of this chapter.

### 1.1 Nuclei Segmentation: Challenges and Significance

The cell is the most fundamental and basic unit of life [1, 2]. A typical human body contains 30 trillion cells, each with its own nucleus containing Deoxyribonucleic acid (DNA) that programs the cell. Cells, like humans, can grow, reproduce, and interact with their surroundings by processing information and responding to environmental stimuli. A plasma membrane, cytoplasm, organelles, and a well-defined nucleus are all features of a eukaryotic cell (Fig.1.1) [3]. The nucleus is an essential part of the cell. It contains genetic material, the chromosomes (DNA + proteins), which contain information about how the cell works. Every cell has a copy of the genetic information, but depending on the function of the cell, some genes are turned off and others are turned on. This explains the variety of cells that make up the human body [2].

Cell nuclei segmentation is important from different biological point of view. For instance, it provides valuable information about the DNA content [4], chromatin condensation [5] and nuclei morphology [6]. This information can be used, for example, to study the cell cycle [4] or to study mutations in proteins associated with cancer [7]. Cell overlapping, image noise and non-uniform acquisition and preparation parameters make the nuclei segmentation procedure a challenging task [5, 6].

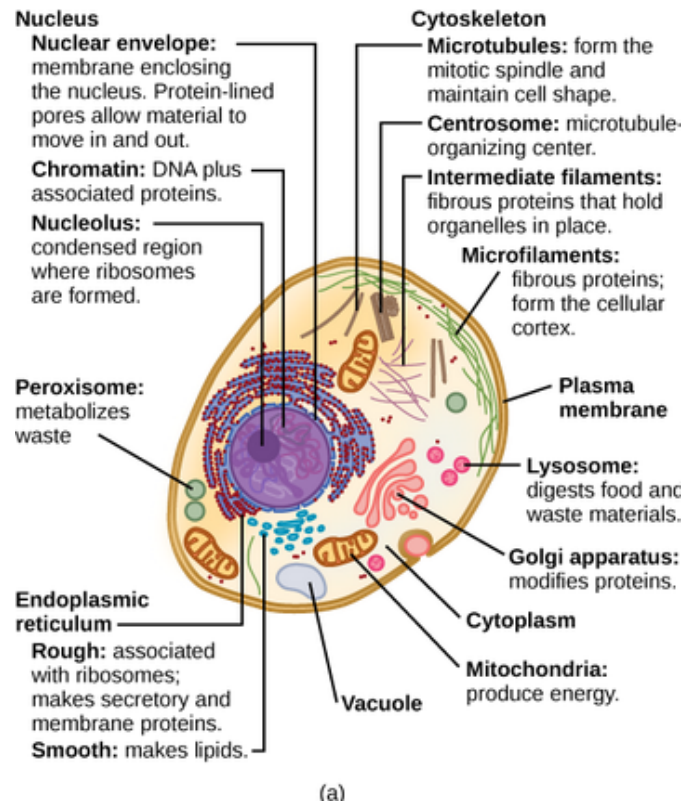


Figure 1.1: Structure of the cell.

Moreover, nuclei segmentation is important for several applications, for instance, it can be used to address the important problem of cell cycle staging. Cell cycle is a fundamental mechanism of living organisms. The main goal is to ensure that when a cell divides it passes correctly the genetic information to the next generation. Additionally, it plays an important role in many diseases. For instance, dysregulation of the cell cycle is at the origin of many diseases, such as cancer, ischemia, neurodegenerative disorders and infection. Therefore, by studying and controlling the cell cycle one can understand the mechanisms of various diseases, and consequently find ways to tackle them. Furthermore, cell cycle control and manipulation can have numerous applications in regenerative medicine [8]. A visualization of the change in nuclear structure is depicted in Fig 1.3. It is found that nuclei can become irregular and begin to fold in the tumor cell, and coarse heterochromatin aggregates are frequently observed. Nucleoli can be enlarged and PML bodies can mislocalize in microspeckles [9].

Among the different types of nuclei, two types are usually the object of particular interest: lymphocyte and epithelial nuclei. Nuclei may look very different according to a number of factors such as nuclei type, malignancy of the disease, and nuclei life cycle as shown in Fig 1.2. Lymphocyte is a type of white blood cell that has a major role in the immune system. Lymphocyte nuclei are inflammatory nuclei having regular shape and smaller size than epithelial nuclei. Nonpathological epithelial nuclei have nearly uniform chromatin distribution with smooth boundary. In high-grade cancer tis-

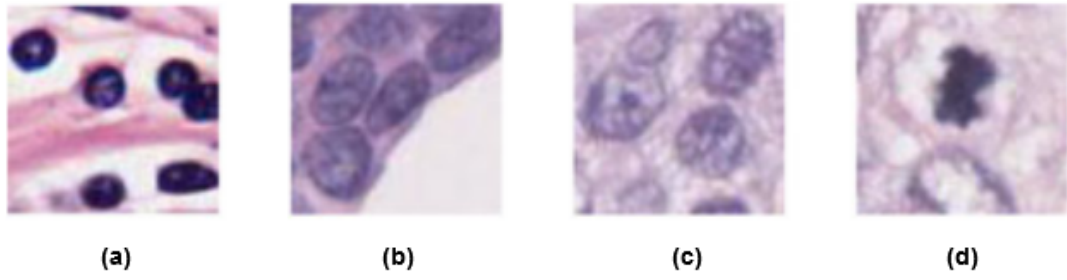


Figure 1.2: Different types of nuclei. (a) Lymphocyte nuclei (b) Epithelial Nuclei (c) Epithelial Nuclei (Cancer). (d) Epithelial Nuclei (Mitosis).

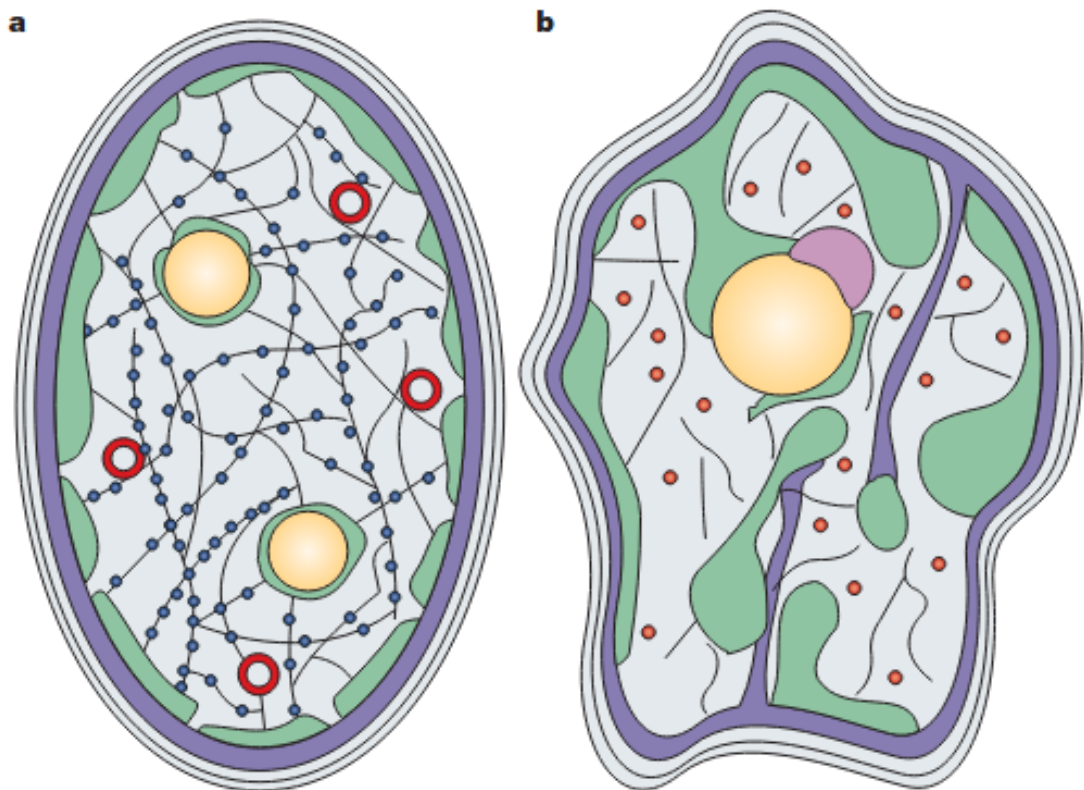


Figure 1.3: Nuclear structure in normal and cancer cells.

sue, epithelial nuclei are larger in size, may have heterogeneous chromatin distribution, irregular boundaries, referred to as nuclear pleomorphism, and clearly visible nucleoli as compared to normal epithelial nuclei. The variation in nuclei shape, size, and texture during nuclei life cycle, mitotic nuclei, is another factor of complexity.

Automated nuclei segmentation is now a well-studied topic for which a large number of methods have been described in the literature and new methodologies continue to be investigated. Segmentation of nuclei in routinely stained histopathological images pose a difficult computer vision problem due to high variability in images caused by a number of factors including differences in slide preparation (dyes concentration, evenness of the cut, presence of foreign artifacts or damage to the tissue sample, etc.) and image acquisition (artifacts introduced by the compression of the image, presence of digital noise, specific features of the slide scanner, etc.). Furthermore, nuclei are often organized in overlapping clusters and have heterogeneous aspects. All these problems make the nuclei segmentation a challenging problem. A successful image processing approach can overcome these issues in a robust way in order to maintain a high level in the quality and accuracy in all situations. To avoid tedious manual analysis of histopathology images, automatic nuclei segmentation schemes are getting importance. An automatic computer aided nucleus segmentation scheme can segment thousands of nucleus within seconds and assists the physicians for quick assessment of cellular activities.

## **1.2 Live-cell Imaging Microscopy Techniques**

When selecting an optical microscopy system for live-cell imaging, three factors should be considered: detector sensitivity (signal-to-noise), specimen viability, and image acquisition speed [10]. The system used should maximize the use of light while utilizing the fewest optical elements in the light path. To improve signal-to-noise ratio, the combination of filters used for live-cell imaging needs to be closely match the spectral profiles of the fluorophores used in experiments. Because of advancements in detector technology, illumination levels can now be reduced. The camera's sensitivity is also an important consideration. This can be accomplished with an intensified camera or a sensitive back-illuminated charged couple device (CCD) camera. Switching between filters or output from a monochromator will reduce acquisition time, especially when imaging multiple fluorophores simultaneously or performing radiometric analysis on a single probe.

There is currently no all-purpose live-cell imaging system that can be used for all possible investigations. As a result, the imaging system can be selected by determining the best parameters while minimizing cell damage or death. The majority of cellular processes take place in three dimensions over time.

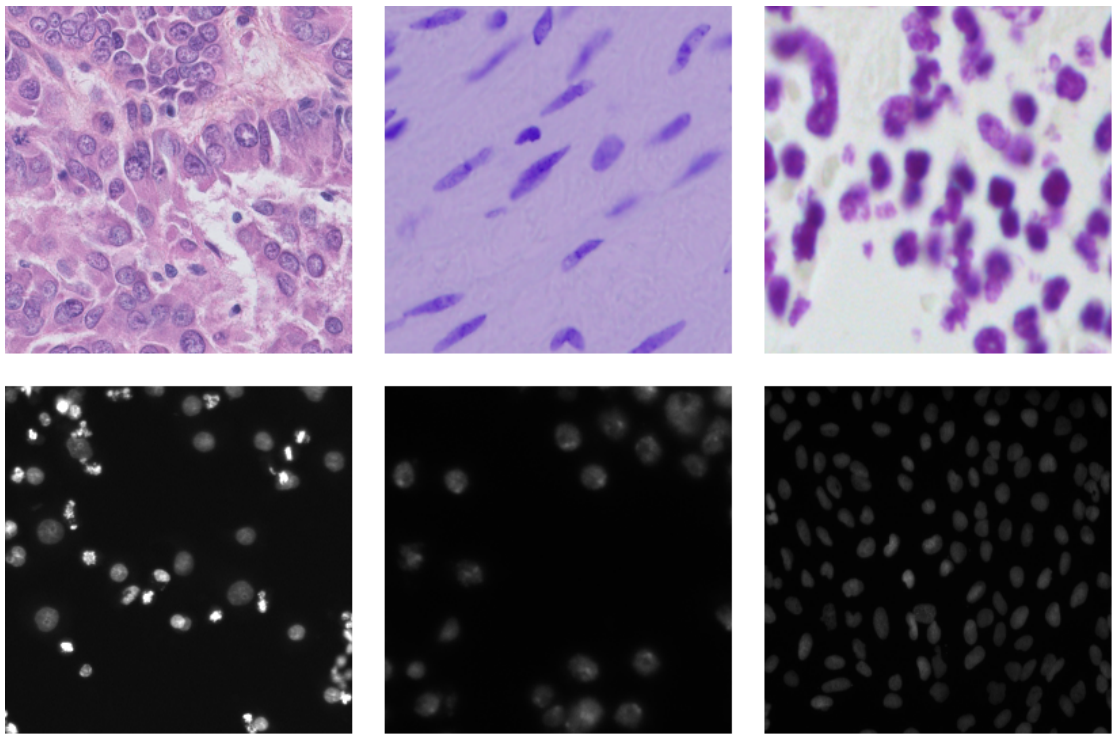


Figure 1.4: Microscopy images extracted from Fluorescence (First row) and Brightfield (second row) microscopy techniques.

Live-cell imaging is currently mainly performed by fluorescence microscopy [11]. Some of the more common systems used for imaging live cells are discussed below; the more widely used systems are discussed first and then the less commonly used systems are described.

### 1.2.1 Fluorescent Microscopy

Fluorescence microscope techniques offer powerful tools for studying almost any cellular process under the microscope [12]. Multicolor imaging is widely used to visualize distinct cellular and subcellular components, as well as other materials of interest, such as nanoparticles. Because of the abundance of synthetic fluorophores, live cell imaging dyes, and fluorescent protein tags available, this approach is feasible. The main experimental challenges in (multicolor) live-cell imaging are keeping cells physiological and minimizing photodamage while extracting data with the highest spatial and temporal resolution possible.

### 1.2.2 Widefield Microscopy

For observing dynamics in live cells, wide-field microscopes provide excellent temporal resolution and submicrometer spatial resolution. Background fluorescence from out-of-focus planes, on the other hand, is not discriminated, resulting in a blurry image



of 3D objects such as cells and tissues [13]. This problem was solved with the introduction of confocal microscopy, in which out-of-focus signals are effectively blocked by a pinhole placed in front of the detector. This key feature of confocal microscopes enables the creation of 3D images of thick specimens by acquiring successions of thin optical sections along the Z-axis [13, 14]. Despite this, laser scanning microscopes have very slow acquisition speeds because the specimen is raster scanned point by point by a single-beam laser light. The most certain way to reduce photodamage is to limit exposure to excitation light as much as possible. Spinning disk confocal microscopy is one of the best solutions for multicolor 3D live-cell imaging in real time (SDCM). SDCM is the most widely used high-resolution imaging method for intracellular dynamics [12]. It combines the benefits of scanning confocal microscopy with high-performance real-time imaging, making it particularly suitable for imaging for quantitative analysis [15].

### 1.3 Literature Survey

Cell nuclei segmentation in histopathology images is a critical step for microscopy-based cellular image analysis, which is the foundation of investigating many biologically related tasks, such as phenotype classification and cell function identification. Given the biological significance of this work, an efficient computer-aided system that can assist pathologists in automating this process is very demanding.

Traditional techniques for nuclei segmentation, also known as handcrafted feature based approaches, include a combination of methods for nuclei detection, which can provide a seed for each nucleus or the area of nuclei, and nuclei splitting [16–21]. The most common and popular method applied for nuclei detection is thresholding based in Otsu’s method [4, 5, 22]. Several shape and size based features such as, area, perimeter, shape factor, aspect ratio, color texture, blue ratio, color histograms, Laplacian of Gaussian response are also extracted from nucleus region to quantify the nucleus morphology and are utilized to guide the models in distinguishing between nuclei and non-nuclei regions [5, 23–26]. Other approaches include, for example K-means clustering, graph cuts based methods, H-maxima transform and combination of Laplacian of Gaussian with Euclidean distance map [4, 5]. However, cell overlapping commonly occurs in biopsy images. Therefore, if nuclei detection step generates the area of nuclei, after this stage many overlapping nuclei appear. In order to split them, several methods can be applied, for example: least square ellipse fitting, concavity detection, edge path selection, Fourier shape descriptor [5, 27]. On the other hand, if the seed of each nucleus has been obtained in the nuclei detection step, region growing [28], or marker-controlled watershed [28, 29] can be applied to obtain the contour of each nucleus. Nonetheless, several issues are associated with these traditional techniques. In some cases manual

tuning of parameters is necessary [4], in other cases the methods developed can be very specific to a given type of tissue or image [30]. Recently, the deep learning based approaches, which have been successfully applied in several tasks of computer vision, like image classification and object detection, are attracting the attention of researchers [30]. These methods have the ability to automatically extract important features from the image and present a good generalization capability, hence their superiority when compared to traditional techniques mentioned above [30].

deep learning frameworks have recently gained popularity as they have shown promising performance in different biomedical image segmentation tasks [31–40]. However, task of nuclei segmentation necessitates consideration of several issues, such as blurriness in the edge region of cells, diversity in image modalities, and variation in intensities across cancer subtypes [5, 41]. Among different deep learning based nuclei segmentation methods, U-net based architectures are most widely used such as U-net [42], U-net+ [43], U-net++ [44], FPN [45], FU-Net [46], and Micro-net [47]. In U-net+ [43] and U-net++ [44], the higher accuracy in nucleus segmentation is achieved by redesigning the encoding branch of the traditional U-net architecture to fuse more image features. A fast U-net architecture (FU-Net) is proposed in [46], where a bottleneck convolution layer is introduced in the encoder and decoder layer of the traditional U-net to make the model computationally efficient and accurate. Another variant of U-net, namely Micro-net, extends the U-net by utilizing a weighted loss function and processing the inputs at multiple resolutions [47]. However, accurate segmentation of nuclei is still a challenging task due to the difficulty in finding the segregation of the clustered nuclei in cell microscopy images [48]. One major concern in U-net based network is the use of downsampling operations, such as max or average pooling that usually ignore the Nyquist sampling theorem, resulting in partial or complete loss of high frequency data details and distorted basic structure presented by the low frequency components [49]. In the search of detail preserving pooling approaches, discrete wavelet transform (DWT) based pooling are reported in several previous works by replacing the pooling operation with DWT operation [49, 50]. However, considering the capability of pooling operation in finding relevant spatial information, it is more appropriate to combine the DWT information in a separate path for better data representation. Furthermore, the non-accountability of boundary or edge information in U-net based architectures may result in loss of prominent features in the edge and even small nucleus region, which is another important issue to be considered.

In recent years, attention module based deep neural networks have achieved tremendous performance in segmenting cell nuclei in histopathology images [51–58]. A hybrid-attention nested U-Net, namely Han-Net, is proposed in [59], which is integrated with a dense network to capture more effective feature information. In [60], a dual encoder attention U-net is proposed which employs a new secondary encoder input to the at-



tention U-net to capture the best attention for the input. A multi-task network based on U-Net is proposed in [61], where a context encoding layer is introduced behind each encoder and its output is processed by an attention module to fuse with the features of the decoding layer. In [62], a deep neural network is proposed, where a robust residual block is introduced for the extraction of high level semantic maps and an attention mechanism is incorporated with the decoder block for robust nuclei segmentation. However, all these attention based approaches consider the down-sampled signal as the attention input, which are prone to noise interruption.

Traditional data augmentation techniques, which mostly rely on performing geometric transformations, such as rotation, zoom, scaling, have been commonly used for training deep learning models for medical image segmentation [63–66]. The generative adversarial network (GAN) provides an alternative approach for generating synthetic images [67]. When used for the image segmentation task, such technique yields an inherent challenge of generating masks for synthetically generated images [68]. In [65], a two-stage GAN is proposed for nucleus segmentation, where the first stage generates synthesis masks and then incorporates them into the second GAN to produce synthetic image. In [66], both the training and testing datasets are augmented followed by the prediction of both of these sets which exhibits promising performance in the nucleus segmentation task. However, despite their dominating performances, GAN-based approaches can perpetuate biases inherent in the dataset or even amplify them [69].

There are several other methods which address the issue of segmenting the overlapping nuclei in histopathology images. For example, in [70], a deep learning approach is proposed to detect superior markers by regressing nuclear distance map for segmenting overlapping nuclei. Another deep neural network is proposed based on the prediction of horizontal and vertical distances of nuclear pixels to their centres of mass to separate clustered nuclei [71]. Although this method can significantly capture the separation of the nuclei, it fails to give more emphasis on boundary information. In ASPPU-net, a modified U-net is proposed to capture multi-scale nuclei features and obtain nuclei context information [72]. However, all of these techniques consider the spatial sampling operation to encode the spatial content of an image into low dimensional space which can eventually distort the low frequency components in the decoding stage resulting in loss of information. Moreover, these kind of architectures are not capable in capturing global contextual information and limited in modelling the long range semantic dependencies due to the small receptive field [73].

## 1.4 Motivation of This Research

The motivation of this thesis is to develop computer-aided schemes to segment cell nuclei in histopathology images. Accurate nuclei segmentation is the starting point for different cell-based analyses, such as cell function identification, its reactions to various treatments, and the underlying biological processes. Hence, automatic cell nuclei segmentation in histopathology images plays a crucial role in this regard. Among different cell nuclei segmentation methods, deep learning-based frameworks have been widely used recently due to their supremacy in terms of accuracy. Different deep learning frameworks are investigated in the nuclei segmentation task previously. Morphological operations are also utilized as the post processing stage in convolutional neural networks to obtain the approximate segmented masks and refine the obtained masks. Attention mechanism-based encoder-decoder architectures are also explored to address the nuclei segmentation task. However, these methods concentrate on utilizing the spatial or channel level information which can be distorted over subsequent downsampling operation of the encoder decoder architectures. Besides, no particular importance was provided in the edge regions of the nucleus regions. Deep neural network based on the prediction of horizontal and vertical distances of nuclear pixels to their centres of mass to separate clustered nuclei is also proposed in several works. Although this method can significantly capture the separation of the nuclei, it fails to give more emphasis on boundary information. Considering all the shortcomings of the previous works, an automatic computer-aided segmentation scheme utilizing a properly guided attention mechanism to prevent the loss of information problem present in the encoder-decoder architecture with more emphasis on boundary level information is in great demand.

## 1.5 Objective of this thesis

The objectives of this research with specific aims are as follows:

- To develop a method to transform different histopathology images into a common intensity level and color statistics that can leverage the segmentation performance.
- To develop an attention mechanism that can utilize multi-resolution features to guide the spatial level information of the encoder-decoder architecture.
- To investigate the performance of different multi-resolution transformations e.g. wavelet, contourlet, and Gabor transforms as the guide signal of the attention mechanism in nuclei segmentation task.

- To propose a boundary preserving unit along with a separate loss function that gives more emphasis on the boundary of the nuclei regions.
- To develop a technique for producing three segmented images from three different paths of the network by minimizing three separate loss functions and combining them to produce the final segmented output.
- To validate the performance of the proposed method by conducting experiments on three publicly available histopathology datasets.

## 1.6 Organization of the Thesis

The rest of the thesis is organized as follows:

In Chapter 2, an automated nuclei segmentation method is proposed based on attention network in which the attention module is guided by the boundary aware feature maps generated by another encoder-decoder network. This boundary extractor network is an encoder decoder network which is utilized to generate boundary aware signal. This boundary aware signal is then injected to the attention module of another encoder-decoder network which acts as the final segmentation network.

In Chapter 3, multi-scale directional contourlet filter driven attention network, namely ConDANet, is proposed which utilizes both the spatial attention of the convoluted output of the encoder-decoder network and multi-directional edge preserved output from contourlet transform. The contourlet filter helps to provide additional edge level information which allows the network to focus on the boundary region of the nuclei cell. Since the contourlet filtered output is directly injected to the network, the chance of losing important edge information can be minimized, which is a major concern in the deep neural network architecture. Additionally, wavelet based pooling technique is employed to prevent the loss of detail information and low frequency components present in the histopathology images. The combined operation of the proposed method is more effective in preserving the small sized nucleus which is present in a compact data space.

In Chapter 4, a boundary aware wavelet guided network, termed as BAWGNet is proposed. Prior to applying the data in the network, a preprocessing step is proposed to handle the lack of data diversity, and the variation in color statistics in different image modalities. Unlike the skip connection strategy of conventional encoder-decoder architecture, we propose an aggregation unit that introduces a guided attention mechanism to combine the encoder and decoder information and also generates boundary Aware information. In the guided attention unit, the spatial domain information at different encoder-decoder levels is further processed under the guidance of the wavelet domain information extracted from discrete wavelet transform. Moreover, the proposed bound-

ary aware unit inside the aggregation unit helps preserve the boundary information of the nucleus by incorporating a separate loss function which gives more attention to the nuclei edge regions. The proposed wavelet guided attention unit and boundary aware unit jointly leverage to identify small nucleus regions and separate the boundary of adjacent individual nuclei very efficiently.

Chapter 5 summarizes the outcome of this thesis with some concluding remarks and possible future works.

## Chapter 2

# Proposed Cell Nuclei Segmentation Framework Using Boundary Aware Feature Map Guided Attention Network

In cell nuclei segmentation, it is very important to extract each nucleus with a high level of precision. When conventional encoder-decoder based deep neural network architectures are used to segment nuclei in a cell, two problems are observed. Firstly, the extracted nucleus boundary does not match accurately with the ground truth with different types of distortions and discontinuities. Secondly, in some cases, neighboring boundaries are very close to each other and actual boundary cannot be found in the extracted boundaries. In order to overcome these problems, in this chapter, attention mechanism and wavelet based sampling strategy are incorporated in the encoder-decoder based segmentation architecture. Moreover, in order to extract boundary information precisely, a separate shallow boundary extractor (BE) encoder-decoder network is used that generates boundary preserving signal to guide the attention mechanism of the segmentation network. In order to get a very precise nucleus boundary estimation, a separate loss function is introduced for the shallow BE network. For this purpose, boundary only mask is generated using image morphological operation on given nucleus mask. Finally, Extensive experimentations have been done over three publicly available datasets containing a large number of subjects with complex nucleus orientations in different organs and wide varieties of patients.

## 2.1 Proposed Method

The cell segmentation task involves a pre-processing step followed by the proposed boundary aware feature map generation scheme and the proposed attention module based segmentation network. Details of these steps are presented in the subsequent subsections.

### 2.1.1 Preprocessing

In histopathology images, there are usually two types of modalities- brightfield and fluorescence used as depicted in Fig.2.1. In order to handle the data diversity present in this task, the images need to be augmented, contrast of histopathology images needs to be enhanced and different modalities of images needs to be transformed into a common intensity level. Three preprocessing steps involved in this proposed method are discussed here.

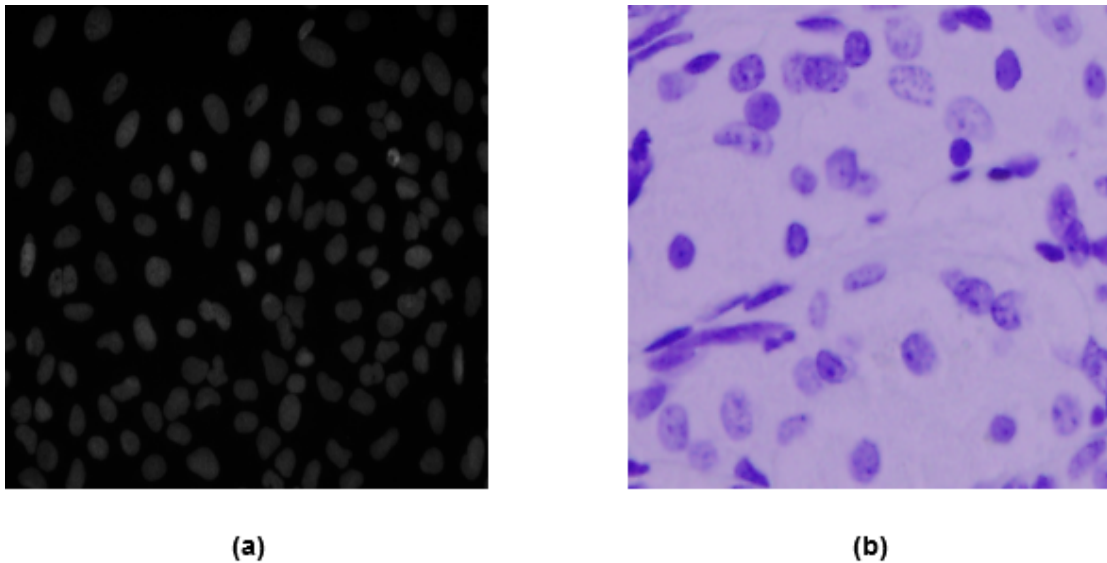


Figure 2.1: Histopathology image sample acquired from (a) Brightfield (b) Fluorescence microscopy techniques.

#### 2.1.1.1 Augmentation

In dealing with histopathology images, one major fact is the variation in images due to different acquisition techniques, which causes variation in angular position of the images as well as the deformation in the shape of the nucleus. Keeping this fact in mind, in order to handle the problem of limited number of samples, image formation based augmentation techniques such as rotation, flipping, and shearing are considered.

### 2.1.1.2 Intensity Level Transformation

One of the critical issues in the nucleus segmentation task is the generalization of the network to work equally well for different image modalities. Due to their variation in visual appearance and intensity level, it is difficult to train a universal model to work equally well for both of these modalities. The lack of abundant data further complicates this process and demands a preprocessing step to transfer all these modalities into a common intensity range. To achieve this objective, LAB color space transformation scheme is employed [74]. Transforming all the three channel images to LAB color space helps to preserve the original structure and maintains similar brightness and color statistics level leveraging the uniformity of data characteristics. At this stage, the two modalities show opposite color statistics (second row of Fig. 2.2).

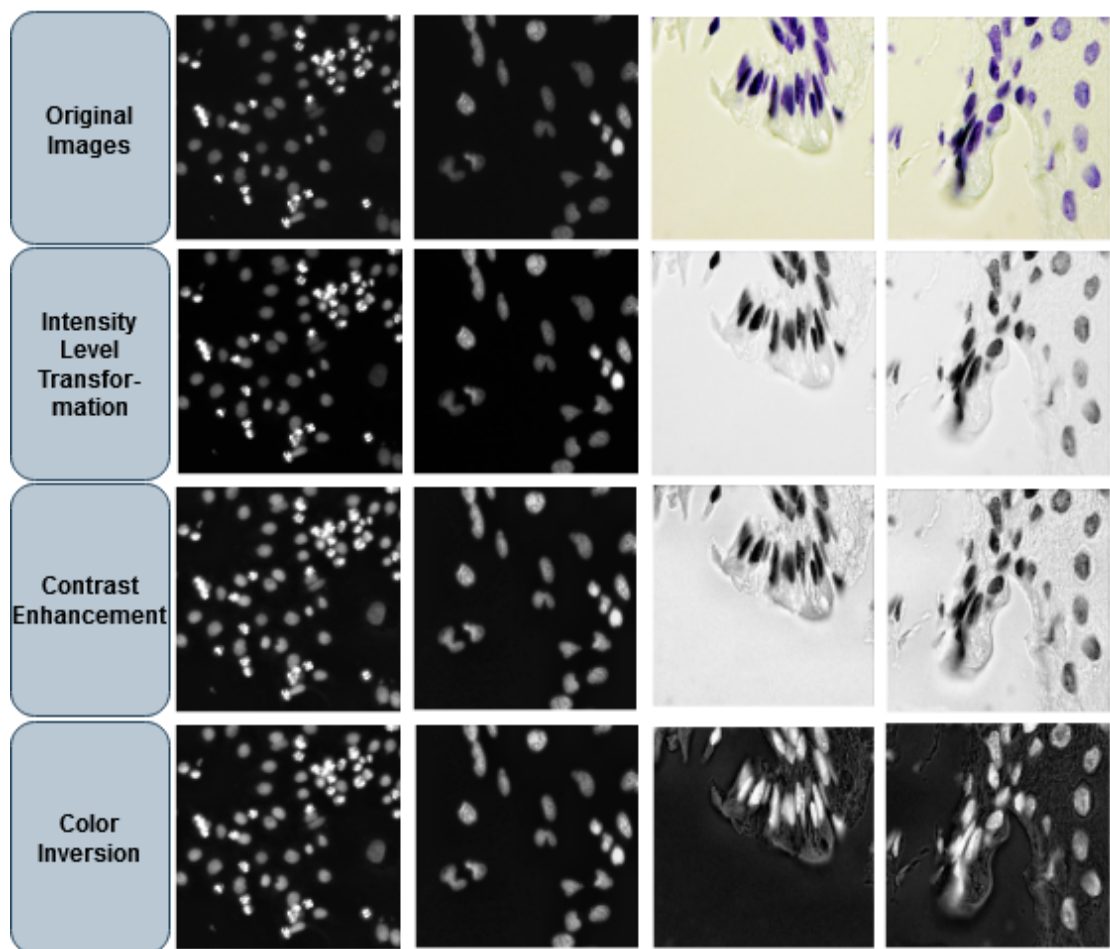


Figure 2.2: Visualization of microscopy images at different stages of preprocessing step.

### 2.1.1.3 Contrast Enhancement and Color Inversion

Usually, the nucleus possesses a small region in the cell space that might be neglected by a nucleus segmentation algorithm. In order to overcome this problem, various types of contrast enhancement techniques are employed that enhance the visibility range of

the small nuclei region. In this proposed method, contrast limited adaptive histogram equalization (CLAHE) algorithm is applied to enhance the contrast of the histopathology images [75]. Ordinary adaptive histogram equalization (AHE) tends to overamplify the contrast in near-constant regions of the image, since the histogram in such regions is highly concentrated. As a result, AHE may cause noise to be amplified in near-constant regions. Contrast Limited AHE (CLAHE) is a variant of adaptive histogram equalization in which the contrast amplification is limited, so as to reduce this problem of noise amplification. Among the two modalities of the histopathology images, brightfield images are widely employed in clinical institutions. The overall intensity level of Fluorescent images are much higher than that of brightfield histopathology images. Hence, a color inversion operation is performed on all images to shift the intensity levels of fluorescent histopathology images. The inversion process of a random pixel  $x_{ij}$  in an image  $X$  with mean intensity level  $\bar{X}$  is as follows

$$x_{ij} = \begin{cases} 255 - x_{ij} & \text{if } \bar{X} > 127 \\ x_{ij} & \text{if } \bar{X} \leq 127 \end{cases} \quad (2.1)$$

This operation ensures that the images with mean intensity below a threshold level (first two columns in Fig. 2.2) have no effect whereas, those having the opposite intensity statistics shift their intensity to match the intensity level of brightfield images.

### 2.1.2 Proposed Boundary Aware Feature Map Guided Attention Network

There are several major challenges lie in nuclei segmentation task such as, intensity variation among images acquired from different imaging techniques and the unclear appearance of the boundary pixels due to blurriness. While the first issue can be solved by traditional image processing techniques like the preprocessing step of our proposed method, the later issue requires special consideration. However, most of the studies in this research direction fail to consider this issue. Hence, to address this concern, a deep learning framework is proposed in this chapter which comprises of two separate networks. The first one is the the primary network which is termed as the boundary extractor network due to its acquisition of edge or boundary level information. It is a traditional encoder-decoder network with less number of parameters compared to the original U-net architecture. The second one is the final segmentation network which is responsible to collect the boundary level information from the primary network and utilize it through an attention module. Additionally, we adapt the wavelet based pooling strategy in our final segmentation network which is effective in preserving the high frequency signal and can prevent the loss of information compared to the traditional



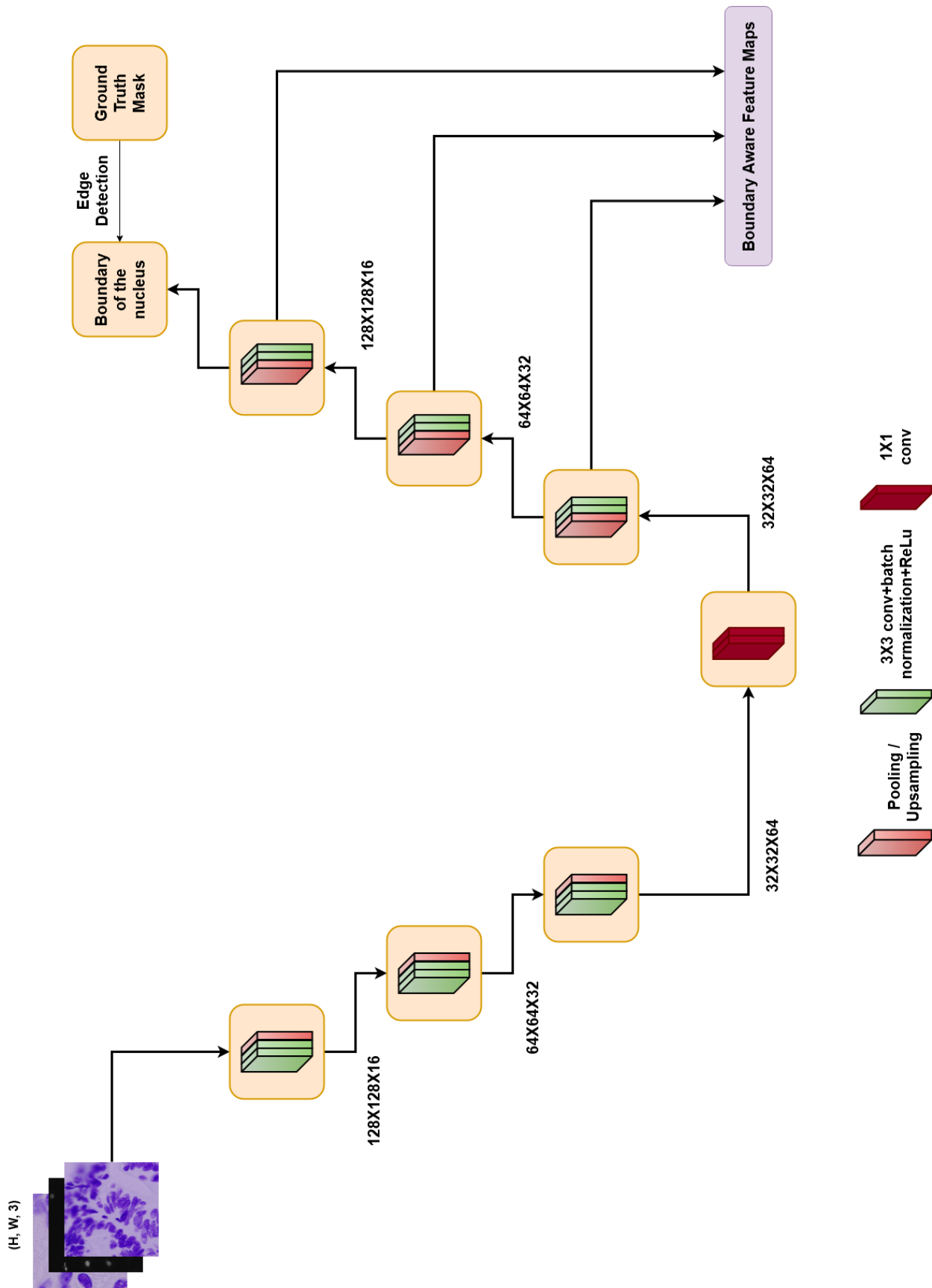


Figure 2.3: Graphical overview of the proposed Boundary Aware Signal Generation Network. Here, the edge regions of the nucleus present in a histopathology images are utilized as the mask.

sampling techniques like max or average pooling and unpooling.

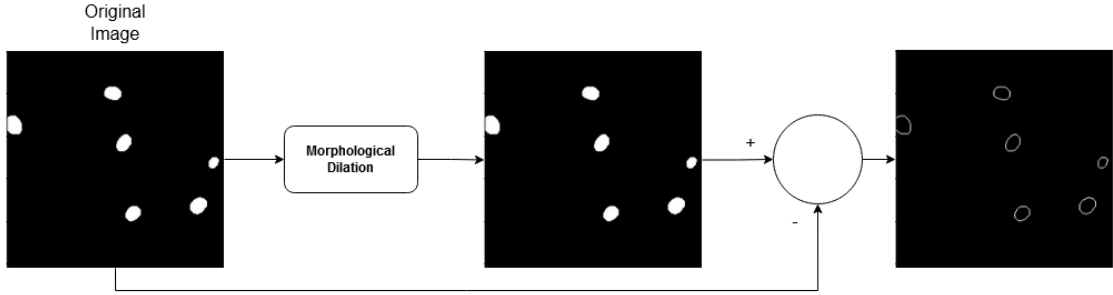


Figure 2.4: Morphological dilation based edge extraction.

### 2.1.2.1 Boundary Extractor Network

The proposed boundary extractor network is a U-net like architecture but with less number of convolutional filters to reduce the computational complexity. The network is depicted in Fig 2.3. Here, instead of using the segmented mask of the nucleus regions, we exploit the synthetic edge of the nucleus regions in a histopathology image as the true mask of this network. The boundary of the nucleus from the ground truth mask is obtained by image dilation operation which is visualized in Fig 2.4 [76]. The encoder and decoder part of this network is similar in structure. Both the encoder and decoder have three blocks which contain two convolutional layers followed by a batch normalization layer and a recursive linear unit (ReLU) activation function. Batch Norm is a normalization technique done between the layers of a Neural Network instead of in the raw data. It is done along mini-batches instead of the full data set. It serves to speed up training and use higher learning rates, making learning easier. The Convolutional layers in each stage contains  $2^{i-1} \times 16$  number of filters with kernel size  $3 \times 3$  (both in encoder and decoder level) where,  $i = 1, 2, 3$ . A  $2 \times 2$  max pooling layer is placed after each convolutional layers to reduce the feature dimension. We extracted the feature maps from both the encoder and decoder of the proposed network and analysed their ability in extracting the boundary information. The feature maps extracted from the encoder part and the decoder part are visualized in Fig. 2.5. After careful inspection, it is found that the decoder part are more efficient in learning the edge characteristics of the nucleus while the encoder mostly concentrate on the overall structural of the nucleus region. Since, the decoder serve the purpose of extracting boundary information more efficiently, the feature maps extracted from the three decoder stages are chosen for the boundary aware signal, which are utilized to guide the second network i.e. the segmentation network.

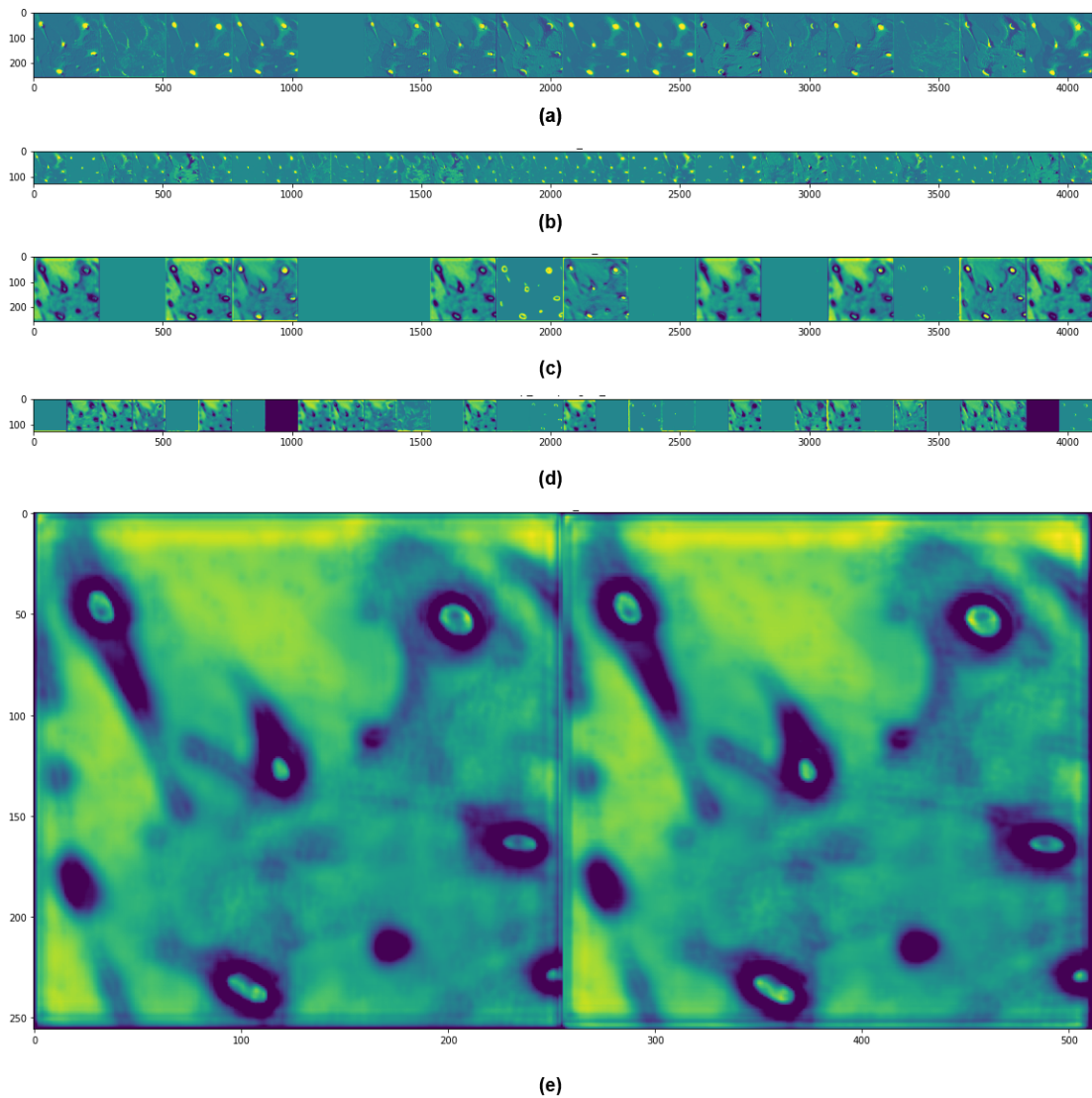


Figure 2.5: Visualization of feature maps at different stages of encoder and decoder. (a-b) feature maps extracted from encoder stage, (c-d) feature maps extracted from decoder stage, (e) a closer look of the feature maps extracted from decoder stage.

### 2.1.2.2 Boundary Aware Feature Map Guided Attention Network

The proposed segmentation network is similar to the boundary aware signal generation network except the attention module part and the wavelet sampling strategy. The detailed architecture is shown in Fig. 2.6. In order to provide edge level information to the network, an attention mechanism is added to the network which is guided by the boundary aware feature map extracted from the previous network containing the edge aware information. Additionally, Wavelet based sampling strategy can overcome the problem of loss of information originated from the traditional down sampling of the encoder part by preserving the spectral content of the encoder unit and reconstruct the original content with a minimal loss in the decoder part.

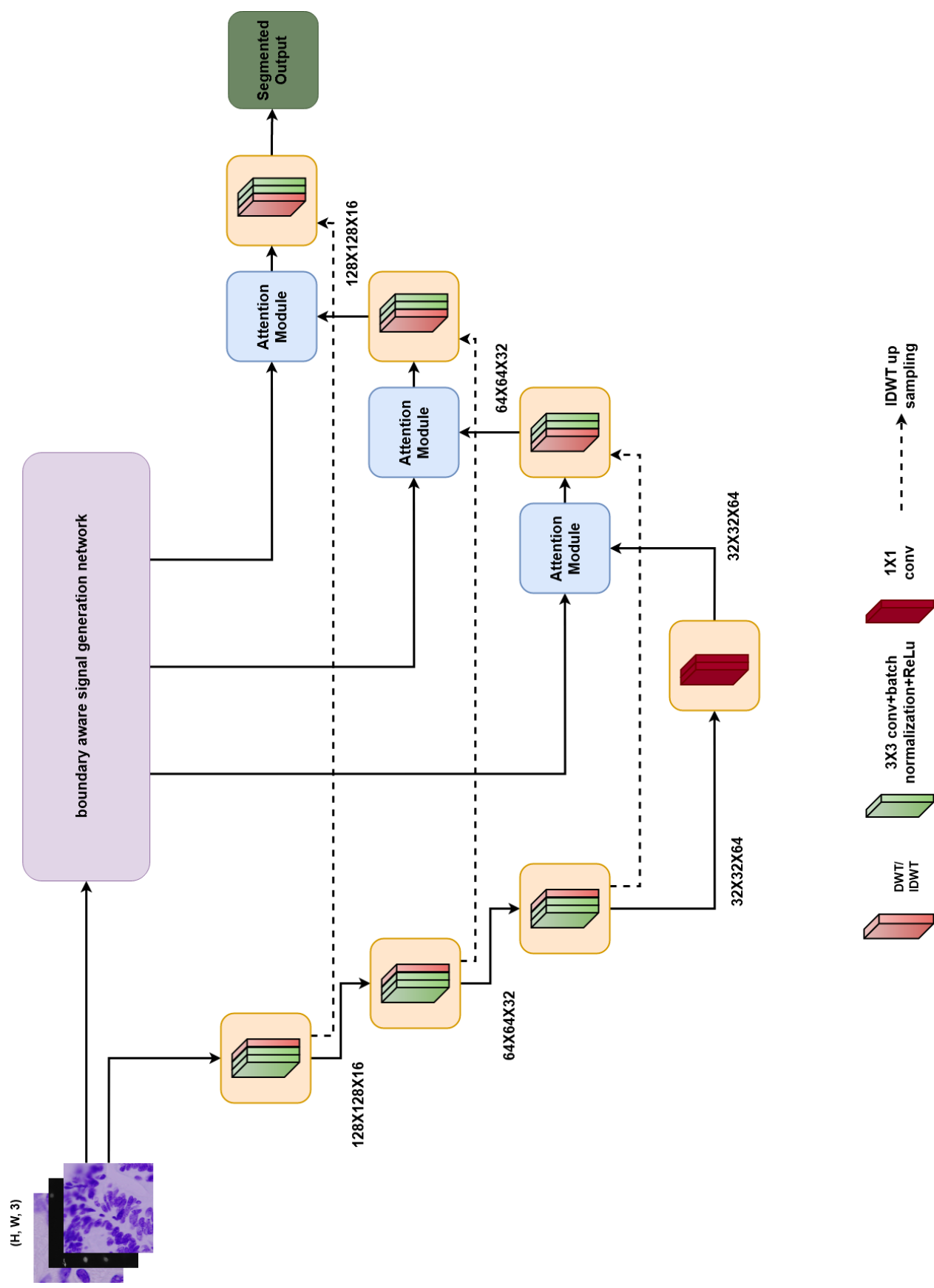


Figure 2.6: Graphical overview of the proposed boundary aware feature map guided attention network.

**Wavelet Based Pooling Strategy** Usually, feature dimension reduction of neural network known as sampling strategy has traditionally been performed by the max or average pooling strategy in well-known encoder-decoder architectures such as U-net and Unet++. These downsampling operations, however, frequently disregard the Nyquist sampling theorem, resulting in aliasing of different data frequency components of the signal. As a result, data details may be lost as a result of these operations. The max-unpooling or bilinear interpolation approaches, on the other hand, are widely used for up-sampling operations in the decoder part. Although they are used to recover data details, they cannot be completely reconstructed. To solve this problem 2D discrete wavelet transform (DWT) and the inverse DWT (IDWT) are adopted in this method. A wavelet-based sampling strategy can address this issue by preserving the spectral content in the encoder unit and reconstructing the original content with minimal loss in the decoder. The input image is first processed through a series of encoding operations using a convolution filter, a batch normalization layer, and a ReLu activation operation. The down-sampling and up-sampling operations are then replaced by the DWT and IDWT operations at each encoder and decoder stage. The DWT decomposes the signal into three high frequency components and one low frequency components as follows:

$$I_{mn} = (f_{mn} * I)_{\downarrow 2}, \quad (2.2)$$

where  $(\cdot)_{\downarrow 2}$  stands for the 2-times downsampling process,  $f_{mn}$  stands for one of the four filters (high and low pass filters) to be applied, and  $*$  denotes the convolution operation. The filtered output is denoted by the subscript  $I_{mn}$ , where  $ll$  denotes a low pass filter and  $lh$ ,  $hl$ , and  $hh$  denote three high pass filters. In order to recover the feature map in the decoder stage the IDWT operation is performed by combining the low frequency component transmitted into the network's subsequent layers and the high frequency components coming from the encoder stages.

$$I = \sum f_{mn} * I_{\uparrow 2} \quad (2.3)$$

Here,  $(\cdot)_{\uparrow 2}$  represents the 2 times upsampling operation.

**Attention Module** In order to guide the spatial level information with the extracted boundary aware information, an attention mechanism is designed in the proposed method which is shown in Fig. 2.7. In this block, the boundary aware feature maps ( $g$ ) to guide the network to leverage the salient edge region of the given images. First,  $g$  and the input tensor ( $X$ ) are linearly transformed using a  $1 \times 1$  convolution block, batch normalized and added together. The introduction of the batch normalization in the attention module plays a significant role here. Since, the extracted feature maps are coming from another network, it is important to keep their values consistent with the feature maps coming

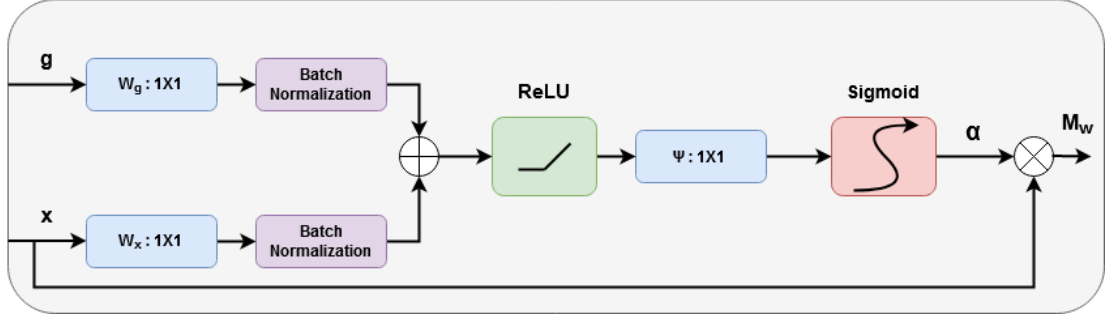


Figure 2.7: A visualization of the attention mechanism used in the segmentation network to combine the boundary aware and spatial domain information.

from the decoder part of our segmentation network. Here, batch normalization acts as the feature normalization that creates consistency between two different features. The convolution operation allows to extract the most promising features from them and the addition operation further enhance the features those are found relevant in both spatial and frequency domain. The resultant output is then passed through a ReLU activation function  $\sigma_1$ , to prevent vanishing gradient problem followed by a  $1 \times 1$  convolution layer  $\phi$  to find the gating coefficient,  $g_{att,c}$ . An additive attention is used to obtain this gating coefficient. The overall process can be represented by the following expression,

$$g_{att,c} = \phi^T(\sigma_1(w_1^T x + w_2^T g + b_g)) + b_\phi, \quad (2.4)$$

where  $w_1$ ,  $w_2$ , and  $\phi$  are the linear transformations and  $\sigma_1$  is the ReLU activation operator which is defined as  $\sigma_1(x) = \max(0, x)$ . Afterward, the attention coefficient  $\alpha$  is obtained by applying a sigmoid activation function  $\sigma_2$  to  $g_{att,c}$ . The reason for selecting sigmoid function over softmax is that the sequential use of softmax function may result in sparse activation of the output [77]. Finally, the output of the wavelet guided attention block ( $M_w$ ) is found from the element wise multiplication of the attention coefficient and the input tensor  $X$ . The final output is as follows,

$$M_{w,i} = \sigma_2(g_{att,c}) \cdot X \quad (2.5)$$

Since, the problem is a binary class segmentation problem, the conventional binary cross-entropy loss is utilized as the loss function of this network which is defined by the following equation:

$$\mathcal{L} = \frac{1}{N} \sum_{j=1}^N p_j \log y_j + (1 - p_j) \log(1 - y_j) \quad (2.6)$$

where  $p_j$  denotes the predicted probability value of  $j^{th}$  sample,  $y_j$  represents the corresponding label, and  $N$  is the total number of samples.

## 2.2 Results and Discussions

In this section, the description of the datasets used in this method, the required setup for our experiment and results found from extensive experimentation on several publicly available datasets are presented and discussed to validate the effectiveness of the proposed scheme.

### 2.2.1 Dataset Description

To verify the effectiveness of the proposed algorithm, three publicly available datasets are employed. The first one is Data Science Bowl (DSB-2018) which was released by Kaggle for competition purpose [78]. The dataset contains a large number of segmented nuclei images that were acquired under different conditions and vary in cell type, shape, magnification, illumination status, and imaging modality (brightfield and fluorescence). The training set of this dataset contains 670 images in which 546 images are brightfield and the rests are fluorescence. The test set contains 65 images and their manual segmentations are provided for the evaluation purpose.

The second dataset used for evaluation purpose is multi-organ nuclei segmentation challenge (MoNuSeg) which contains a diverse set of Hematoxylin-Eosin (HE) stained tissue images [79]. It contains 30 images with around 22,000 nuclear boundary annotations for training and 14 images with 7,000 nuclear boundary annotations for testing purpose. This dataset was collected from 18 hospitals and a wide range of patients. The size of all the images is  $1000 \times 1000$ .

The final dataset is the triple negative breast cancer (TNBC) dataset which consists of 50 images with a total of 4,022 annotated cells, including normal epithelial and myoepithelial breast cells (localized in ducts and lobules), invasive carcinomatous cells, fibroblasts, endothelial cells, adipocytes, macrophages and inflammatory cells (lymphocytes and plasmocytes) [70]. The size of the images is  $500 \times 500$ . The sample images collected from these three datasets are visualized in Fig 2.8.

### 2.2.2 Experimental Setup

In the data augmentation step, the images are augmented by three operations. The images are rotated by  $30^\circ$  and  $60^\circ$ , flipped horizontally, and sheared by following affine

transformation  $\begin{bmatrix} 1 & 0 & 0 \\ s & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ , where  $s$  is selected as 0.3. During the training period, dif-

ferent hyperparameters are chosen based on the performance of the model. Adam optimizer with a learning rate of 0.001 and a decaying rate of 0.99 after each 10 epochs



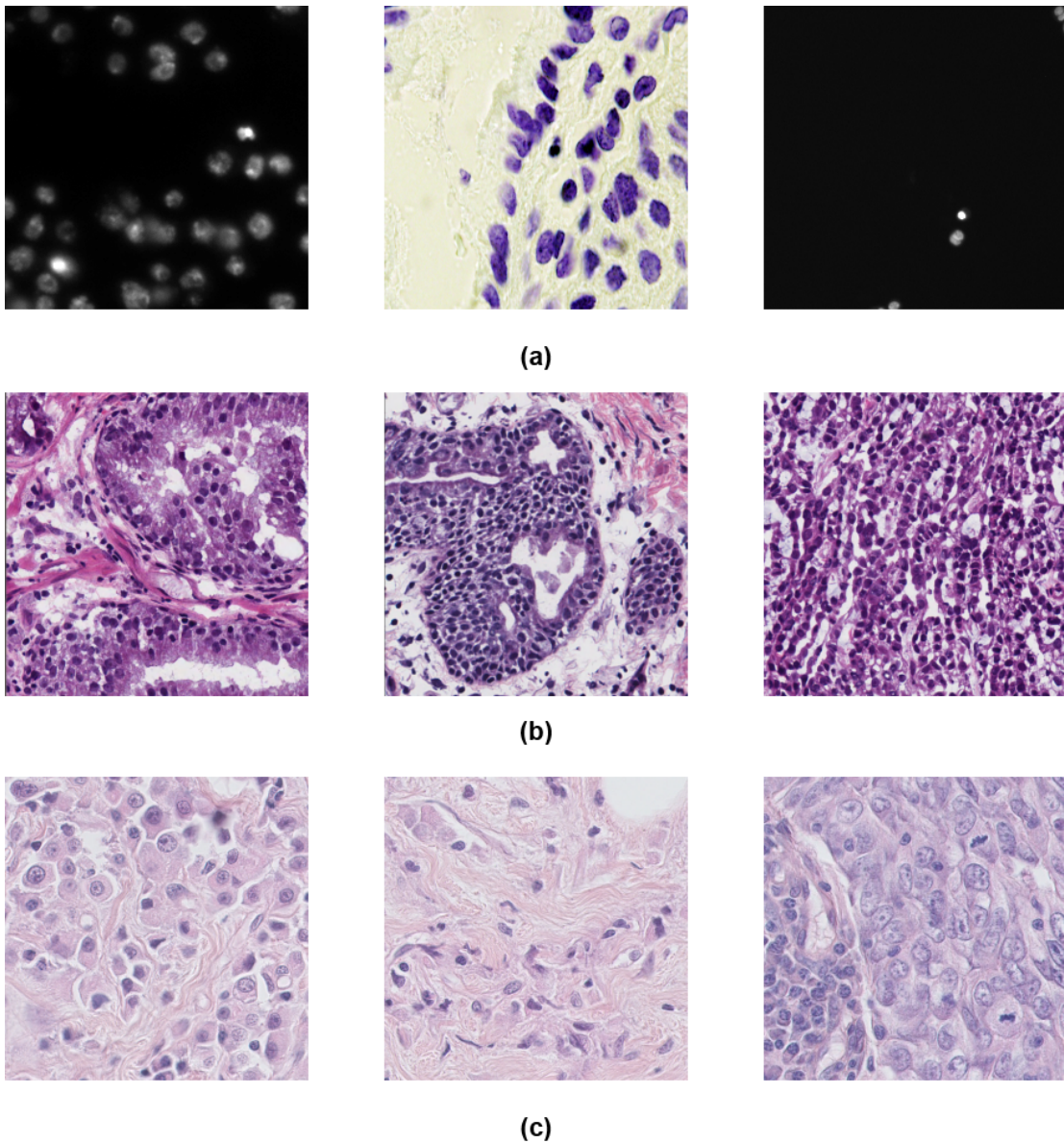


Figure 2.8: A visualization of the samples collected from (a) Data Science Bowl 2018, (b) MoNuSeg and (c) TNBC datasets.

is chosen for the primary network and the network is trained for 50 epochs. Adam is a method for efficient stochastic optimization that only requires first-order gradients with little memory requirement. The method computes individual adaptive learning rates for different parameters from estimates of first and second moments of the gradients; the name Adam is derived from adaptive moment estimation. This method is designed to combine the advantages of two popular methods: AdaGrad [80], which works well with sparse gradients, and RMSProp [81], which works well in on-line and non-stationary settings. Some of Adam's advantages are that the magnitudes of parameter updates are invariant to rescaling of the gradient, its stepsizes are approximately bounded by the stepsize hyperparameter, it does not require a stationary objective, it works with sparse gradients, and it naturally performs a form of step size annealing. The secondary or segmentation network is optimised with same hyperparameters, but is trained for 100



epochs. All the experimentations have been implemented on the Google cloud platform with NVIDIA P-100 GPU as the hardware accelerator.

For the DSB dataset, all the images are resized into  $256 \times 256$  to feed into our network. The size of each images of the MoNuSeg dataset  $1000 \times 1000$ . To make it compatible with the input size of our network, each images are zero padded to change the size of the network to  $1024 \times 1024$  and then 16 non-overlapping image patches of size  $256 \times 256$  are extracted. These cropped images are then acted as the inputs of the network. Finally, for TNBC dataset, the size of the images are  $512 \times 512$ . Hence, each images are tiled to four image patches with size of  $256 \times 256$  are employed for the training purpose. During the testing phase, the similar patches are employed in the model and their segmented images are merged to obtain the segmentation results of the whole image.

Several standard evaluation metrics are utilized to assess the performance of the proposed model which are stated as follows:

$$\text{Dice} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}} \quad (2.7)$$

$$\text{IOU} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \quad (2.8)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (2.9)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (2.10)$$

The dice score is a statistic used to gauge the similarity of two samples [82]. Intersection over Union (IoU) is a measure that specifies the amount of overlap between the predicted and ground truth bounding box. Here, TP, FP, and FN represent true positive, false positive, and false negative, respectively. For training and testing purpose of the DSB-2018 and TNBC datasets, the images are splitted as 80:10:10 for training, validation, and testing. Moreover, the results on the test set of MoNuSeg dataset are also reported and we compare it with the state of the art image segmentation as well as nuclei segmentation networks to corroborate the effectiveness of the proposed network.

### 2.2.3 Analysis of the Segmentation Performance

At first, an ablation study is provided to describe the performance of each of the modules of the proposed methodology. Eventually, the performance of the proposed network is compared with that obtained by the other state-of-the-art segmentation networks and its superiority is explained in qualitative and quantitative manner.

Table 2.1: Ablation study of the proposed Network on three datasets.

Prepr.	Wavelet Pooling	Attention Unit	DSB-2018		MoNuSeg		TNBC	
			Dice	IOU	Dice	IOU	Dice	IOU
X	X	X	82.46	73.52	73.08	58.92	66.51	51.41
✓	X	X	84.34	76.97	76.02	62.97	70.45	54.83
✓	✓	X	85.40	77.57	78.01	65.09	71.98	55.55
✓	✓	✓	88.05	79.98	80.96	69.92	74.25	55.87

Table 2.2: Performance comparison of different architectures in the DSB-2018 dataset (8:1:1 split setup)

Network	Dice(%)	IOU(%)	Precision(%)	Recall(%)
Segnet [83]	75.48	62.36	84.08	68.47
U-Net [42]	86.75	76.97	85.55	87.98
nucleAIzer [84]	86.92	75.89	82.32	90.91
Attention U-net [51]	87.30	77.51	79.97	94.81
CENet [85]	87.25	76.50	79.21	91.90
FPN [45]	87.34	77.97	80.34	95.67
Our Method	<b>88.05</b>	<b>79.98</b>	<b>81.07</b>	<b>95.81</b>

### 2.2.3.1 Ablation Study

To represent the individual performance contributed by different components of our proposed model, an ablation study is performed under different settings of the method. The summary of this study is depicted in Table.2.1. Here, we utilized two standard performance evaluation metric, dice coefficient and intersection over union (IOU) to represent the improvement contributed by each module. When utilizing preprocessing methods, dice score improvement between 2.28 and 5.92 percent is seen for the three datasets compared to the baseline. The IOU performance also increases by 4.7% to 6.9%. This improvement in performance demonstrates the efficacy of the preparing the data to the model’s input and removing discrepancies between various the imaging modalities used in histopathology. Additionally, the wavelet pooling module in parallel produces somewhat improved DSB-2018 results (upto 1.25 percent), while a better improvement is found in the MoNuSeg and TNBC datasets (2.17–2.61%). Finally, we integrates the attention module into the network. Integrating the attention module with the wavelet pooling module, the best results are obtained in all three datasets with an improvement between 4.36-4.74% in terms of dice score and 4.17-8.91% in terms IOU. Here, while wavelet pooling works as a simple pooling strategy on the convoluted signal in intermediate layers, the attention module utilizes the additional boundary aware information to the network and works as an effective tool to identify the contour around the nucleus region.

### 2.2.3.2 Quantitative Analysis

In order to compare the effectiveness of our proposed method, a comparative analysis is presented for the TNBC, MoNuSeg, and DSB-2018 datasets in Table 2.2, Table 2.3,

Table 2.3: Performance comparison of different architectures in the MoNuSeg dataset (8:1:1 split setup)

Network	Dice(%)	IOU(%)	Precision(%)	Recall(%)
Unet [42]	74.67	60.89	71.24	78.44
DIST [70]	77.31	63.77	75.31	79.41
Segnet [83]	77.44	63.3	73.91	81.32
Attention U-net [51]	78.67	66.51	77.87	82.33
CENet [85]	79.35	65.86	77.86	81.55
FPN [45]	80.02	66.89	74.98	82.19
Our Method	<b>80.96</b>	<b>69.92</b>	<b>78.82</b>	<b>87.21</b>

Table 2.4: Performance comparison of different architectures in the TNBC dataset (8:1:1 split setup)

Network	Dice(%)	IOU(%)	Precision(%)	Recall(%)
U-Net [42]	68.61	52.92	65.94	72.54
Mask-RCNN [86]	70.54	52.97	66.27	75.39
DIST [70]	70.51	56.34	66.82	74.63
Micro-Net [47]	71.23	53.71	66.52	76.61
Attention U-net [51]	71.43	54.21	70.01	76.03
FCN [87]	72.67	50.62	67.53	78.53
CEnet [83]	73.88	54.93	71.11	76.71
Our Method	<b>74.25</b>	<b>55.87</b>	<b>71.32</b>	<b>78.42</b>

and Table 2.4, respectively. It is noticed that our proposed network outperforms all the previously proposed segmentation networks by a significant amount in all the evaluation metrics. The proposed networks exhibits performance improvement over FCN, Unet, Segnet and other benchmark networks in nucleus segmentation task. Aside from these networks, it has also shown its superiority compared to the state-of -the-art nuclei segmentation network, CE-Net. Using the proposed framework, 4.25%, 2.15%, and 1.47% dice improvements are achieved over the results obtained in CE-net for the TNBC, MoNuSeg, and DSB-2018 datasets, respectively. The proposed network’s robustness and efficacy in determining the nuclei region can be corroborated by the high scores of other metrics reported. The proposed network’s ability to provide additional boundary aware information from the feature generation network through the attention mechanism provides additional information to the network. Moreover, with the help of wavelet pooling strategy and the preprocessing steps, it performs efficiently in capturing the small nuclei regions and thus improves both the precision and recall.

### 2.2.3.3 Qualitative Analysis

Quantitative analysis alone cannot always define the effectiveness and superiority of a method. The efficacy of the model can only be identified if the model performs outstandingly in the critical cases of a particular problem. Hence, the segmentation performance of different networks along with our proposed network in some challenging cases are presented in Fig.2.9. The presence of other cell organelles with similar in-

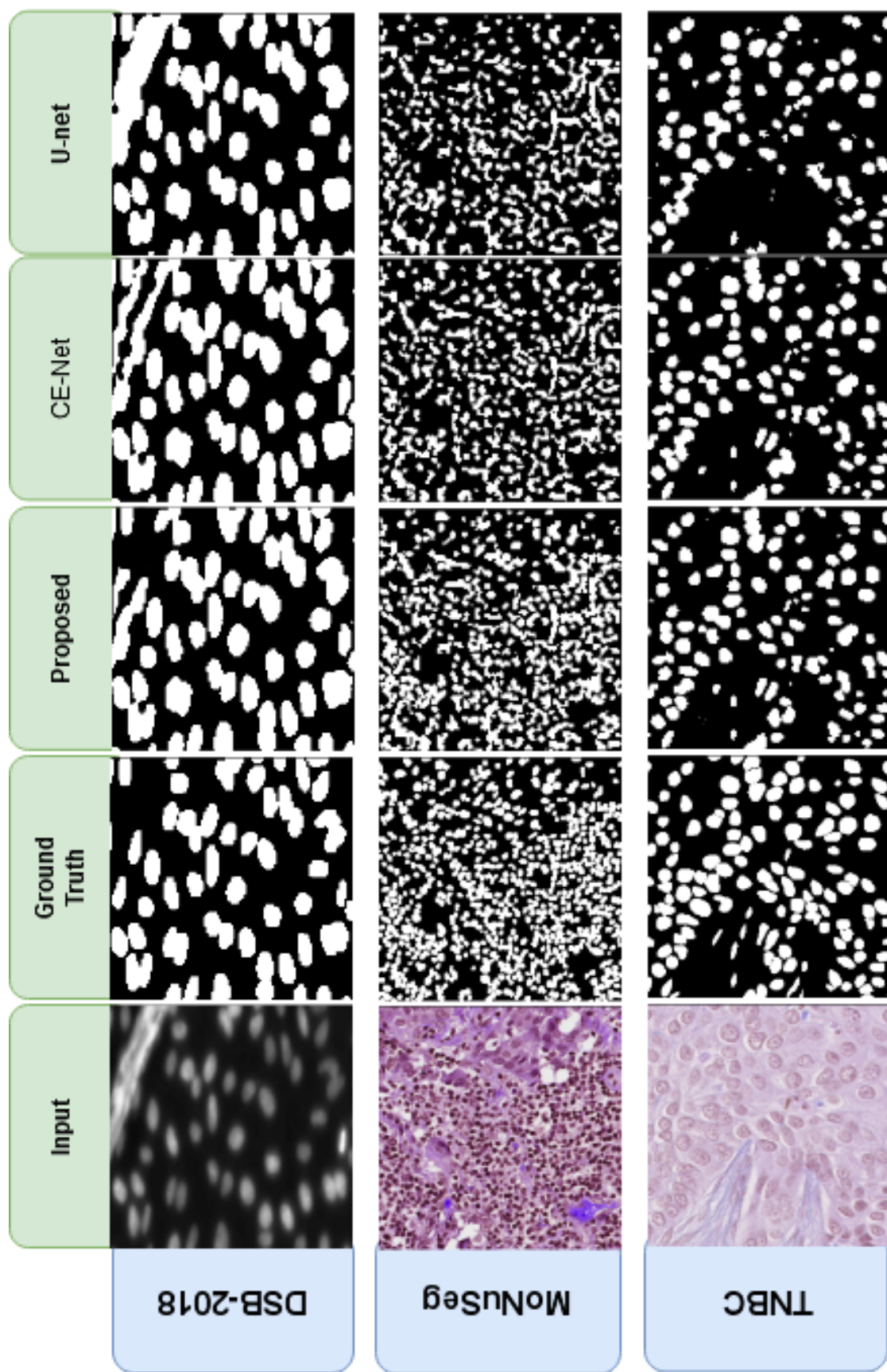


Figure 2.9: Visualization of the nuclei segmentation performance of some of the state-of-the-art networks in three datasets.

Table 2.5: Computational Efficiency Analysis of Numerous Architectures along with the Performance of the proposed method.

Methods	Number of Parameters (millions)	Inference Time (seconds)
Unet	31.0	0.1
MicroNet	184	0.18
DIST	8	0.12
SegNet	14.7	0.17
Attention U-net	8.727	0.09
CENet	38.8	0.12
FPN	8.4	0.137
Our Proposed	3.4	0.1

tensity levels to the nucleus, as well as the position of the nucleus in close proximity to each other, might be responsible for a significant number of false positive cases. It is evident from the segmentation performance of the other networks that all of them struggle to address these issues. On the other hand, our proposed method with its efficient utilization of spatial and boundary level information considerably addresses these issues and significantly shows its performance in the challenging situation of nuclei segmentation task. The computational efficiency of different networks are summarized in Table 2.5. Although the proposed method utilizes a separate architecture to extract the boundary preserving signal, the utilization of the shallow encoder decoder network limits the overall computational cost. Also, the network works almost equally faster with respect to the the Attention U-net with almost 2.5 times memory efficiency.

### 2.3 Conclusion

In this paper, we address two major issues present in the nuclei characteristics. The first one is the small sized nuclei region and the other one is the blurred edge region. Both of them are challenging situations because of their presence in low dimensional space for which they can be excluded during the segmentation operation. To handle these issues, two necessary modifications are conducted in the traditional encoder-decoder architecture. First, a boundary aware feature map generation network is introduced which produces edge level feature map that are used to guide our segmentation network through an attention mechanism. Additionally, it is found that the wavelet pooling based sampling strategy can solve the problem of identifying the small-sized nuclei regions by its effective reconstruction capability in the decoder part which can not be possible in the sampling technique like max-pooling or average-pooling operation. The qualitative and quantitative analysis of the proposed method with other modern deep learning architectures including the attention based network and context aware deep learning architectures clearly demonstrates the superior performance of this network.

## Chapter 3

# Proposed Cell Nuclei Segmentation Scheme Based on Contourlet Driven Attention Mechanism

A typical human body contains 30 trillion cells, each with its nucleus containing DNA that programs the cell [78]. The morphology and nucleus polymorphism presented on slides of Hematoxylin and Eosin (HE) stained microscopic tissue images allow doctors to identify cell functions and even grade a patient's cancer stage. As a result, accurate nuclei segmentation from histopathology images can help doctors treat patients more effectively and assess treatment efficacy. Usually the manual segmentation of nuclei in cell histopathology images is exhausting and time-consuming for clinicians which creates a demand for an accurate and efficient computer-aided technique. Despite previous efforts to automate this process, it still remains a difficult task due to the presence of overlapping nuclei and complicated boundary properties [41, 88]. Therefore, appropriate automation in reliable and accurate cell nuclei segmentation in medical image analysis is highly desired to assist expert clinicians and increase the efficacy of this task.

In this chapter, a deep learning-enabled segmentation framework, namely ConDANet, is developed which consists of an attention mechanism driven by the contourlet transform of the histopathology images. Contourlet transform not only exploits the advantage of the multi-scale and time frequency localization properties of wavelets but also provides a high degree of directionality. These features enhance the proposed attention mechanism to extract the fine edge details of the nuclei regions. Additionally, the wavelet pooling strategy instead of the traditional max-pooling and average-pooling operation of a convolutional neural network which preserves the original textural content of the nucleus in histopathology images and prevents the loss of information in the subsequent sampling operation of the encoder-decoder part of the network. Finally, the proposed



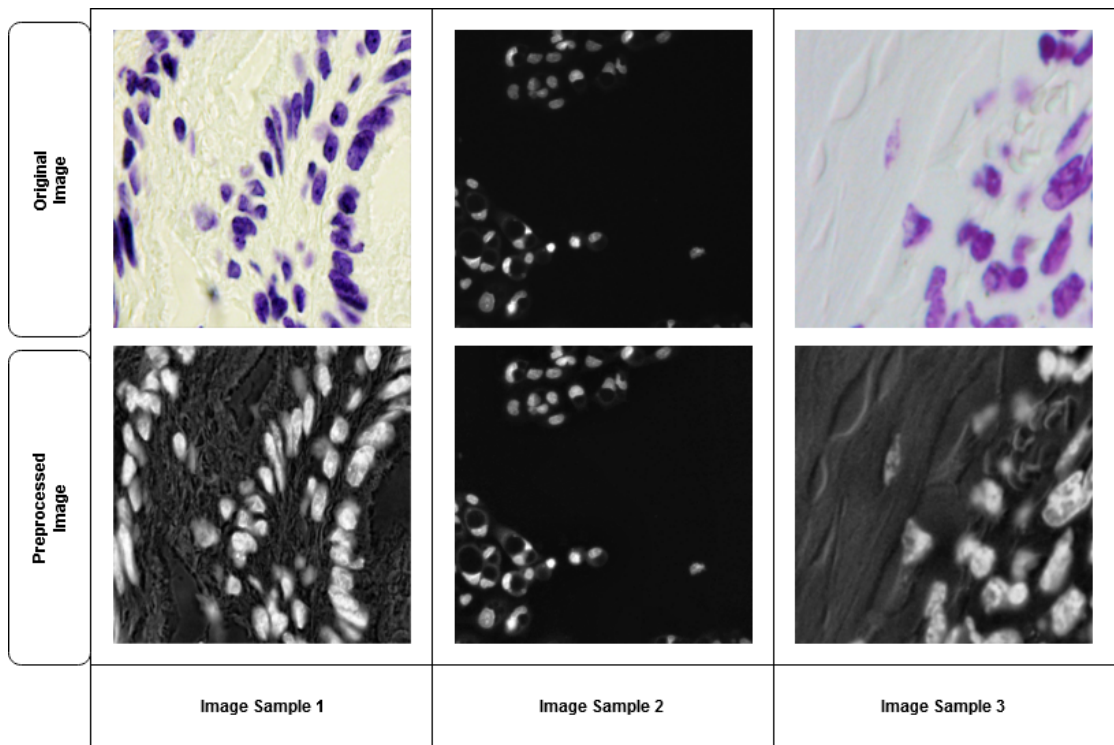


Figure 3.1: Visualization of microscopy images at different stages of preprocessing step.

method is employed for analyzing three publicly available histopathology datasets to manifest its effectiveness of it in segmenting nuclei from cellular images extracted from a wide variety of organs and patients.

### 3.1 Proposed Method

The cell segmentation task involves a pre-processing step followed by the proposed contourlet driven attention network (ConDANet). Details of these steps are presented in the subsequent subsections.

#### 3.1.1 Preprocessing

Usually, the two most commonly employed microscopy systems are fluorescent and brightfield microscopy which are used to visualize a single molecules and whole cell structure. The images acquired from these two modalities vary in their intensity level, color statistics and contrast information. In order to reduce such variations among images acquired from two different modalities, transformation of these images into similar intensity level and color statistics is required. Moreover, considering the lack of data availability is a major concern. To address these issues, a set of preprocessing operations are employed in this method. At first, several augmentation techniques are applied

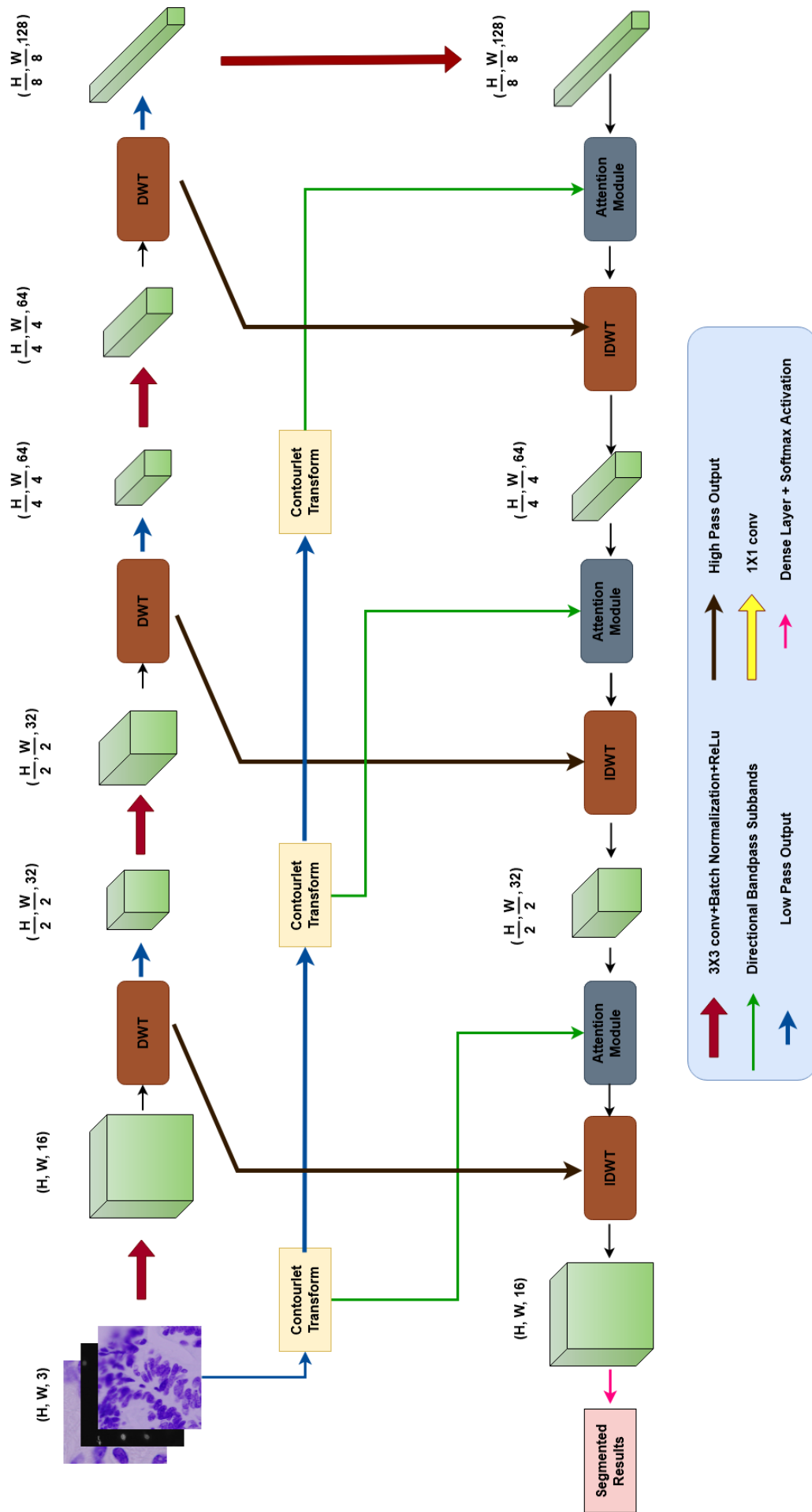


Figure 3.2: Graphical overview of the proposed contourlet driven attention network (ConDANet).



to handle the data scarcity issue. Next, to handle the variation in intensity level across the modalities, all the three channel images are transformed to LAB color space which ensures the preservice of the original cell structure and maintains the uniformity along the data [74]. To enhance the visibility range of each of the nuclei more specifically in the brightfield images, contrast level image enhancement is an important step. Hence, contrast limited adaptive histogram equalization (CLAHE) is applied to enhance the contrast of the histopathology images [75]. At this stage, the two modalities namely the fluorescent and the brightfield microscopy images show opposite intensity characteristics where intensity levels of nuclei in fluorescent histopathology images are much lower than that of brightfield histopathology images . Hence, the following color inversion process is applied. For a random pixel,  $x_{ij}$  in an image,  $X$  with mean intensity level,  $\bar{X}$ , the inversion process can be expressed as:

$$x_{ij} = \begin{cases} 255 - x_{ij} & \text{if } \bar{X} > 127 \\ x_{ij} & \text{if } \bar{X} \leq 127 \end{cases} \quad (3.1)$$

At this stage, all the images from two different modalities show similar intensity level and color statistics as shown in Fig.3.1.

### 3.1.2 Proposed Contourlet Driven Attention Network

Here, a contourlet driven attention network, namely ConDANet, is proposed which utilizes directional multi-resolution signal to inject additional fine scale detail features to the network. The traditional U-net like architecture is utilized as the baseline network of this proposed method which comprises of an encoding part and a decoding part. In this traditional encoder-decoder architecture, two separate modifications are introduced as follows:

- Our primary objective is to inject multi-level directional feature to guide the spatial level information found from the traditional U-net architecture. The motivation of this approach is to enrich the spatial information with the edge-aware information to make the model more capable in capturing the faded boundary region of the nuclei structure. To serve this purpose, contourlet transform is exploited to produce edge aware transformed signal. However, instead of working with the contourlet transformed images as the input of the network, we propose an attention mechanism which takes the spatial domain information at different levels of encoder-decoder architecture as the input and the contourlet transformed images as the controlling signal which assists the model to focus on the edge portion, eventually captures the fine edge details of the network.
- The subsequent downsampling operation in the encoder part of the network can

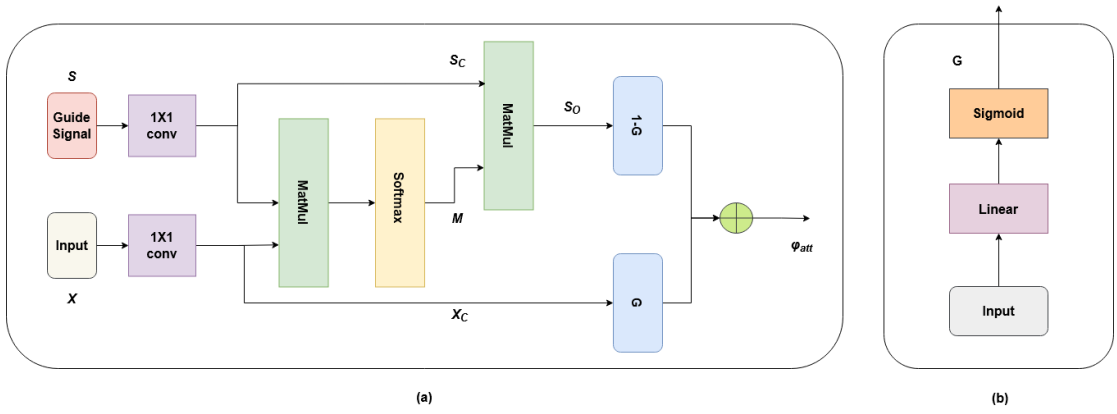


Figure 3.3: Visualization of contourlet driven attention mechanism.

have an unwanted impact in the spatial level information, especially on the objects present in a small spatial domain. Due to the capability of wavelet transform in preserving the original texture of the image in a small spatial domain, a wavelet pooling is adopted instead of the traditional max or average pooling operation in this method.

The detailed architecture of the proposed network is shown in Fig.3.2.

### 3.1.2.1 Wavelet Based Pooling Strategy

Our backbone network is an encoder.-decoder network similar to U-net like architectures. However, instead of using the traditional up and down-sampling operation in the encoder-decoder part, the 2D discrete wavelet transform (DWT) and the inverse DWT (IDWT) are utilized as the sampling method. Traditionally, the well-known encoder-decoder architectures, such as U-net and Unet++, perform the down-sampling operation by the max or average pooling strategy. However, these down-sampling operations usually neglect the Nyquist sampling theorem resulting in aliasing of different data frequency components of signal. As a result, data details may lost in these operations. On the other hand, for up-sampling operation in the decoder part, the max-unpooling or bilinear interpolation approach are widely used. Although they are applied to recover the data details, that cannot be reconstructed completely. Wavelet based sampling strategy can overcome this problem by preserving the spectral content in the encoder unit and reconstruct the original content with a minimal loss in the decoder part. First, the input image is passed through a sequence of encoding operation with convolution filter, batch normalization layer, and ReLu activation operation. Next, at each encoder and decoder stages, the down-sampling and up-sampling operations are replaced by the DWT and IDWT operations. The DWT decomposes the signal into three high frequency components and one low frequency components as follows:

$$X_{ij} = (f_{ij} * X)_{\downarrow 2}, \quad (3.2)$$

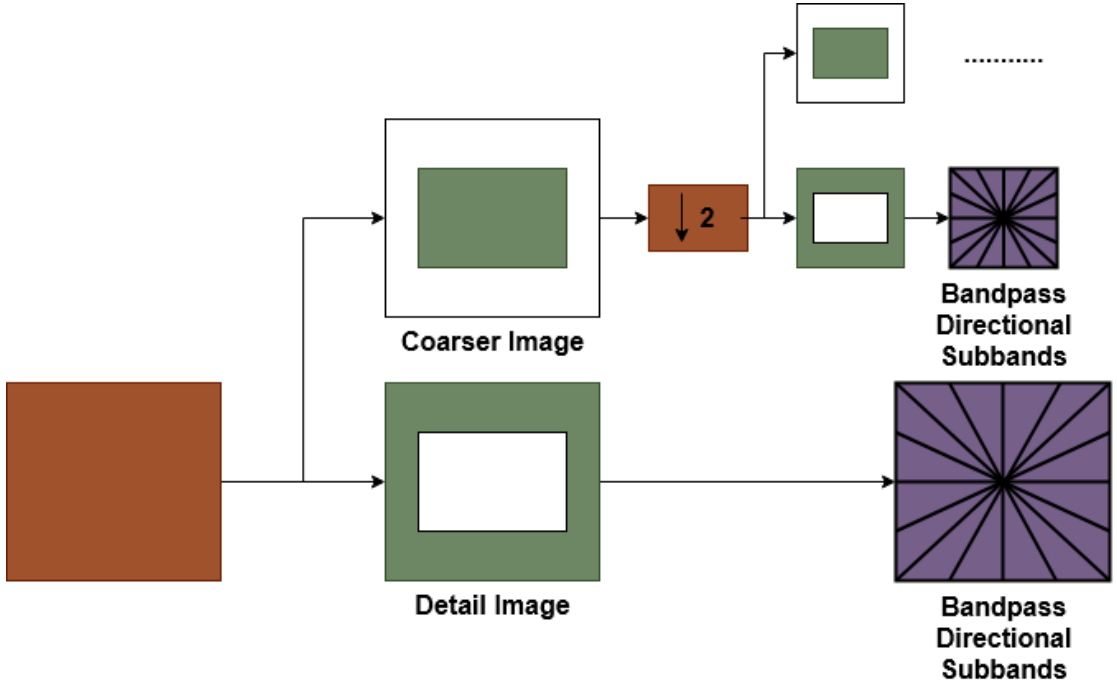


Figure 3.4: Contourlet filter bank.

where  $*$  denotes convolution operation,  $(\cdot)_{\downarrow 2}$  represents the 2 times downsampling operation, and  $f_{ij}$  represents one of the four filters (high and low pass filters) to be used.  $X_{ij}$  is the corresponding filtered output where, Subscript ‘ $ij$ ’ refers to filter type i.e.  $ll$  for low pass filter, and  $lh$ ,  $hl$ , and  $hh$  for three high pass filters. The low frequency component is transmitted into the subsequent layers of the network and the high frequency components are combined with the decoding stage outputs to recover the feature map using IDWT operation which is as follows:

$$X = \sum f_{ij} * X_{\uparrow 2} \quad (3.3)$$

Here,  $(\cdot)_{\uparrow 2}$  represents the 2 times upsampling operation.

### 3.1.2.2 Contourlet Driven Attention Mechanism (ConDAM)

Utilization of DWT operation instead of the max or average pooling operation in the encoder-decoder architecture cannot always ensure the prevention of loss of information, especially in edge regions where the image quality is not prominent. It may create unwanted noise and hamper the segmentation performance. In order to provide edge level information to the network, an attention mechanism is required which is driven by a controlling signal containing the edge aware information at multiple directions. To accomplish this objective, a contourlet transform driven attention mechanism is proposed here. The proposed attention mechanism is depicted in Fig.3.4 and described in the following subsections.

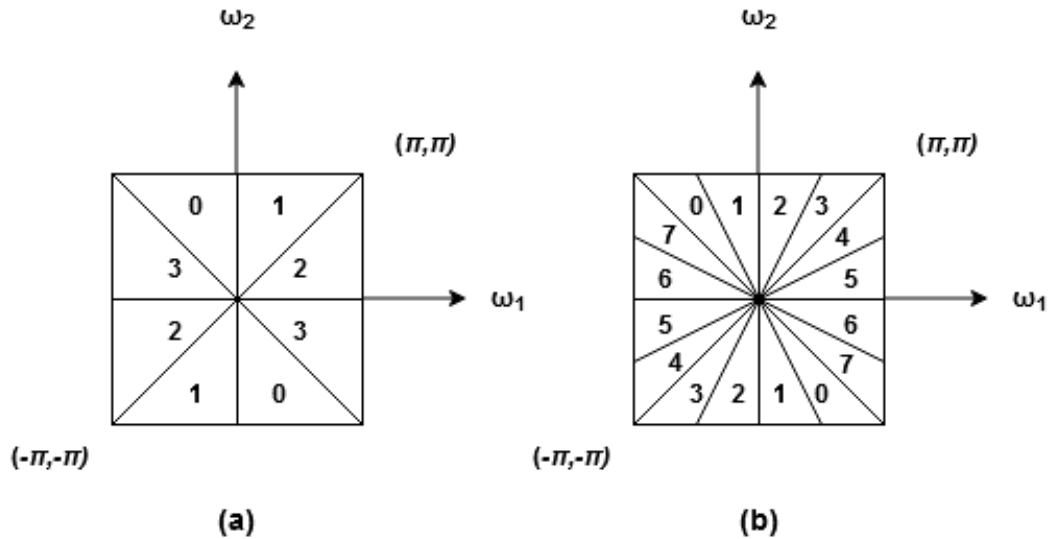


Figure 3.5: Directional filter bank. Frequency partitioning where level (a)  $l = 2$  and there are  $2^2 = 4$  real wedge-shaped frequency bands, and (b)  $l = 3$  and there are  $2^3 = 8$  real wedge-shaped frequency bands.

**Controlling Signal** The contourlet transform is an effective method which can realize multi-resolution and direction decomposition flexibly and grasp essential information from a two dimensional image [89–91]. It is well suited for multiscale edge-based image analysis due to its superior performance in representing the image salient features such as edges, lines, curves, and contours [92–94]. The contourlet operation consists of two separate filtering techniques. The first one is a Laplacian pyramid filter which can extract the multi-resolution features of an image and the other one is the directional filter banks that can decompose the image into multi directional components. A typical contourlet filter with a combination of a Laplacian pyramid (LP) and a directional filter bank (DFB) is shown in Fig.3.4. At first, the Laplacian Pyramid decomposition at each level generates a lowpass filtered coarser image and a bandpass image resulting from the difference between the original and the coarse image, which is termed as detail image. The LP decomposition at each level generates a downsampled lowpass version of the original and the difference between the original and the prediction, resulting in a bandpass image. Fig. 3.7 depicts this decomposition process, where  $H$  and  $G$  are called (lowpass) analysis and synthesis filters, respectively, and  $M$  is the sampling matrix. The process can be iterated on the coarse (downsampled lowpass) signal. Note that in multidimensional filter banks, sampling is represented by sampling matrices; for example, downsampling  $x[n]$  by  $M$  yields  $x_d[n] = x[Mn]$ , where  $M$  is an integer matrix [95]. Bandpass images from the LP are fed into a DFB so that directional information can be captured. The scheme can be iterated on the coarse image. The combined result is a double iterated filter bank structure, named contourlet filter bank, which decomposes images into directional subbands at multiple scales. Next, the DFB is applied on the detail image to capture the directional information of the edge signal. DFB is basically a

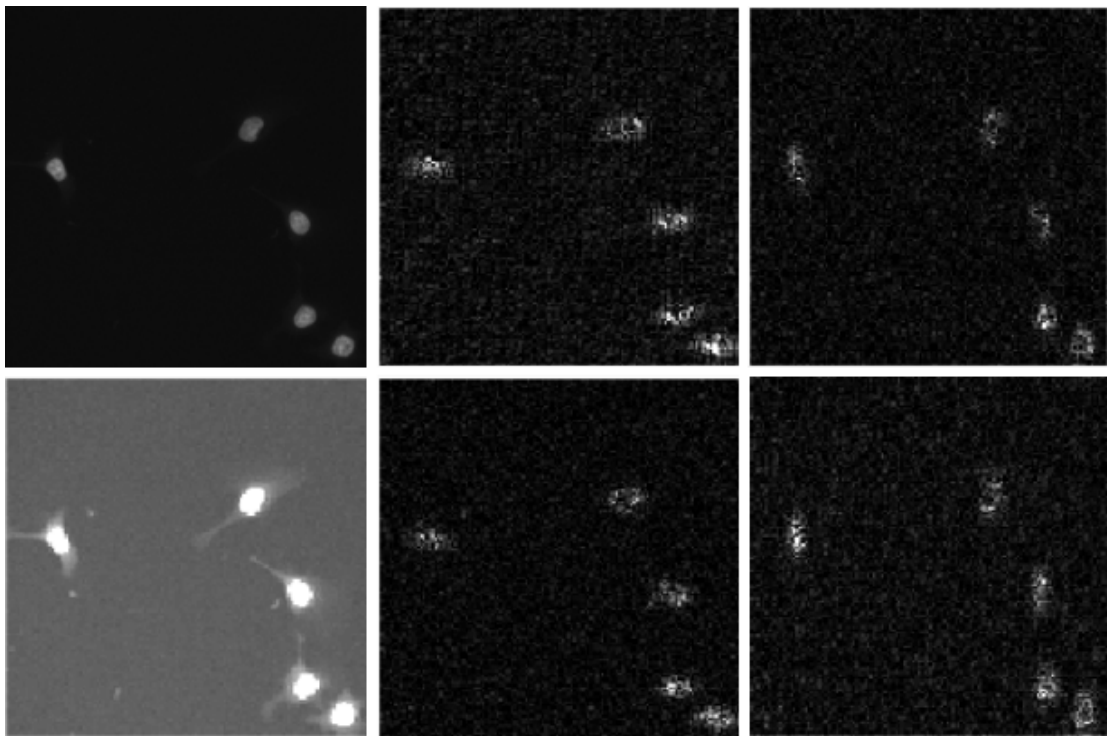


Figure 3.6: Example of the contourlet transform on the histopathology image. For clear visualization, each image is only decomposed into one pyramidal level, which are then decomposed into four directional subbands. Small coefficients are shown in black while large coefficients are shown in white.

$l$ -level binary decomposition of a signal that leads to a  $2^l$  subbands with wedge-shaped frequency partitioning as shown in Fig.3.5 [96]. In our case, two level directional filter bank decomposition is followed which creates one coarser image and four band pass filtered directional image components as shown in Fig.3.6.

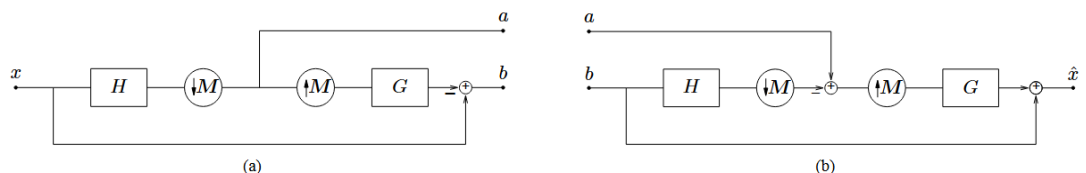


Figure 3.7: Laplacian pyramid. (a) One level of decomposition. The outputs are a coarse approximation  $a[n]$  and a difference  $b[n]$  between the original signal and the prediction. (b) The new reconstruction scheme for the Laplacian pyramid

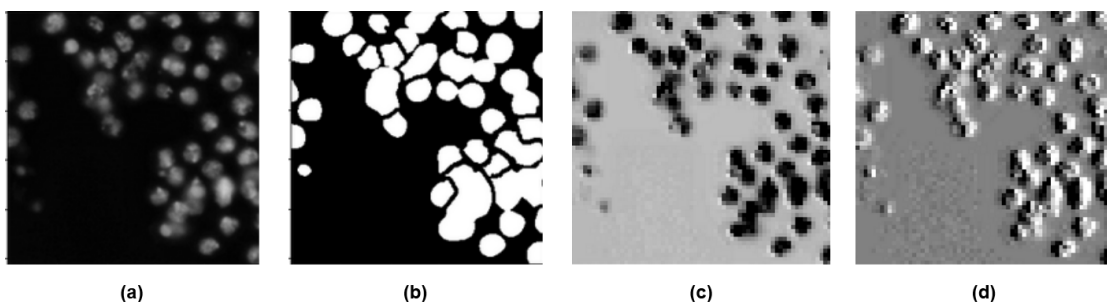


Figure 3.8: Effect of Contourlet driven attention module (CDAM). (a) Original image (b) Ground truth (b) feature map without CDAM (d) feature map with CDAM

**Driven Attention Module** An attention mechanism is proposed to incorporate the contourlet transformed bandpass directional information ( $S$ ) of image data with the low level semantics of the decoder as shown in Fig.3.3. Initially, the directional components coming from the countourlet transform operation are passed through a  $1 \times 1$  convolutional operation and concatenated with each other to form the controlling signal. On the other hand, the decoder level information is utilized as the input of our attention block and also linearly transformed using a  $1 \times 1$  convolution block. The resultant controlling and input signal are then multiplied and passed through a softmax activation function to create a new feature map  $M$ . The softmax operation is performed to create the probability distribution over the controlling signal. The resultant feature map is further multiplied with the convoluted controlling signal ( $S_C$ ) to extract the prominent regions found from the controlling signal. The overall operation can be summarized by the following equation:

$$S_O = \text{Softmax}(\phi_1(X)\phi_2(S)^T)\phi_2(S) \quad (3.4)$$

Here,  $\phi_1$  and  $\phi_2$  are the  $1 \times 1$  convolution operation. At this stage, to regulate the degree of enhancement of input signal, a gating mechanism is proposed to prevent information overload. The gating mechanism,  $G$  comprises of a sigmoid activation function followed by a linear layer. The gating mechanism creates the final feature map  $\phi_{att}$  which can be represented as follows:

$$\phi_{att} = G.X + (1 - G).S_O \quad (3.5)$$

Since, the problem is a binary class segmentation problem, the conventional binary cross-entropy loss is utilized as the loss function of this network which is defined by the following equation:

$$\mathcal{L} = \frac{1}{N} \sum_{j=1}^N p_j \log y_j + (1 - p_j) \log(1 - y_j) \quad (3.6)$$

where  $p_j$  denotes the predicted probability value of  $j^{th}$  sample,  $y_j$  represents the corresponding label, and  $N$  is the total number of samples.

## 3.2 Results and Discussion

The datasets used for the evaluation of the proposed method are described briefly in this section. Next, the simulation details are discussed thoroughly. Finally, the performance of the modules of the proposed method and the method itself are presented and discussed compared to the other state-of-the-art methods.

### 3.2.1 Datasets

Three publicly available datasets are utilized in this work to verify the functioning of the proposed model.

#### 3.2.1.1 Data Science Bowl (DSB-2018)

The first one is released as a Kaggle competition dataset known as data science bowl (DSB-2018) [78]. The training set comprises of total 670 images microscopy images that were acquired under different conditions and vary in cell type, shape, magnification, illumination status, and imaging modality (brightfield and fluorescence). The distribution of nuclei particle size varies from 21 to 1037 pixels. The data can be found in [97]

#### 3.2.1.2 MoNuSeg Dataset

Multi-organ nuclei segmentation challenge (MoNuSeg) [79] is selected as the second dataset evaluated in this work. It contains a diverse set of Hematoxylin-Eosin (HE) stained tissue images where 30 images with around 22,000 nuclear boundary annotations is available for training and 14 images with 7,000 nuclear boundary annotations are available for testing purpose. The histopathology images of this dataset are collected from wide range of patients and their resolution is  $1000 \times 1000$ . The data can be found in [98]

#### 3.2.1.3 TNBC Dataset

The final dataset is the triple negative breast cancer (TNBC) dataset [70], which includes 50 images with a total of 4,022 annotated cells, including normal epithelial and myoepithelial breast cells (localized in ducts and lobules), invasive carcinomatous cells, fibroblasts, endothelial cells, adipocytes, macrophages, and inflammatory cells (lymphocytes and plasmocytes). The images are  $500 \times 500$  in size. The data can be found in [99]

### 3.2.2 Training and Implementation Details

In order to match the image resolution of the datasets used in this method with the input size of our proposed model, some modifications are done on those three datasets. The images in the DSB-2018 dataset are resized into  $256 \times 256$  to feed into the network. Since, the image size of the MoNuSeg dataset is  $1000 \times 1000$ , each image is zero padded

to convert the image size to  $1024 \times 1024$  to make it divisible by the input size of the network. Then each of these images are cropped to 16 non-overlapping images with size of  $256 \times 256$  and finally these image patches are fed into the network. Similarly, for TNBC dataset, the images are converted to  $512 \times 512$  and then four image patches with size of  $256 \times 256$  are obtained. In the preprocessing step, two commonly used augmentation techniques —rotation and flipping, are followed in our methods. Following these two techniques, the original images are rotated by  $30^\circ$  and  $60^\circ$ , and flipped horizontally, respectively. To make our proposed model perform better for the nuclei segmentation task, different hyperparameters are chosen based on empirical analysis. The network is trained for 70 epochs. To optimize the network, Adam optimizer with a learning rate of 0.005 is used and a decaying rate of 0.99 after each 10 epochs is chosen to reduce the loss of the network. The model is implemented on the Google cloud platform with NVIDIA P-100 GPU as the hardware accelerator.

To assess the model performance on the datasets, several performance evaluation metrics are utilized. Since, it is a segmentation task, we have selected the most commonly used metrics of the typical segmentation tasks which are as follows [100]:

$$\text{Dice} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}} \quad (3.7)$$

$$\text{IOU} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \quad (3.8)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (3.9)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (3.10)$$

where, TP, FP, and FN represent true positive, false positive, and false negative predictions, respectively. Since, only the training data are available for the DSB-2018 and TNBC datasets, 80% of the training images are considered for training stage and the rest are equally divided for validation and testing purpose. Moreover, the result on the test set of MoNuSeg are also reported and compared with the state of the art segmentation networks to validate the effectiveness of the proposed network.

### 3.2.3 Ablation Study of the proposed nuclei segmentation Network

To represent the individual performance contributed by different components of our proposed model, an ablation study is performed under different settings of the method. The summary of this study is depicted in Table.3.1. Here, we utilized two standard performance evaluation metric, dice coefficient and intersection over union (IOU) to represent the improvement contributed by each module. Compared to the baseline,



dice score improvement between 2.28-5.92% is observed over the three dataset while using preprocessing steps. The IOU performance also improves between 4.7-6.9%. This performance improvement indicates the effectiveness of the preprocessing steps in preparing the data to feed into the model and removing the inconsistencies between different modalities of the histopathology images. Moreover, having the wavelet pooling module in parallel brings slightly better results (1.25% improvement) in DSB-2018, while obtains a better improvement (2.17-2.61%) in MoNuSeg and TNBC dataset. The reason for this scenario is found by observing the complex cases of the three datasets. The overlapping nuclei and the small sized nucleus cases are less likely in DSB-2018 dataset compared to the other two. That's why the impact of the wavelet pooling is less promising in the DSB-2018. Next, we have utilized the proposed CBGA module without WP approach and find a 2.98-3.7% improvement in dice score is achieved. Finally, integrating the ConDaM with the wavelet pooling module, the best results are obtained in all three datasets with an improvement between 4.36-4.74% in terms of dice score and 4.17-8.91% in terms IOU compared to the case when ConDAM module is not applied. Here, while wavelet pooling works as a simple pooling strategy on the convoluted signal in intermediate layers, the ConDAM injects the additional wavelet domain information to the network and works as an effective tool to identify the contour around the nucleus region.

### 3.2.4 Comparison to the State-of-the-art Methods

In order to corroborate the effectiveness of the proposed method in segmenting nuclei region compared to the other popular deep neural frameworks, the experimental results of several state-of-the-art networks are reported for DSB-2018, MoNuSeg, and TNBC datasets in Table.4.1, Table.4.2, and Table.4.3, respectively. In the DSB-2018 dataset, compared to the other encoder-decoder networks that were used for medical image segmentation purpose- i.e., U-net, FPN, and Segnet - our network achieves an overall performance improvement of 2.48%, 2%, and 17.8% in terms of dice score, respectively. The huge difference in performance improvement is because of the liability of Segnet in rounding edges which is already proved in [101]. On the other hand, compared to the attention based network like Attention U-net, a 2% improvement is observed. This improvement proves the superiority of contourlet driven attention mechanism instead of the attention of the encoding signal. Finally, we also compared our model with some of the recent context aware deep learning frameworks applied for the nuclei segmentation task i.e., CENet and nucleAIzer- in which context information is utilized in order to elevate the performance. After comparing the result with our technique, an average of 1.9-2.28% improvement is observed. In addition to dice score comparison, three other standard performance evaluation metrics are also utilized which provides the sim-

Table 3.1: Ablation study of the proposed network on three datasets.

Pre-processing	wavelet pooling	ConDAM	DSB-2018		MoNuSeg		TNBC	
			Dice	IOU	Dice	IOU	Dice	IOU
X	X	X	82.46	73.52	73.08	58.92	66.51	51.41
✓	X	X	84.34	76.97	76.02	62.97	70.45	54.83
✓	✓	X	85.40	77.57	78.01	65.09	71.98	55.55
✓	X	✓	87.95	80.21	81.02	69.92	74.65	56.12
✓	✓	✓	88.91	80.81	81.71	70.89	75.12	56.87

Table 3.2: Performance comparison of different architectures in the DSB-2018 dataset (8:1:1 split setup).

Network	Dice(%)	IOU(%)	Precision(%)	Recall(%)
Segnet [83]	75.48	62.36	84.08	68.47
U-Net [42]	86.75	76.97	85.55	87.98
nucleAIzer [84]	86.92	75.89	82.32	90.91
Attention U-net [51]	87.30	77.51	79.97	94.81
CENet [85]	87.25	76.50	79.21	91.90
FPN [45]	87.34	77.97	80.34	95.67
ConDANet	<b>88.91</b>	<b>80.81</b>	<b>82.11</b>	<b>96.73</b>

ilar observation as well. The similar scenario is observed in other two datasets as well which is summarized in Table.4.2 and Table.4.3.

### 3.2.5 Qualitative evaluation

In addition to quantitative performance measurement, a qualitative analysis of the proposed method is conducted by visual representation of the segmented nucleus obtained by using different state-of-the-art methods along with our proposed approach. Fig.4.6 displays the segmentation performance on three separate cases obtained from three datasets. In spite of the close results found in the quantitative study, some interesting observations are found from the visual expectation. Here, we considered three different network architectures for the comparison purpose which are —the basic encoder-decoder architecture U-net, the attention based network attention U-net, and the context

Table 3.3: Performance comparison of different architectures in the MoNuSeg dataset (8:1:1 split setup).

Network	Dice(%)	IOU(%)	Precision(%)	Recall(%)
FCN [87]	79.77	62.45	73.21	87.46
Unet [42]	74.67	60.89	71.24	78.44
DIST [70]	77.31	63.77	75.31	79.41
Segnet [83]	77.44	63.3	73.91	81.32
Attention U-net [51]	78.67	66.51	76.71	82.33
CENet [85]	79.35	65.86	77.86	81.55
FPN [45]	80.02	66.89	74.98	82.19
ConDANet	<b>81.71</b>	<b>70.89</b>	<b>79.95</b>	<b>88.33</b>

Table 3.4: Performance comparison of different architectures in the TNBC dataset (8:1:1 split setup).

Network	Dice(%)	IOU(%)	Precision(%)	Recall(%)
U-Net [42]	68.61	52.92	65.94	72.54
Mask-RCNN [86]	70.54	52.97	66.27	75.39
DIST [70]	70.51	56.34	66.82	74.63
Micro-Net [47]	71.23	53.71	66.52	76.61
Attention U-net [51]	71.43	54.21	70.01	76.03
FCN [87]	72.67	50.62	67.53	78.53
CENet [83]	73.88	54.93	71.11	76.71
ConDANet	<b>75.12</b>	<b>56.87</b>	<b>72.02</b>	<b>79.13</b>

aware network CENet. First, we consider a comparatively simple case from DSB-2018 dataset with fairly separable and no clustered or adjacent nuclei. Since the edge characteristics are distinguishable in this particular case, the three networks perform quite well in this segmentation task except for some false positive and negative cases around the contour of some nuclei. Afterward, two comparatively difficult scenarios are considered from MoNuSeg and TNBC dataset. These two cases contain several challenging issues, for example, the presence of small nuclei with confounding edge characteristics and the clustered nuclei regions. In the case from MoNuSeg dataset (second row of Fig.4.6), the proposed method is found very effective in identifying the edge characteristics and small regions compared to the best performing method CENet. On the other hand, the other two methods- attention U-net and U-net architecture barely distinguish the adjacent nuclei separation and many false positive cases which are wrongly identified as the nuclei regions. The most possible reason for this finding is the non-accountability of giving attention to the contour regions and the subsequent down-sampling operations of traditional pooling operation in both of these architectures. The same scenario is happened in the case from TNBC dataset (third row of Fig.4.6), where a fairly clear picture of segmentation performance of the proposed method along with the other state-of-the-art architectures is observed. In this case, the U-net performs badly in identifying the edge regions of the nucleus thus creating false negative cases. The other two architectures perform in a similar manner with less portion of false negative cases because of their context aware nature and attention mechanism. However, these methods mistakenly identify some of the regions as nucleus region whereas our method successfully discards those regions because of its capability of identifying boundaries of the object of interest. The computational efficiency of different networks are summarized in Table 3.5. From the computational perspective it can be concluded that the network possesses less number of parameters and performs faster with respect to the other network. The network takes equal time to infer with respect to the Attention U-net with almost 4.84 times memory efficiency.

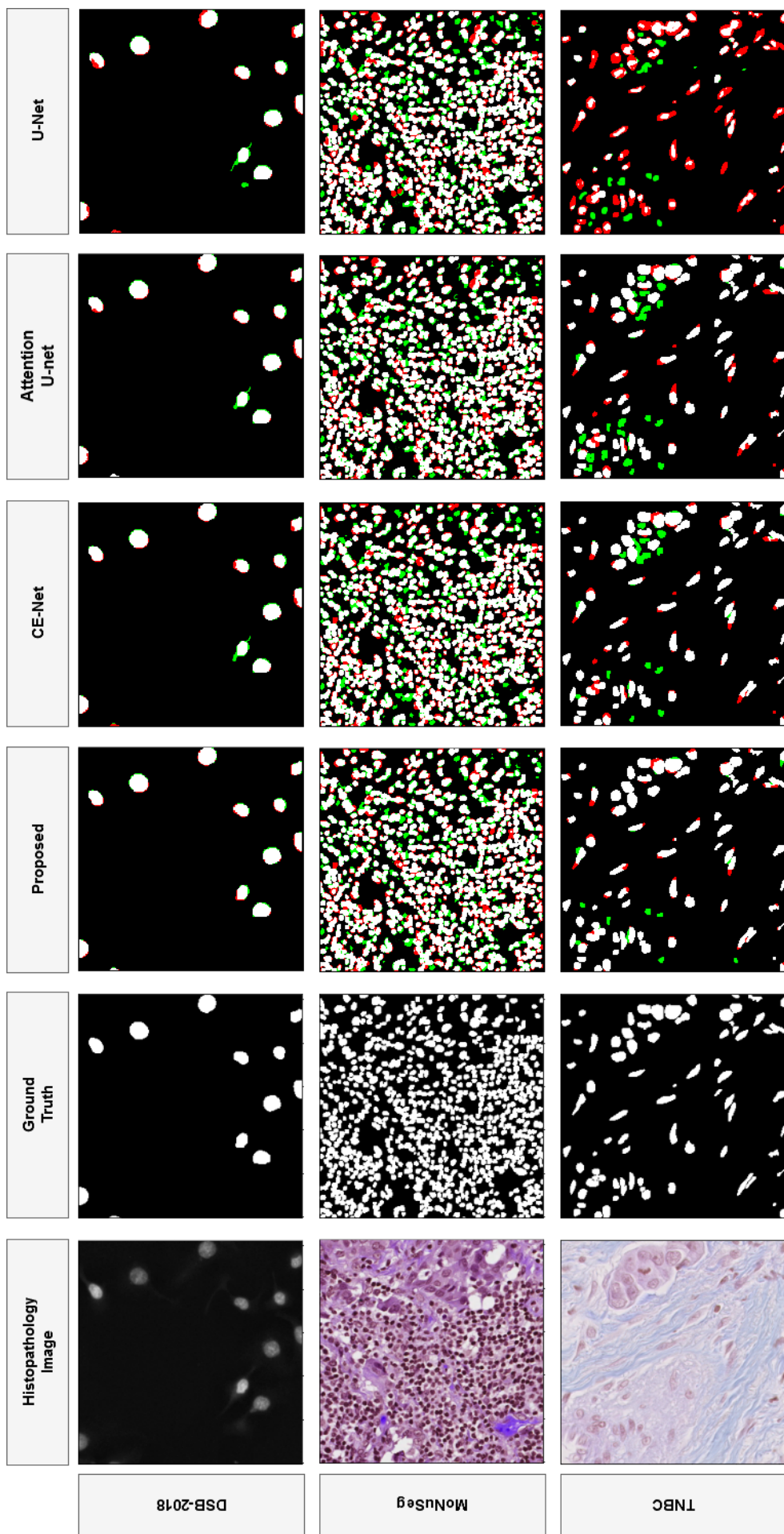


Figure 3.9: Visualization of the nuclei segmentation performance of some of the state-of-the-art networks in three datasets. The inputs of the first two rows correspond to DSB-2018 dataset, the next two rows correspond to MoNuSeg dataset, and the final two rows correspond to TNBC dataset. Here, the 'green' color is used to represent false positive cases and the 'red' color is used to represent false negative cases.

Table 3.5: Computational Efficiency Analysis of Numerous Architectures along with the Performance of the proposed method.

Methods	Number of Parameters	Inference Time
Unet	31.0	0.1
MicroNet	184	0.18
DIST	8	0.12
SegNet	14.7	0.17
Attention U-net	8.727	0.09
CENet	38.8	0.12
FPN	8.4	0.137
ConDANet	1.8	0.09

### 3.3 Conclusion

In this Chapter, a contourlet driven attention based deep neural network architecture is proposed to address the nuclei segmentation task in histopathology images. The proposed architecture addresses two major issues present in the nuclei characteristics. The first one is the small sized nuclei region and the other one is the blurred edge region. Both of them are challenging situations because of their presence in low dimensional space for which they can be excluded during the segmentation operation. To handle these issues, two necessary modifications are conducted in the traditional encoder-decoder architecture. First, a contourlet guided attention module introduced which guides the convoluted output signal by the edge aware directional filtered output from contourlet transform. The additional attention to the contour of the nuclei regions has a significant impact in the performance which is observed in our analysis. Additionally, it is found that the wavelet pooling based sampling strategy can solve the problem of identifying the small-sized nuclei regions by its effective reconstruction capability in the decoder part which can not be possible in the sampling technique like max-pooling or average-pooling operation. The qualitative and quantitative analysis of the proposed method with other modern deep learning architectures including the attention based network and context aware deep learning architectures clearly demonstrates the superior performance of this network.

## Chapter 4

# Proposed Cell Nuclei Segmentation Scheme Based on Boundary Aware Wavelet Guided Network

In this chapter, a boundary aware wavelet guided network, termed as BAWGNet is proposed. Prior to applying the data in the network, a preprocessing step is proposed to handle the lack of data diversity, and the variation in color statistics in different image modalities. Unlike the skip connection strategy of conventional encoder-decoder architecture, we propose an aggregation unit that introduces a guided attention mechanism to combine the encoder and decoder information and also generates boundary Aware information. In the guided attention unit, the spatial domain information at different encoder-decoder levels is further processed under the guidance of the wavelet domain information extracted from discrete wavelet transform. Moreover, the proposed boundary aware unit inside the aggregation unit helps preserve the boundary information of the nucleus by incorporating a separate loss function which gives more attention to the nuclei edge regions. The proposed wavelet guided attention unit and boundary aware unit jointly leverage to identify small nucleus regions and separate the boundary of adjacent individual nuclei very efficiently. Finally, Extensive experimentations have been done over three publicly available datasets containing a large number of subjects with complex nucleus orientations in different organs and wide varieties of patients.

### 4.1 Proposed Cell Nuclei Segmentation Scheme

In this section, the proposed methodology is described in details including the preprocessing steps, the network architecture, and the hybrid loss function.

### 4.1.1 Preprocessing

Data generalisability is a major concern in convolutional neural network specifically when the image data is collected from multiple sources. Since different microscopy techniques are used to collect the histopathology images, it is important to make all the data consistent with each other and prepare them for the network. In the cell imaging data, two most commonly used acquisition techniques are fluorescent and brightfield microscopy which are used to prepare the live-cell images to capture the single molecules and whole cell structure. Due to the difference in their acquisition procedures, the images from these two modalities vary in their intensity level, color statistics and contrast information. Hence, it is important to process them and make them prepared for computational analysis. In order to do this, a set of preprocessing steps are adopted in this method. At first, all the microscopy images acquired from different acquisition techniques are converted into LAB color space to maintain the consistency among the data. Next, the images are contrast enhanced by applying the contrast level adaptive histogram equalization (CLAHE) technique to enhance the visual appearance and contrast of the nucleus structure in the histopathology images [75]. Up to this position, the Fluorescent and Brightfield microscopy images show exactly opposite intensity characteristics, where the intensity of the nuclei region is darker in Fluorescent microscopy images and brighter in the Brightfield images. To solve this issue, a color inversion process is applied which leaves the Brightfield images as before but opposes the color characteristics of the Fluorescent images. The overall process can be summarised by the following equation:

$$i_{mn} = \begin{cases} 255 - i_{mn} & \text{if } \bar{I} > 127 \\ i_{mn} & \text{if } \bar{I} \leq 127 \end{cases} \quad (4.1)$$

where,  $i_{mn}$  is the intensity of the (m,n) spatial position of image,  $I$  and  $\bar{I}$  is the mean intensity.

### 4.1.2 Proposed Boundary Aware Wavelet Guided Network

In this paper, a boundary Aware wavelet guided network is proposed to segment the nuclei in the cell structure using histopathology images. U-net architecture is used as the backbone of this network that consists of an encoding part, a decoding part. The structures of the encoding and decoding part are similar but converse in their internal arrangements. In order to represent the basic structure of this network, a simplified diagram is shown in Fig. 4.1. The proposed structure consists of three basic operations as follows:

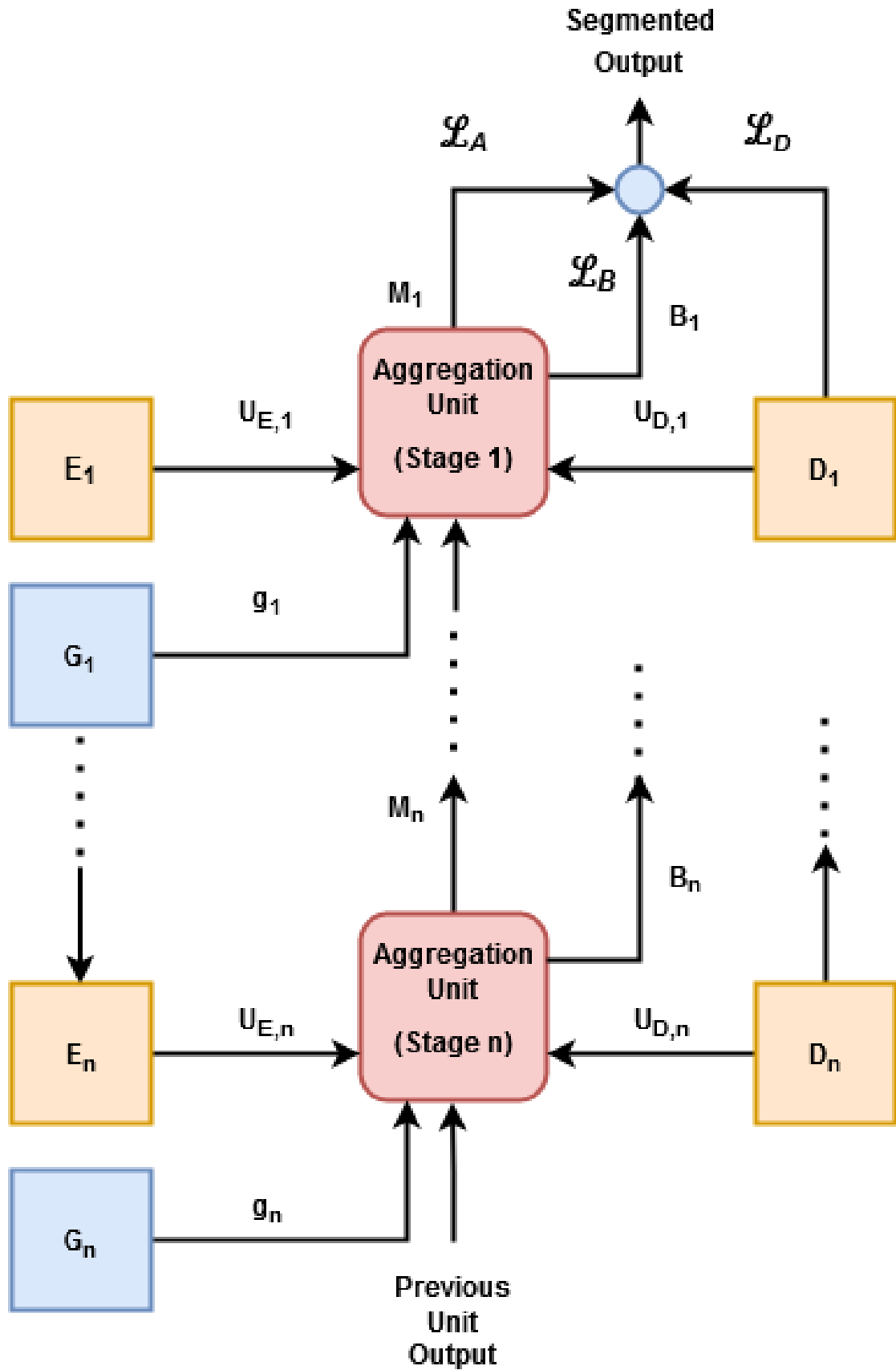


Figure 4.1: A simplified representation of the proposed architecture. Here,  $E_i$ ,  $D_i$ , and  $G_i$  correspond to the encoder, decoder and guide unit at  $i^{th}$  stage of the network, respectively.



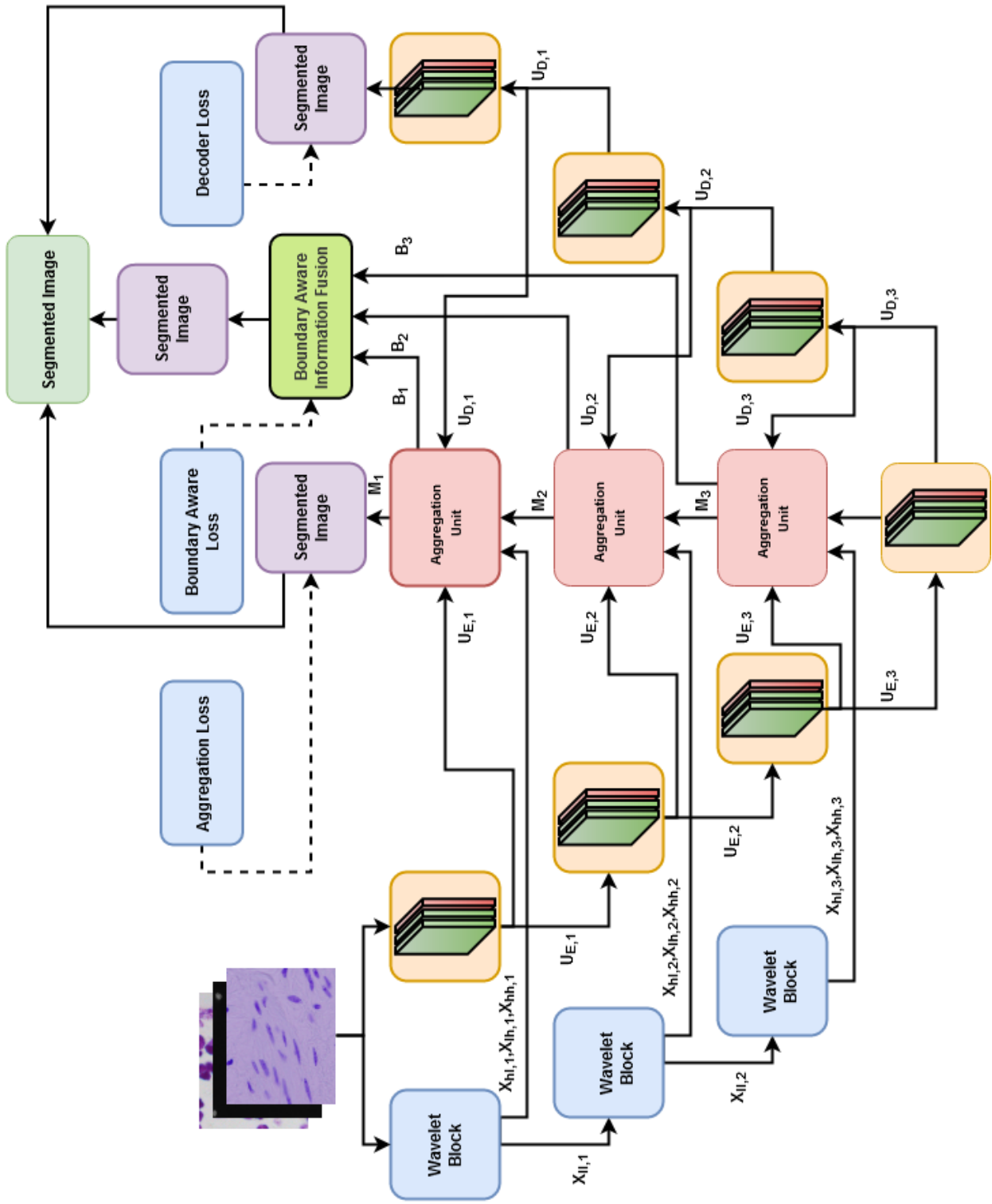


Figure 4.2: Graphical overview of the proposed wavelet attention U-net. Here, three separate loss functions are utilized to optimize the overall network and three individual segmented results are considered to get the final segmentation result. Boundary Aware Information Fusion' unit (yellow block) is shown in detail in Fig.4.3

- The primary objective is to utilize spatial information obtained at different stages of encoder and decoder. We developed a processing block called aggregation unit (AU) which is the heart of our proposed network. At a particular stage  $i$  ( $i \in 1, 2, 3$ ) of the encoder-decoder architecture, it incorporates the spatial domain information ( $U_{E,i}$  and  $U_{D,i}$ ) present in the cell structure from encoder and decoder block ( $E_i$  and  $D_i$ ). Moreover, a decoder loss function,  $\mathcal{L}_D$  is defined to produce a segmented output from this part.
- Instead of directly utilize the encoder decoder information, we followed an attention mechanism approach. In this regard, our idea is to incorporate wavelet domain information along with spatial domain information in the attention mechanism. Hence, a guide signal,  $g$  containing frequency level information is utilized to guide the spatial information through an attention mechanism inside the AU. AU at the final stage produces another segmented output,  $M$  by minimizing a loss function  $\mathcal{L}_A$ . It is to be mentioned that we propose to use wavelet transform to get  $g$ .
- Finally, a boundary Aware unit is proposed inside the AU, which utilizes the morphological features to preserve the boundary or edge information of the nucleus region and produces a boundary Aware map  $B$ . By combining boundary information from AUs at different stages, another segmented output is produced. A boundary Aware loss function,  $\mathcal{L}_B$  is also incorporated with this part.

The detailed architecture of our proposed network is shown in Fig. 4.2. In the backbone encoder-decoder structure, both the encoder and decoder have three blocks and the outputs at each stage are denoted as  $U_{E,i}$  and  $U_{D,i}$ , respectively; where  $i \in [1, 2, 3]$ . Each block contains three convolutional layers followed by batch normalization layer and ‘ReLU’ activation function. Convolutional layers in each block (same encoder and decoder level  $i$ ) have  $3 \times 3$  kernel size with number of kernel equal to  $2^{i-1} \times 32$ . A  $2 \times 2$  max pooling layer is placed after each convolutional layers.

As the previously mentioned guide signal, we propose to utilize discrete wavelet transform (DWT) which provides relevant frequency information at different levels of decomposition. Another advantage of DWT is that at every level of DWT, the change in dimension of the input level exactly corresponds to the change occurs at every stage of the encoder. Here in our case, the 2D- DWT decomposes the 2D histopathology image ( $X$ ) as follows:

$$X_{ab} = (f_{ab} * X)_{\downarrow 2}, \quad (4.2)$$

where  $*$  denotes convolution operation,  $(\cdot)_{\downarrow 2}$  represents the 2 times downsampling operation, and  $f_{ab}$  represents one of the four filters to be used and  $X_{ab}$  represents the corresponding filtered output. Subscript ‘ $ab$ ’ refers to filter type i.e.  $ll$  for low pass

filter, and  $lh$ ,  $hl$ , and  $hh$  for three high pass filters.

The network consists of three AU blocks that receives the feature maps from the encoder and decoder blocks of the backbone network at the same level  $i$ , the high frequency components ( $X_{lh}$ ,  $X_{hl}$ , and  $X_{hh}$ ) at a particular level of the DWT, and the output ( $M_{i-1}$ ) from the previous AU block. Each of these main stream blocks consists of two separate units. The first unit is the wavelet guided attention unit (WGAU) to guide the signal from encoder-decoder architecture by the wavelet information into the network and the other is boundary Aware unit (BAU) to learn the morphological features. In what follows these major units are explained in subsequent sections.

#### 4.1.2.1 Wavelet Guided Attention Unit (WGAU)

The wavelet guided attention unit is separately shown in detail in Fig.4.3a. It is utilized to fuse the high frequency components of image data with the low level textures of encoder and high level semantics of decoder of the backbone network. Initially, a  $1 \times 1$  convolutional block is used to extract features individually from the high frequency components of 2D-DWT ( $X_{lh,i}$ ,  $X_{hl,i}$ ,  $X_{hh,i}$ ). The resultant features are concatenated to create a new feature map,  $g_i$ . Parallely, the feature maps from encoder and decoder block of the backbone network are incorporated to capture the textures and semantics representations which is represented as  $U_{c,i}$ . Afterward, this feature map is integrated to the previous main stream block output,  $M_{i-1}$ . Prior to this integration,  $M_{i-1}$  is passed through a  $2 \times 2$  upsampling block to recover the resolution. Also, a dilated convolutional layer with dilation rate 2 is applied to capture the global information. The two simultaneous operation produces  $M_{i,dil}$ . Afterwards,  $M_{i,dil}$  is added with  $U_{c,i}$  to produce the integrated output  $X$ . Finally,  $X$  and  $g$  are applied as the input of the attention block.

In order to guide the spatial level information with the extracted wavelet domain information ( $g$ ), an attention block is designed in the proposed method which is shown in Fig. 4.4. In this block, the wavelet component ( $g$ ) is used as the gating vector which guides the network to leverage the salient region of the given images. It also contains the contextual information that can cut off the lower level feature responses in the natural image classification task. First, the gated vector ( $g$ ) and the input tensor ( $X$ ) are linearly transformed using a  $1 \times 1$  convolution block and added together. The convolution operation allows to extract the most promising features from them and the addition operation further enhance the features those are found relevant in both spatial and frequency domain. The resultant output is then passed through a ReLU activation function  $\sigma_1$ , to prevent vanishing gradient problem followed by a  $1 \times 1$  convolution layer  $\phi$  to find the gating coefficient,  $g_{att,c}$ . An additive attention is used to obtain this gating coefficient. The overall process can be represented by the following expression,

$$g_{att,c} = \phi^T(\sigma_1(w_1^T x + w_2^T g + b_g)) + b_\phi, \quad (4.3)$$

where  $w_1$ ,  $w_2$ , and  $\phi$  are the linear transformations and  $\sigma_1$  is the ReLU activation operator which is defined as  $\sigma_1(x) = \max(0, x)$ . Afterward, the attention coefficient  $\alpha$  is obtained by applying a sigmoid activation function  $\sigma_2$  to  $g_{att,c}$ . The reason for selecting sigmoid function over softmax is that the sequential use of softmax function may result in sparse activation of the output [77]. Finally, the output of the wavelet guided attention block ( $M_w$ ) is found from the element wise multiplication of the attention coefficient and the input tensor  $X$ . The final output is as follows,

$$M_{w,i} = \sigma_2(g_{att,c}) \cdot X \quad (4.4)$$

#### 4.1.2.2 Boundary Aware Unit (BAU)

In order to focus on the boundary region of the nuclei in histopathology images, a boundary aware unit (BAU) is developed inside the AU which operates on the output of the WGAU ( $M_{w,i}$ ) as shown in Fig.4.3a. BAU first uses a set of  $1 \times 1$  and  $3 \times 3$  convolutional operations on  $M_{w,i}$  to extract  $B_{c,i}$ . Afterwards, it is passed into a  $1 \times 1$  convolutional layer to get the boundary prediction map  $B_i$  (as shown in Fig.4.3a). At each stage  $i$  of the aggregation block corresponding boundary prediction ( $B_i$ ) will be generated. However, the size of this predicted map will change at each stage due to up-sampling/downsampling operation of the encoder/decoder architecture. As a result in order to combine boundary Aware output of different stages,  $B_i$  at lower stages ( $i > 1$ ) need to be upsampled as shown in Fig.4.3b. For example, for the  $i^{\text{th}}$  stage, output  $B_i$  needs to be upsampled by  $2^{i-1} \times 2^{i-1}$ . After  $3 \times 3$  convolution operation on the combined output, the final boundary aware map,  $B$  is obtained. The  $B$  is learned under the supervision of the boundary constraint map  $Y_B$  which is shown in Fig 4.5.  $Y_B$  is created in such a way so that it can help the network to focus more on the boundary pixels instead of equal consideration of all the pixels inside the nucleus region. Let  $e \in E$  denote the point collection of nuclei boundary and  $p$  denote the pixel point collection in a microscopy image.  $Y_B(i, j)$  can be computed by optimizing the function as follows:

$$Y_B(i, j) = f(\min_{e \in E} d(e; p)) \quad (4.5)$$

where  $d(e; p)$  denotes the Cartesian distance between each pixel point  $p(i, j)$  inside the nucleus and the boundary pixel  $e$  of that nucleus region, and  $f(x)$  is defined as  $f(x) = \max(x) - x$ . Intuitively, the pixel value  $Y_B(i, j)$  will be larger if the pixel point  $p$  is closer to the edge. Finally, utilizing  $Y_B(i, j)$  and output  $B$ , an additional loss function  $L_B$  is proposed to get the segmentation result from this path. It is expected that the segmentation obtained based on this boundary constraint map can preserve the

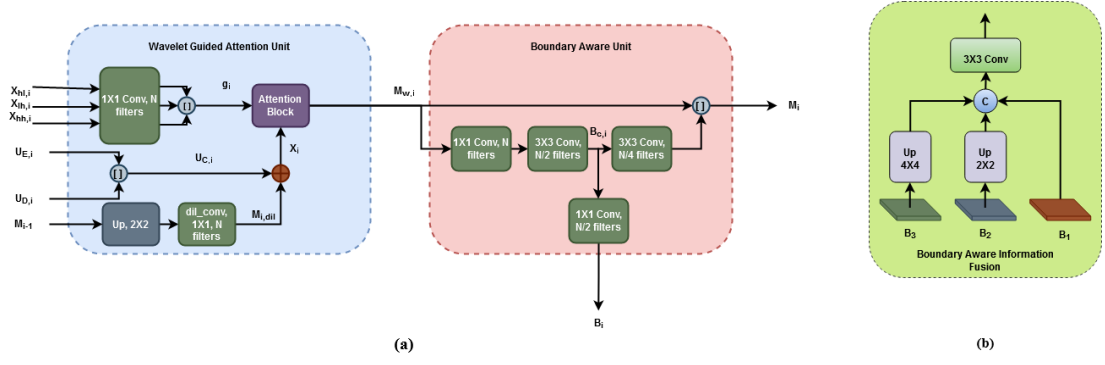


Figure 4.3: (a) Proposed aggregation unit (AU): It consists of two major parts- wavelet guided attention unit (WGUAU) and boundary aware unit (BAU). (b) Boundary Aware information fusion: It fuses multi-level boundary information  $S_i$  ( $i= 1, 2, 3$ ) generated from BAU.

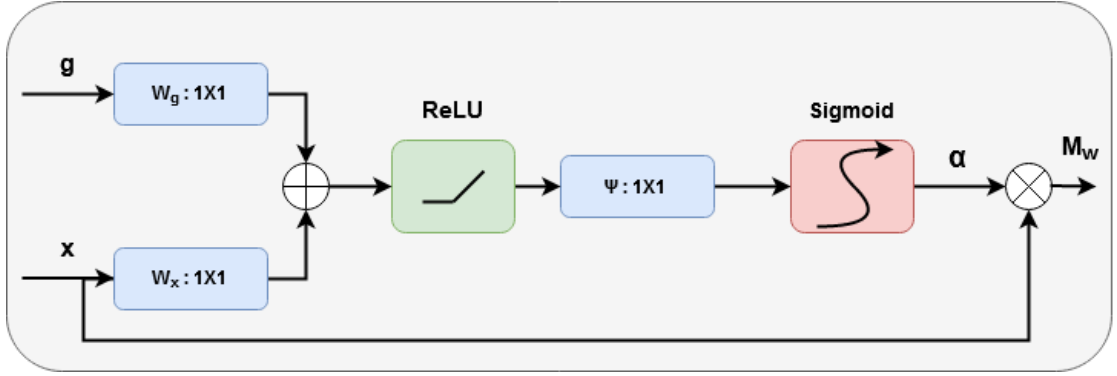


Figure 4.4: A visualization of the attention mechanism used in WGUAU to combine the wavelet and spatial domain information.

nucleus boundary region more precisely. As a result, when such a boundary aware operation is conducted jointly with the baseline Unet based segmentation and aggregation block based segmentation, a superior nuclei segmentation is obtained. Finally, three segmented binary outputs are found from the proposed network. Since it is a binary class segmentation problem it is more appropriate to follow majority voting approach. Hence in this method, we used the majority voting strategy to produce the final nuclei segmentation output from the three segmented outputs.

### 4.1.3 Hybrid Loss Function

Since, there are three loss functions, namely the decoder loss ( $\mathcal{L}_M$ ), the aggregation loss ( $\mathcal{L}_A$ ), and the boundary aware loss ( $\mathcal{L}_B$ ), a hybrid loss function is employed summing all the three loss functions. The total loss  $\mathcal{L}$  is defined as following,

$$\mathcal{L} = \mathcal{L}_M + \mathcal{L}_A + \mathcal{L}_B. \quad (4.6)$$

Here,  $\mathcal{L}_A$  and  $\mathcal{L}_M$  are defined by the conventional binary cross-entropy loss,  $\mathcal{L}_C$  which

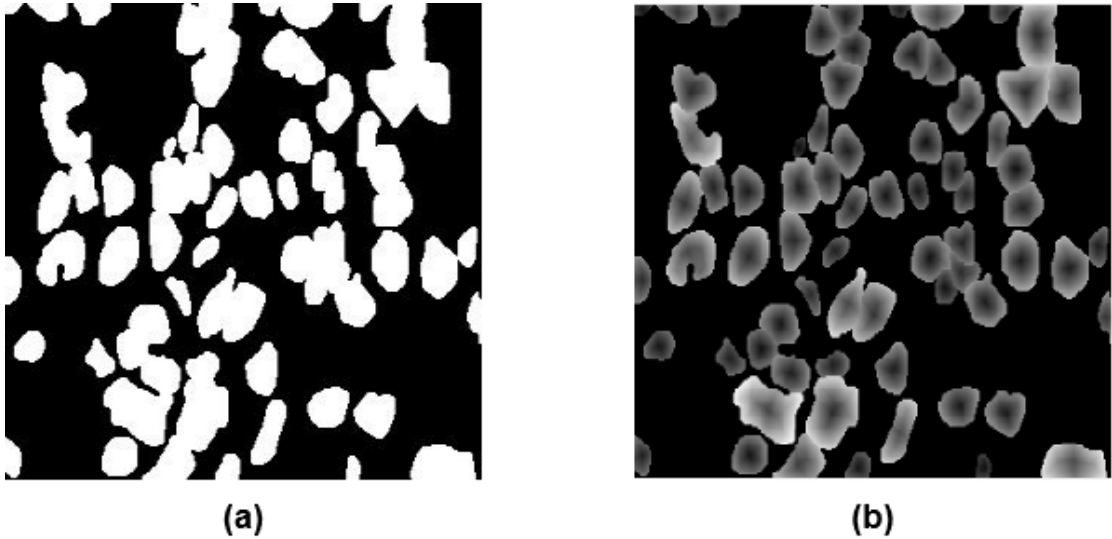


Figure 4.5: A visualization of the (a) original segmented output and (b) the corresponding boundary constraint map.

is defined by the following equation:

$$\mathcal{L}_c = \frac{1}{N} \sum_{i=1}^N p_i \log y_i + (1 - p_i) \log(1 - y_i) \quad (4.7)$$

where  $p_i$  denotes the predicted probability value of  $i^{th}$  sample and  $y_i$  represents the corresponding label.

Conversely, boundary aware loss is needed to constrain the whole unit to pay attention to boundary-focused morphological information, which is defined as the mean square error between  $B$  and  $Y_B$

$$\mathcal{L}_B = l_{mse}(B; Y_B) \quad (4.8)$$

where  $l_{mse}$  is the mean squared error loss.

## 4.2 Results and Discussions

In this section, the description of the datasets used in this method, the required setup for our experiment and results found from extensive experimentation on several publicly available datasets are presented and discussed to validate the effectiveness of the proposed scheme.

### 4.2.1 Dataset Description

To verify the effectiveness of the proposed algorithm, three publicly available datasets are employed. The first one is Data Science Bowl (DSB-2018) which was released by Kaggle for competition purpose [78]. The dataset contains a large number of seg-

Table 4.1: Ablation study of the proposed network on three datasets.

Prepr.	WGAU	BAU	DSB-2018		MoNuSeg		TNBC	
			Dice	IOU	Dice	IOU	Dice	IOU
X	X	X	83.50	74.57	74.28	60.23	68.21	52.24
✓	X	X	87.20	78.26	79.54	66.97	72.45	57.23
✓	✓	X	89.40	80.57	83.67	71.02	75.98	59.89
✓	✓	✓	90.82	82.43	85.74	74.89	78.57	61.90

mented nuclei images that were acquired under different conditions and vary in cell type, shape, magnification, illumination status, and imaging modality (brightfield and fluorescence). The training set of this dataset contains 670 images in which 546 images are brightfield and the rests are fluorescence. The test set contains 65 images and their manual segmentations are provided for the evaluation purpose.

The second dataset used for evaluation purpose is multi-organ nuclei segmentation challenge (MoNuSeg) which contains a diverse set of Hematoxylin-Eosin (HE) stained tissue images [79]. It contains 30 images with around 22,000 nuclear boundary annotations for training and 14 images with 7,000 nuclear boundary annotations for testing purpose. This dataset was collected from 18 hospitals and a wide range of patients. The size of all the images is  $1000 \times 1000$ .

The final dataset is the triple negative breast cancer (TNBC) dataset which consists of 50 images with a total of 4,022 annotated cells, including normal epithelial and myoepithelial breast cells (localized in ducts and lobules), invasive carcinomatous cells, fibroblasts, endothelial cells, adipocytes, macrophages and inflammatory cells (lymphocytes and plasmocytes) [70]. The size of the images is  $500 \times 500$ .

## 4.2.2 Experimental Setup

For the augmentation techniques used in our methods, the original images are rotated by  $30^\circ$  and  $60^\circ$ , flipped horizontally, and sheared by following affine transformation

$$\begin{bmatrix} 1 & 0 & 0 \\ s & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

where  $s = 0.3$  is selected. To make our proposed model perform better for the nuclei segmentation task, different hyperparameters are chosen based on empirical analysis. Adam optimizer with a learning rate of 0.001 and a decaying rate of 0.99 after each 10 epochs is chosen to reduce the loss of the network. The network is trained for 100 epochs on each dataset which is found sufficient in these cases. All the experiments have been implemented on the Google cloud platform with NVIDIA P-100 GPU as the hardware accelerator.

For the DSB dataset, all the images are resized into  $256 \times 256$  to feed into our network. For the implementations of pre-state-of-the-art methods, Segnet, Micro-Net, and

Table 4.2: Performance comparison of different architectures in the TNBC dataset (8:1:1 split setup).

Network	Dice(%)	IOU(%)	Precision(%)	Recall(%)
U-Net [42]	68.61	52.92	65.94	72.54
Mask-RCNN [86]	70.54	52.97	66.27	75.39
DIST [70]	70.51	56.34	66.82	74.63
Micro-Net [47]	71.23	53.71	66.52	76.61
FCN [87]	72.67	50.62	67.53	78.53
Segnet [83]	73.88	54.93	71.11	76.71
HoVer-Net [71]	74.32	57.13	68.91	80.61
Our Method	<b>78.57</b>	<b>61.90</b>	<b>73.45</b>	<b>81.90</b>

HoVer-Net, as per the requirement stated there, the input size of the images are resized to  $224 \times 224$ ,  $252 \times 252$ , and  $270 \times 270$ , respectively [47, 71, 83]. For MoNuseg dataset, each image is zero padded to convert the size to  $1024 \times 1024$  and then 16 non-overlapping image patches of size  $256 \times 256$  are created. Each image patch thus created is considered as an independent image in the training process. Finally, for TNBC dataset, since the size of the images are  $512 \times 512$ , four image patches from a single image are obtained, which are employed for the training purpose. During the testing phase, the similar patches are employed in the model and their segmented images are merged to obtain the segmentation results of the whole image. The batch size are selected as 32, 16, and 16 for the DSB-2018, MoNuSeg, and TNBC datasets, respectively.

Several standard evaluation metrics are utilized to assess the performance of the proposed model which are stated as follows:

$$\text{Dice} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}} \quad (4.9)$$

$$\text{IOU} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \quad (4.10)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (4.11)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (4.12)$$

where, TP, FP, and FN represent true positive, false positive, and false negative predictions, respectively. The images available in the DSB-2018 and TNBC data are splitted as 80:10:10 for training, validation, and testing. Moreover, the results on the stage-1 test set of DSB-2018 and on the test set of MoNuSeg are also reported and compared with the state of the art segmentation networks to validate the superiority of the proposed network.



Table 4.3: Performance comparison of different architectures in the MoNuSeg dataset (8:1:1 split setup).

Network	Dice(%)	IOU(%)	Precision(%)	Recall(%)
FCN [87]	79.77	62.45	73.21	87.46
Unet [42]	74.67	60.89	71.24	78.44
U-net++ [44]	75.68	62.01	72.12	79.60
DIST [70]	77.31	63.77	75.31	79.41
Segnet [83]	77.44	63.3	73.91	81.32
HoVer-Net [71]	79.63	66.47	77.02	82.42
MicroNet [47]	80.24	68.02	77.98	82.63
NucleiSegNet [62]	83.59	72.06	78.95	88.81
Our Method	<b>85.74</b>	<b>74.89</b>	<b>79.95</b>	<b>92.43</b>

Table 4.4: Performance comparison of different architectures in the DSB-2018 dataset

Network	8:1:1 split			
	Dice(%)	IOU(%)	Precision(%)	Recall(%)
FCN [87]	74.12	60.67	83.01	66.94
U-Net [42]	86.75	76.97	85.55	87.98
Segnet [83]	75.48	62.36	84.08	68.47
HoVer-Net [71]	89.42	80.14	87.92	90.97
Micro-Net [47]	89.23	79.97	85.78	92.96
FPN [45]	87.34	77.97	80.34	95.67
U-Net++ [44]	89.74	79.92	82.34	98.60
Our Method	<b>90.82</b>	<b>82.43</b>	<b>88.56</b>	<b>98.65</b>

### 4.2.3 Analysis of the Segmentation Performance

At first, an ablation study is provided to describe the performance of each of the modules of the proposed methodology. Eventually, the performance of the proposed network is compared with that obtained by the other state-of-the-art segmentation networks and its superiority is explained in qualitative and quantitative manner.

Table 4.5: Performance comparison of different architectures in the DSB-2018 dataset.

Network	Test Set			
	Dice(%)	IOU(%)	Precision(%)	Recall(%)
FCN [87]	71.86	60.69	80.56	64.85
U-Net [42]	81.2	74.28	82.85	79.61
Segnet [83]	73.84	62.06	82.01	67.15
HoVer-Net [71]	83.57	74.97	83.85	83.29
Micro-Net [47]	83.79	73.13	82.12	85.52
FPN [45]	82.75	72.56	82.56	82.94
U-Net++ [44]	83.98	73.56	84.56	83.40
Our Method	<b>85.45</b>	<b>75.61</b>	<b>87.45</b>	<b>86.53</b>

### 4.2.3.1 Ablation Study

As it is mentioned earlier, a traditional U-net architecture is utilized as the baseline network of our proposed model. Over the top of that, three additional techniques/modules are incorporated and their individual and joint performances are compared to analyze the contribution of each of them. The ablation study is summarized in Table 4.1 where only Dice and IOU metrics are utilized for evaluation purpose. After analyzing the table, it is found that 4.4%, 7.1%, and 6.2% improvements in dice score are achieved after incorporating the preprocessing step in DSB-2018, MoNuSeg, and TNBC datasets, respectively. The preprocessing steps allow to reduce the variations in intensities across cell subtypes and leverage the model to distinguish between the nucleus and non-nucleus characteristics. Moreover, the introduction of WGAU to guide the spatial features by the frequency domain information from DWT through an attention mechanism contributes to an efficient representation of the features by combining both the spatial and frequency level information and provide feature maps with fine scale details. This unit contributes to additional 2.5%, 5.2%, and 4.9% improvements in the three datasets, respectively. Finally, the boundary Aware unit provides the edge information to the network to capture the boundary or edge level information. By providing edge information, it indirectly represents the edge or boundary representation of the blocks and increases the performance by 1.6%, 2.5%, and 3.4% dice improvement in DSB-2018, MoNuSeg, and TNBC databases, respectively. Similar scenario is observed in the perspective of another standard evaluation metric, IOU, in all three databases.

### 4.2.3.2 Quantitative Analysis

In order to compare the effectiveness of our proposed method, a comparative analysis is presented for the TNBC, MoNuSeg, and DSB-2018 datasets in Table 4.2, Table 4.3, and Table 4.5, respectively. It is noticed that our proposed network outperforms all the previously proposed segmentation networks by a significant amount in all the evaluation metrics. The proposed networks exhibits performance improvement over FCN, Unet, Segnet and other benchmark networks in nucleus segmentation task. Aside from these networks, it has also shown its superiority compared to some specific networks previously employed for nuclei segmentation task such as, HoVer-Ner and NucleiSeg-Net. Using the proposed framework, 4.25%, 2.15%, and 1.47% dice improvements are achieved over the results obtained in HoVer-net for the TNBC, MoNuSeg, and DSB-2018 datasets, respectively. The proposed network's robustness and efficacy in determining the nuclei region can be corroborated by the high scores of other metrics reported. The proposed network's ability to provide additional wavelet domain information from DWT through the attention mechanism provides additional information to the network. Moreover, with the help of boundary aware unit, it performs efficiently in

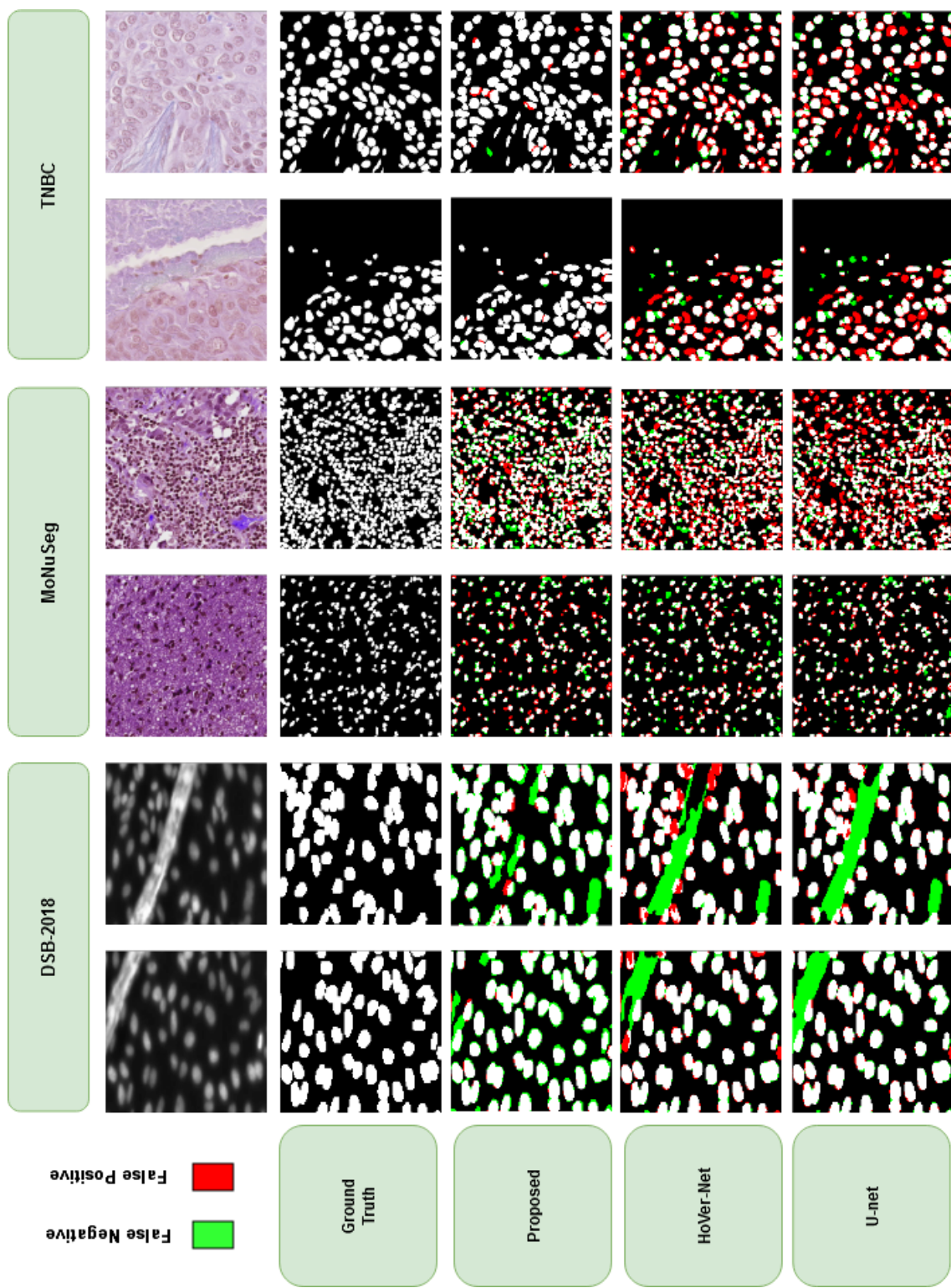


Figure 4.6: Visualization of the nuclei segmentation performance of some of the state-of-the-art networks in three datasets. The inputs of the first two rows correspond to DSB-2018 dataset, the next two rows correspond to MoNuSeg dataset, and the final two rows correspond to TNBC dataset.

capturing the small nuclei regions and thus improves both the precision and recall.

Table 4.6: Computational Efficiency Analysis of Numerous Architectures along with the Performance of the proposed method.

Methods	Number of Parameters	Inference Time
Unet	31.0	0.1
MicroNet	184	0.18
DIST	8	0.12
SegNet	14.7	0.17
Attention U-net	8.727	0.09
CENet	38.8	0.12
FPN	8.4	0.137
HoVerNet	7.7	0.15
NucleiSegNet	15.7	0.10
BAWGNet	1.5	0.09

#### 4.2.3.3 Qualitative Analysis

Quantitative analysis alone cannot always define the effectiveness and superiority of a method. The efficacy of the model can only be identified if the model performs outstandingly in the critical cases of a particular problem. Hence, the segmentation performance of different networks along with our proposed network in some challenging cases are presented in Fig.4.6. The presence of other cell organelles with similar intensity levels to the nucleus, as well as the position of the nucleus in close proximity to each other, might be responsible for a significant number of false positive cases. It is evident from the segmentation performance of the other networks that all of them struggle to address these issues. On the other hand, our proposed method with its efficient utilization of spatial and frequency level information along with the boundary information considerably addresses these issues and significantly shows its performance in the challenging situation of nuclei segmentation task. The computational efficiency of different networks are summarized in Table 4.6. The efficient incorporation of boundary aware

Table 4.7: Performance of the proposed method on DSB dataset (10 fold cross validation)

Fold No.	Dice	IOU	Precision	Recall
1	88.91	81.03	86.98	85.23
2	89.67	81.23	87.40	85.76
3	88.65	81.12	87.11	85.34
4	90.92	83.03	88.89	86.97
5	91.51	83.78	89.43	87.51
6	88.93	80.93	86.99	85.02
7	89.86	81.34	87.81	85.91
8	90.25	82.24	88.34	86.23
9	90.51	82.39	88.87	86.91
10	89.61	81.58	87.78	85.83

Table 4.8: Comparison of the three proposed architectures on three datasets (DSB-2018, MoNuSeg, and TNBC)

Name of the Method	DSB-2018			MoNuSeg			TNBC					
	Dice	IOU	Precision	Recall	Dice	IOU	Precision	Recall	Dice	IOU	Precision	Recall
<b>Boundary Aware Feature Map Guided Attention Network</b>	88.05	79.98	81.07	95.81	80.96	69.92	78.82	87.21	74.25	55.87	71.32	78.42
<b>Contourlet Guided Attention Network</b>	88.91	80.81	82.11	96.73	81.71	70.89	79.95	88.33	75.12	56.87	72.02	79.13
<b>Boundary Aware Wavelet Guided Attention Network</b>	<b>90.82</b>	<b>82.43</b>	<b>88.56</b>	<b>98.65</b>	<b>85.74</b>	<b>74.89</b>	<b>79.95</b>	<b>92.43</b>	<b>78.57</b>	<b>61.90</b>	<b>73.45</b>	<b>81.90</b>

unit with the attention mechanism makes the model computationally efficient with respect to the other models. Also, the inference time of the network is really noteworthy in comparison to the state-of-the-art HoVerNet and NucleiSegNet. The network is 1.66 times faster than the HoVerNet. Although the network infers equally faster with respect to the NucleiSegNet, the memory efficiency of this network is 10.46 times better.

Additionally, a ten fold cross validation technique is utilized to validate the consistency among the data present in individual folds. The analysis is done on DSB-2018 dataset and the results found from different folds of data is summarized in Table 4.7. It can be concluded from the table that the results are consistent with each other at different folds with a minimal deviation. Moreover, the performance of this method is compared with the other two proposed methods of this thesis and summarised in Table 4.8. It is found that a significant improvement over the other two methods is achieved because of the introduction of three separate loss functions and the proposed aggregation unit.

### **4.3 Conclusion**

Accurate nuclei segmentation is a challenging task due to varying characteristics of nuclei in contrast level and color statistics, and varieties of acquisition-specific modalities. The proposed BAWGNet architecture overcomes these challenges by introducing an aggregation unit, which includes the WGAU and BAU. The proposed preprocessing steps in our work are found very effective to increase the performance of the segmentation task. Guiding the spatial level information by the high-frequency wavelet coefficients through the WGAU compensates the loss of information resulting from the subsequent downsampling operation of traditional encoder-decoder architecture. On top of that, the proposed boundary aware unit preserves the spatial content and improves the discriminative capability of intermediate features for separating the adhesion and clustered nuclei. From experimentation, it is found that the incorporation of the WGAU and BAU provides an average of 5.3% improvement on three benchmark datasets of histopathology images. The dominating performance of the proposed method compared to the other conventional encoder-decoder architecture and modern deep learning architectures demonstrates the superiority of this network in the accurate segmentation of nuclei in cell microscopy images.

# Chapter 5

## Conclusions

### 5.1 Contribution of This Thesis

The major contributions of the thesis are described below:

- In this thesis, several deep learning frameworks are proposed which can segment both the simple and complex structured nuclei from histopathology images by creating an edge/boundary level awareness in the networks. To create this boundary level awareness three different boundary preserving guiding signal generation schemes are proposed.
- A shallow encoder-decoder architecture is utilized which preserves boundary preserving signal to guide the attention mechanism of the proposed boundary aware feature map guided attention network. A separate loss function is used to minimize the primary network by utilizing the boundary only mask as the ground truth.
- Contourlet transform based boundary preserving signal generation scheme is also explored in this thesis. The features found from the contourlet transform are found very effective in extracting the fine edge details of the nuclei regions.
- One of the critical issues in the nucleus segmentation task is the generalization of the network to work equally well for different image modalities. Due to their variation in visual appearance and intensity level, it is difficult to train a universal model to work equally well for both of these modalities. The lack of abundant data further complicates this process and demands a preprocessing step to transfer all these modalities into a common intensity range. We have proposed an effective preprocessing step to bring all the image modalities in a common intensity range so that they are consistent with each other.

- One major concern in U-net based network is the use of downsampling operations, such as max or average pooling that usually ignore the Nyquist sampling theorem, resulting in partial or complete loss of high frequency data details and distorted basic structure presented by the low frequency components. In the proposed networks, we considered both the wavelet based pooling and guiding mechanism to address the issue which is found very effective in preserving the content of the nucleus structure.
- In the proposed approaches, multiple loss functions are considered to make the final prediction which improves the segmentation performance. Jointly optimized and hybrid loss functions are developed to utilize this multiple loss functions.
- The feature maps generated at different encoder and decoder stages are visualized to show the boundary aware information generated by the network.
- Extensive experimentations have been done over three datasets and the results are compared with the state-of -the-art methods to corroborate their superiority and effectiveness.

## **5.2 Future Prospects of Our Work**

Although the proposed schemes presented in this thesis achieve significant performance, some other studies can also be conducted in future. While consistent performance on nucleus segmentation has been achieved, the proposed segmentation approaches will be expanded with the incorporation of diverse datasets such as patient-based studies taking into account age, gender, health conditions, and geographical locations of the patients. These factors can provide additional information which may assist a better performance. Hence, for a better understanding of the nature of the cellular structure, an in-depth, closer, patient-specific study should be conducted. Our current study only focuses to segment the nuclei in a histopathology images. However, as a future work, a separate study can be explored to determine whether the nucleus is cancerous or not by utilizing the morphological information. The segmentation of the nucleus along with the identification of the nucleus type will assist the clinicians to assess the current stage of cancer which eventually makes the diagnosis process more efficient.



# References

- [1] R. Seruca, J. Suri, and J. Sanches, *Fluorescence Imaging and Biological Quantification*. CRC Press, 2017.
- [2] H. F. Lodish, A. Berk, C. Kaiser, M. Krieger, M. P. Scott, A. Bretscher, H. L. Ploegh, and P. T. Matsudaira, *Molecular cell biology*. W.H. Freeman, 2008.
- [3] E. Enger *et al.*, *Concepts in Biology'2007 Ed.* Rex Bookstore, Inc., 2003.
- [4] F. Xing and L. Yang, “Robust nucleus/cell detection and segmentation in digital pathology and microscopy images: a comprehensive review,” *IEEE reviews in biomedical engineering*, vol. 9, pp. 234–263, 2016.
- [5] H. Irshad, A. Veillard, L. Roux, and D. Racoceanu, “Methods for nuclei detection, segmentation, and classification in digital histopathology: a review—current status and future potential,” *IEEE reviews in biomedical engineering*, vol. 7, pp. 97–114, 2013.
- [6] A. Ferro, T. Mestre, P. Carneiro, I. Sahumbaiev, R. Seruca, and J. M. Sanches, “Blue intensity matters for cell cycle profiling in fluorescence dapi-stained images,” *Laboratory Investigation*, vol. 97, no. 5, pp. 615–625, 2017.
- [7] T. Mestre, J. Figueiredo, A. S. Ribeiro, J. Paredes, R. Seruca, and J. M. Sanches, “Quantification of topological features in cell meshes to explore e-cadherin dysfunction,” *Scientific reports*, vol. 6, no. 1, pp. 1–12, 2016.
- [8] R. Seruca, J. S. Suri, and J. M. Sanches, *Fluorescence imaging and biological quantification*. CRC Press, 2017.
- [9] D. Zink, A. H. Fischer, and J. A. Nickerson, “Nuclear structure in cancer cells,” *Nature reviews cancer*, vol. 4, no. 9, pp. 677–687, 2004.
- [10] E. C. Jensen, “Overview of live-cell imaging: requirements and methods used,” *The Anatomical Record: Advances in Integrative Anatomy and Evolutionary Biology*, vol. 296, no. 1, pp. 1–8, 2013.

- [11] G. Crivat and J. W. Taraska, “Imaging proteins inside cells with fluorescent tags,” *Trends in biotechnology*, vol. 30, no. 1, pp. 8–16, 2012.
- [12] A. Ettinger and T. Wittmann, “Fluorescence live cell imaging,” *Methods in cell biology*, vol. 123, pp. 77–94, 2014.
- [13] A. S. Stender, K. Marchuk, C. Liu, S. Sander, M. W. Meyer, E. A. Smith, B. Neupane, G. Wang, J. Li, J.-X. Cheng *et al.*, “Single cell optical imaging and spectroscopy,” *Chemical reviews*, vol. 113, no. 4, pp. 2469–2527, 2013.
- [14] J. Pawley, *Handbook of biological confocal microscopy*. Springer Science & Business Media, 2006, vol. 236.
- [15] N. Naredi-Rainer, J. Prescher, A. Hartschuh, and D. C. Lamb, “Confocal microscopy,” *Fluorescence Microscopy*, vol. 1075, pp. 175–213, 2013.
- [16] C. McQuin, A. Goodman, V. Chernyshev, L. Kametsky, B. A. Cimini, K. W. Karhohs, M. Doan, L. Ding, S. M. Rafelski, D. Thirstrup *et al.*, “Cellprofiler 3.0: Next-generation image processing for biology,” *PLoS biology*, vol. 16, no. 7, p. e2005970, 2018.
- [17] C. A. Schneider, W. S. Rasband, and K. W. Eliceiri, “Nih image to imagej: 25 years of image analysis,” *Nature methods*, vol. 9, no. 7, pp. 671–675, 2012.
- [18] S. Berg, D. Kutra, T. Kroeger, C. N. Straehle, B. X. Kausler, C. Haubold, M. Schiegg, J. Ales, T. Beier, M. Rudy *et al.*, “Ilastik: interactive machine learning for (bio) image analysis,” *Nature methods*, vol. 16, no. 12, pp. 1226–1232, 2019.
- [19] F. De Chaumont, S. Dallongeville, N. Chenouard, N. Hervé, S. Pop, T. Provoost, V. Meas-Yedid, P. Pankajakshan, T. Lecomte, Y. Le Montagner *et al.*, “Icy: an open bioimage informatics platform for extended reproducible research,” *Nature methods*, vol. 9, no. 7, pp. 690–696, 2012.
- [20] I. Belevich, M. Joensuu, D. Kumar, H. Vihinen, and E. Jokitalo, “Microscopy image browser: a platform for segmentation and analysis of multidimensional datasets,” *PLoS biology*, vol. 14, no. 1, p. e1002340, 2016.
- [21] M. N. Gurcan, T. Pan, H. Shimada, and J. Saltz, “Image analysis for neuroblastoma classification: segmentation of cell nuclei,” in *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2006, pp. 4844–4847.
- [22] H. Irshad, “Automated mitosis detection in histopathology using morphological and multi-channel statistics features,” *Journal of pathology informatics*, vol. 4, no. 1, p. 10, 2013.

- [23] H. Chang, J. Han, A. Borowsky, L. Loss, J. W. Gray, P. T. Spellman, and B. Parvin, “Invariant delineation of nuclear architecture in glioblastoma multi-forme for clinical and molecular association,” *IEEE Trans. Med. Imag.*, vol. 32, no. 4, pp. 670–682, 2012.
- [24] H. Kong, M. Gurcan, and K. Belkacem-Boussaid, “Partitioning histopathological images: an integrated framework for supervised color-texture segmentation and cell splitting,” *IEEE Trans. Med. imag.*, vol. 30, no. 9, pp. 1661–1677, 2011.
- [25] M. Zhang, T. Wu, and K. M. Bennett, “Small blob identification in medical images using regional features from optimum scale,” *IEEE Trans. Biomed. Eng.*, vol. 62, no. 4, pp. 1051–1062, 2014.
- [26] S. Petushi, F. U. Garcia, M. M. Haber, C. Katsinis, and A. Tozeren, “Large-scale computations on histology images reveal grade-differentiating parameters for breast cancer,” *BMC medical imaging*, vol. 6, no. 1, pp. 1–11, 2006.
- [27] O. S. Al-Kadi, “Texture measures combination for improved meningioma classification of histopathological images,” *Pattern recognition*, vol. 43, no. 6, pp. 2043–2053, 2010.
- [28] Y. Cui, G. Zhang, Z. Liu, Z. Xiong, and J. Hu, “A deep learning algorithm for one-step contour aware nuclei segmentation of histopathology images,” *Medical & biological engineering & computing*, vol. 57, no. 9, pp. 2027–2043, 2019.
- [29] M. Veta, P. J. Van Diest, R. Kornegoor, A. Huisman, M. A. Viergever, and J. P. Pluim, “Automatic nuclei segmentation in h&e stained breast cancer histopathology images,” *PloS one*, vol. 8, no. 7, p. e70221, 2013.
- [30] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, and J. Garcia-Rodriguez, “A review on deep learning techniques applied to semantic segmentation,” *arXiv preprint arXiv:1704.06857*, 2017.
- [31] N. Tajbakhsh, L. Jeyaseelan, Q. Li, J. N. Chiang, Z. Wu, and X. Ding, “Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation,” *Medical Image Analysis*, vol. 63, pp. 101 693–101 722, 2020.
- [32] T. Falk, D. Mai, R. Bensch, Ö. Çiçek, A. Abdulkadir, Y. Marrakchi, A. Böhm, J. Deubner, Z. Jäckel, K. Seiwald *et al.*, “U-net: deep learning for cell counting, detection, and morphometry,” *Nature Methods*, vol. 16, no. 1, pp. 67–70, 2019.
- [33] P. Huang, J. Wang, J. Zhang, Y. Shen, C. Liu, W. Song, S. Wu, Y. Zuo, Z. Lu, and D. Li, “Attention-aware residual network based manifold learning for white blood cells classification,” *IEEE J. Biomed. Health Inform.*, vol. 25, no. 4, pp. 1206–1214, 2020.

- [34] A. Esteva, K. Chou, S. Yeung, N. Naik, A. Madani, A. Mottaghi, Y. Liu, E. Topol, J. Dean, and R. Socher, “Deep learning-enabled medical computer vision,” *NPJ digital medicine*, vol. 4, no. 1, pp. 1–9, 2021.
- [35] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghahfarokan, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, “A survey on deep learning in medical image analysis,” *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.
- [36] F. Isensee, P. F. Jaeger, S. A. Kohl, J. Petersen, and K. H. Maier-Hein, “nnu-net: a self-configuring method for deep learning-based biomedical image segmentation,” *Nature methods*, vol. 18, no. 2, pp. 203–211, 2021.
- [37] A. Srivastava, D. Jha, S. Chanda, U. Pal, H. D. Johansen, D. Johansen, M. A. Riegler, S. Ali, and P. Halvorsen, “Msr-net: A multi-scale residual fusion network for biomedical image segmentation,” *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 5, pp. 2252–2263, 2021.
- [38] N. K. Tomar, D. Jha, M. A. Riegler, H. D. Johansen, D. Johansen, J. Rittscher, P. Halvorsen, and S. Ali, “Fanet: A feedback attention network for improved biomedical image segmentation,” *IEEE Transactions on Neural Networks and Learning Systems*, 2022.
- [39] N. R. Gudhe, H. Behravan, M. Sudah, H. Okuma, R. Vanninen, V.-M. Kosma, and A. Mannermaa, “Multi-level dilated residual network for biomedical image segmentation,” *Scientific Reports*, vol. 11, no. 1, pp. 1–18, 2021.
- [40] N. Ibtehaz and M. S. Rahman, “Multiresunet: Rethinking the u-net architecture for multimodal biomedical image segmentation,” *Neural networks*, vol. 121, pp. 74–87, 2020.
- [41] Y. Al-Kofahi, W. Lassoued, W. Lee, and B. Roysam, “Improved automatic detection and segmentation of cell nuclei in histopathology images,” *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 4, pp. 841–852, 2009.
- [42] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *Proc. Int. Conf. Medical Image Computing and Computer-assisted Intervention*. Springer, 2015, pp. 234–241.
- [43] F. Long, “Microscopy cell nuclei segmentation with enhanced U-Net,” *BMC Bioinformatics*, vol. 21, no. 1, pp. 1–12, 2020.
- [44] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, “Unet++: Redesigning skip connections to exploit multiscale features in image segmentation,” *IEEE Trans. on Medical Imaging*, vol. 39, no. 6, pp. 1856–1867, 2019.

- [45] Z. Cheng and A. Qu, “A fast and accurate algorithm for nuclei instance segmentation in microscopy images,” *IEEE Access*, vol. 8, pp. 158 679–158 689, 2020.
- [46] B. Olimov, K. Sanjar, S. Din, A. Ahmad, A. Paul, and J. Kim, “FU-net: fast biomedical image segmentation model based on bottleneck convolution layers,” *Multimedia Systems*, pp. 1–14, 2021.
- [47] S. E. A. Raza, L. Cheung, M. Shaban, S. Graham, D. Epstein, S. Pelengaris, M. Khan, and N. Rajpoot, “Micro-Net: A unified model for segmentation of various objects in microscopy images,” *Medical Image Analysis*, vol. 52, pp. 160–173, 2019.
- [48] S. Gehlot, A. Gupta, and R. Gupta, “EDNFC-Net: Convolutional neural network with nested feature concatenation for nuclei-instance segmentation,” in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 1389–1393.
- [49] Q. Li, L. Shen, S. Guo, and Z. Lai, “WaveCNet: Wavelet integrated cnns to suppress aliasing effect for noise-robust image classification,” *IEEE Trans. Image Process.*, vol. 30, pp. 7074–7089, 2021.
- [50] T. Williams and R. Li, “Wavelet pooling for convolutional neural networks,” in *International Conference on Learning Representations*, 2018.
- [51] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz *et al.*, “Attention U-net: Learning where to look for the pancreas,” *arXiv preprint arXiv:1804.03999*, 2018.
- [52] P. Zhao, J. Zhang, W. Fang, and S. Deng, “Scau-net: spatial-channel attention u-net for gland segmentation,” *Frontiers in Bioengineering and Biotechnology*, vol. 8, p. 670, 2020.
- [53] H. Ali, L. Cui, J. Feng *et al.*, “Msal-net: improve accurate segmentation of nuclei in histopathology images by multiscale attention learning network,” *BMC Medical Informatics and Decision Making*, vol. 22, no. 1, pp. 1–11, 2022.
- [54] Q. Jin, Z. Meng, C. Sun, H. Cui, and R. Su, “Ra-unet: A hybrid deep attention-aware network to extract liver and tumor in ct scans,” *Frontiers in Bioengineering and Biotechnology*, p. 1471, 2020.
- [55] H. Zhang, H. Zhu, and X. Ling, “Polar coordinate sampling-based segmentation of overlapping cervical cells using attention u-net and random walk,” *Neurocomputing*, vol. 383, pp. 212–223, 2020.

- [56] Z. Liu, H. Wang, S. Zhang, G. Wang, and J. Qi, “Nas-scnn: Neural architecture search-based spatial and channel joint attention module for nuclei semantic segmentation and classification,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2020, pp. 263–272.
- [57] M. Sahasrabudhe, S. Christodoulidis, R. Salgado, S. Michiels, S. Loi, F. André, N. Paragios, and M. Vakalopoulou, “Self-supervised nuclei segmentation in histopathological images using attention,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2020, pp. 393–402.
- [58] J. Zhang, Z. Liu, B. Du, J. He, G. Li, and D. Chen, “Binary tree-like network with two-path fusion attention feature for cervical cell nucleus segmentation,” *Computers in Biology and Medicine*, vol. 108, pp. 223–233, 2019.
- [59] H. He, C. Zhang, J. Chen, R. Geng, L. Chen, Y. Liang, Y. Lu, J. Wu, and Y. Xu, “A hybrid-attention nested UNet for nuclear segmentation in histopathological images,” *Frontiers in Molecular Biosciences*, vol. 8, p. 6, 2021.
- [60] A. Vahadane, B. Atheeth, and S. Majumdar, “Dual encoder attention U-net for nuclei segmentation,” in *Proc. 43rd Annual Int. Conf. the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2021, pp. 3205–3208.
- [61] J. Zhao, Y. He, S. Zhao, J. Huang, and W. Zuo, “AL-net: Attention learning network based on multi-task learning for cervical nucleus segmentation,” *IEEE J. Biomed. Health Inform.*, early access, 2021. doi:10.1109/JBHI.2021.3136568.
- [62] S. Lal, D. Das, K. Alabhya, A. Kanfode, A. Kumar, and J. Kini, “NucleiSeg-net: Robust deep learning architecture for the nuclei segmentation of liver cancer histopathology images,” *Computers in Biology and Medicine*, vol. 128, pp. 104075–104082, 2021.
- [63] Y. Yuan, M. Chao, and Y.-C. Lo, “Automatic skin lesion segmentation using deep fully convolutional networks with jaccard distance,” *IEEE transactions on medical imaging*, vol. 36, no. 9, pp. 1876–1886, 2017.
- [64] J.-G. Lee, S. Jun, Y.-W. Cho, H. Lee, G. B. Kim, J. B. Seo, and N. Kim, “Deep learning in medical imaging: general overview,” *Korean journal of radiology*, vol. 18, no. 4, pp. 570–584, 2017.
- [65] S. Pandey, P. R. Singh, and J. Tian, “An image augmentation approach using two-stage generative adversarial network for nuclei image segmentation,” *Biomedical Signal Processing and Control*, vol. 57, p. 101782, 2020.

- [66] N. Moshkov, B. Mathe, A. Kertesz-Farkas, R. Hollandi, and P. Horvath, “Test-time augmentation for deep learning-based cell segmentation on microscopy images,” *Scientific Reports*, vol. 10, no. 1, pp. 1–7, 2020.
- [67] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets,” *Advances in neural information processing systems*, vol. 27, 2014.
- [68] T. Neff, C. Payer, D. Štern, and M. Urschler, “Generative adversarial networks to synthetically augment data for deep learning based image segmentation,” in *Proceedings of the OAGM workshop*, 2018, pp. 22–29.
- [69] N. Jain, L. Manikonda, A. O. Hernandez, S. Sengupta, and S. Kambhampati, “Imagining an engineer: On gan-based data augmentation perpetuating biases,” *arXiv preprint arXiv:1811.03751*, 2018.
- [70] P. Naylor, M. Laé, F. Reyal, and T. Walter, “Segmentation of nuclei in histopathology images by deep regression of the distance map,” *IEEE Trans. on Medical Imaging*, vol. 38, no. 2, pp. 448–459, 2018.
- [71] S. Graham, Q. D. Vu, S. E. A. Raza, A. Azam, Y. W. Tsang, J. T. Kwak, and N. Rajpoot, “HoVer-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images,” *Medical Image Analysis*, vol. 58, p. 101563, 2019.
- [72] T. Wan, L. Zhao, H. Feng, D. Li, C. Tong, and Z. Qin, “Robust nuclei segmentation in histopathology using asppu-net and boundary refinement,” *Neurocomputing*, vol. 408, pp. 144–156, 2020.
- [73] A. Reza, H. Moein, W. Yuli, and M. Dorit, “Contextual attention network: Transformer meets u-net,” *arXiv preprint arXiv:2203.01932*, 2022.
- [74] R. C. Gonzalez, *Digital Image Processing*. Pearson Education India, 2009.
- [75] A. M. Reza, “Realization of the contrast limited adaptive histogram equalization (clahe) for real-time image enhancement,” *Journal of VLSI Signal Processing Systems for Signal, Image and Video Technology*, vol. 38, no. 1, pp. 35–44, 2004.
- [76] M. Coster and J.-L. Chermant, “Image analysis and mathematical morphology for civil engineering materials,” *Cement and Concrete Composites*, vol. 23, no. 2-3, pp. 133–151, 2001.
- [77] A. Martins and R. Astudillo, “From softmax to sparsemax: A sparse model of attention and multi-label classification,” in *Proc. Int. Conf. Machine Learning*, 2016, pp. 1614–1623.

- [78] J. C. Caicedo, A. Goodman, K. W. Karhohs, B. A. Cimini, J. Ackerman, M. Haghighi, C. Heng, T. Becker, M. Doan, C. McQuin *et al.*, “Nucleus segmentation across imaging experiments: the 2018 data science bowl,” *Nature Methods*, vol. 16, no. 12, pp. 1247–1253, 2019.
- [79] N. Kumar, R. Verma, S. Sharma, S. Bhargava, A. Vahadane, and A. Sethi, “A dataset and a technique for generalized nuclear segmentation for computational pathology,” *IEEE Trans. on Medical Imaging*, vol. 36, no. 7, pp. 1550–1560, 2017.
- [80] J. Duchi, E. Hazan, and Y. Singer, “Adaptive subgradient methods for online learning and stochastic optimization.” *Journal of machine learning research*, vol. 12, no. 7, 2011.
- [81] G. Hinton, N. Srivastava, and K. Swersky, “Neural networks for machine learning lecture 6a overview of mini-batch gradient descent,” *Cited on*, vol. 14, no. 8, p. 2, 2012.
- [82] T. A. Sorensen, “A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on danish commons,” *Biol. Skar.*, vol. 5, pp. 1–34, 1948.
- [83] V. Badrinarayanan, A. Kendall, and R. Cipolla, “Segnet: A deep convolutional encoder-decoder architecture for image segmentation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, 2017.
- [84] R. Hollandi, A. Szkalitsy, T. Toth, E. Tasnadi, C. Molnar, B. Mathe, I. Grexa, J. Molnar, A. Balind, M. Gorbe *et al.*, “nucleaizer: a parameter-free deep learning framework for nucleus segmentation using image style transfer,” *Cell Systems*, vol. 10, no. 5, pp. 453–458, 2020.
- [85] Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao, T. Zhang, S. Gao, and J. Liu, “Ce-net: Context encoder network for 2d medical image segmentation,” *IEEE trans. on med. imag.*, vol. 38, no. 10, pp. 2281–2292, 2019.
- [86] K. He, G. Gkioxari, P. Dollár, and R. Girshick, “Mask R-CNN,” in *Proc. IEEE Int. Conf. Computer Vision*, 2017, pp. 2961–2969.
- [87] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” in *Proc. IEEE Conference Computer Vision and Pattern Recognition*, 2015, pp. 3431–3440.
- [88] H. A. Phoulady, D. Goldgof, L. O. Hall, and P. R. Mouton, “A framework for nucleus and overlapping cytoplasm segmentation in cervical cytology extended



- depth of field and volume images,” *Computerized Medical Imaging and Graphics*, vol. 59, pp. 38–49, 2017.
- [89] M. N. Do and M. Vetterli, “The contourlet transform: an efficient directional multiresolution image representation,” *IEEE Trans. on imag. process.*, vol. 14, no. 12, pp. 2091–2106, 2005.
- [90] M. Golpardaz, M. S. Helfroush, and H. Danyali, “Nonsubsampled contourlet transform-based conditional random field for sar images segmentation,” *Signal Processing*, vol. 174, p. 107623, 2020.
- [91] Z. Wang, X. Li, H. Duan, Y. Su, X. Zhang, and X. Guan, “Medical image fusion based on convolutional neural networks and non-subsampled contourlet transform,” *Expert Systems with Applications*, vol. 171, p. 114574, 2021.
- [92] M. H. Asmare, V. S. Asirvadam, and A. F. M. Hani, “Image enhancement based on contourlet transform,” *Signal, Image and Video Processing*, vol. 9, no. 7, pp. 1679–1690, 2015.
- [93] W. Li, Q. Lin, K. Wang, and K. Cai, “Improving medical image fusion method using fuzzy entropy and nonsampling contourlet transform,” *International Journal of Imaging Systems and Technology*, vol. 31, no. 1, pp. 204–214, 2021.
- [94] L. A. Akter and G.-R. Kwon, “Integration of contourlet transform and canny edge detector for brain image segmentation,” in *2018 Tenth International Conference on Ubiquitous and Future Networks (ICUFN)*. IEEE, 2018, pp. 798–800.
- [95] M. Vetterli and J. Kovacevic, *Wavelets and subband coding*. Prentice-hall, 1995, no. BOOK.
- [96] M. N. Do and M. Vetterli, “Framing pyramids,” *IEEE Transactions on Signal Processing*, vol. 51, no. 9, pp. 2329–2342, 2003.
- [97] “Find the nuclei in divergent images to advance medical discovery, howpublished = <https://www.kaggle.com/c/data-science-bowl-2018>, note = .”
- [98] “A Multi-organ Nucleus Segmentation Challenge, howpublished = <https://monuseg.grand-challenge.org/>, note = .”
- [99] “Triple-Negative Breast Cancer Database, howpublished = <https://rgcb.res.in/tncbdb/>, note = .”
- [100] L. R. Dice, “Measures of the amount of ecologic association between species,” *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.

- [101] P. Vianna, R. Farias, and W. C. de Albuquerque Pereira, “U-net and segnet performances on lesion segmentation of breast ultrasonography images,” *Research on Biomedical Engineering*, vol. 37, no. 2, pp. 171–179, 2021.