Deep Learning Based Segmentation of Multiple Tissue Structures in Histology Images for Predictive Cancer Analytics

By

Afia Rasool
Fall 2018-MS(CS)-00000276977

Supervisor

Dr. Muhammad Moazam Fraz
Department of Computing

A thesis submitted in partial fulfillment of the requirements for the degree of Masters in Computer Science (MS CS)

In

School of Electrical Engineering and Computer Science, National University of Sciences and Technology (NUST),
Islamabad, Pakistan.
(August, 2020)

Approval

It is certified that the contents and form of the thesis entitled "Deep learning based segmentation of multiple tissue structures in histology Images for predictive cancer analytics" submitted by AFIA RASOOL have been found satisfactory for the requirement of the degree.

Advisor: Dr. Muhammad Moazam Fraz

Signature: Moazam Fraz

Date: 27-Jul-2020

Committee Member 1:Dr. Muhammad Shahzad

Signature: M. SHAHZAD

Date: 27-Jul-2020

Committee Member 2:Dr. Qaiser Riaz

Signature: Qaiser Riaz

Date: 28-Jul-2020

Committee Member 3:Dr. Arsalan Ahmad

Signature:

Date: 28-Jul-2020

Thesis Acceptance Certificate

Certified that final copy of MS/MPhil thesis entitled "Deep learning based segmentation of multiple tissue structures in histology Images for predictive cancer analytics" written by AFIA RASOOL, (Registration No 00000276977), of SEECS has been vetted by the undersigned, found complete in all respects as per NUST Statutes/Regulations, is free of plagiarism, errors and mistakes and is accepted as partial fulfillment for award of MS/M Phil degree. It is further certified that necessary amendments as pointed out by GEC members of the scholar have also been incorporated in the said thesis.

Signature:______________________________

Name of Supervisor: Dr. Muhammad Moazam Fraz

Date: 27-Jul-2020

Signature (HOD):_________________________

Date:___________________________________

Signature (Dean/Principal):_________________

Date:___________________________________

Dedication

Dedicated to my Motherland.

Certificate of Originality

I hereby declare that this submission titled "Deep learning based segmentation of multiple tissue structures in histology Images for predictive cancer analytics" is my own work. To the best of my knowledge it contains no materials previously published or written by another person, nor material which to a substantial extent has been accepted for the award of any degree or diploma at NUST SEECS or at any other educational institute, except where due acknowledgement has been made in the thesis. Any contribution made to the research by others, with whom I have worked at NUST SEECS or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except for the assistance from others in the project's design and conception or in style, presentation and linguistics, which has been acknowledged. I also verified the originality of contents through plagiarism software.

Author Name: AFIA RASOOL

Signature: Afia Rasool

Acknowledgement

I pay my gratitude to Allah almighty for His abundant help to finish this task. I am thankful to Dr. Muhammad Moazam Fraz for his sincere and timely guidance throughout my research journey. His intellect as well as moral support has been a huge source of inspiration for me. I extend my thanks to the teachers of my department for providing me with the academic base which paved a way towards completion of my research. My utmost gratitude and prayers go to my parents and to both of my brothers for encouraging me to execute my best throughout my life and my educational career.

Table of Content

List of Abbreviations

MVD	Micro-Vessel Density
LVI	Lymphovascular invasion
PNI	Perineural Invasion
CNN	Convolutional Neural Network
H & E	Haematoxylin-Eosin
IHC	Immunohistochemistry
ROI	Region of Interest
RBCs	Red Blood Cells
NVSSD	Nerve and micro-Vessel Semantic Segmentation Dataset
WSI	Whole Slide Image
RGB	Red Green Blue
ASPP	Atrous Spatial Pyramid Pooling
FPN	Feature Pyramid Network
FCN	Fully Convolutional Network
FAIR	Facebook AI Research
ASAP	Automated Slide Analysis Platform
GIMP	GNU Image Manipulation Program
LSTM	Long Short Term Memory

List of Tables

List of Figures

List of Graphs

Abstract

According to a survey of global health metrics, cancer is one of top five leading lethal diseases around the world. It has the capability to proliferate into other parts of body and often recurs again after the treatment or removal from body. This immensely growing issue is now-a-days a hot topic among pathologists and researchers community. Among the analytic factors to study tumor aggressiveness and disease recurrence, density of micro-vessels (MVD), Lymphovascular invasion (LVI) and Perineural Invasion (PNI) are considered key prognostic factors. The manual identification of micro-vessels and nerves is time consuming, laborious and highly prone to human error. Computational pathology is an emerging field striving to improve patient care by incorporating modern algorithms to the traditional analysis procedures of microscopic slides. To overcome the challenges of multi-scale, multi-shape and slight intensity variant histopathology structures, a deep neural network based hybrid semantic segmentation architecture is proposed. It comprises the fundamentals of encoder-decoder structure with the essence of parallel path network. The framework is specifically designed to improve the accuracy by focusing mega to minor object details during every block of segmentation network. The encoder uses Multi-scale feature extraction block made up of ResNeXt Blocks. This organization is effective to encode coarse to fine grained features from all specifications and dimensions while limiting the number of learnable parameters. The decoder is a combination of feature fusion and feature erudition while step by step mapping them back to pixel map. Monte-Carlo Dropout based uncertainty maps are also generated at prediction time. The proposed architecture is trained and tested on generated Nerve and micro-Vessel Semantic Segmentation Dataset (NVSSD). The trained architecture outperformed the existing state-of-the-art networks like FCN, Unet, SegNet, Deeplabv3+.

Keywords: *Deep neural network, Computational pathology, semantic segmentation, multi-scale feature extraction*.

CHAPTER 1

Introduction

This chapter is an introduction to semantic segmentation, computational pathology, our problem statement and solution statement.

1.1 Semantic Segmentation in Computer Vision

Artificial Intelligence, a branch of Computer science aims at the development of intelligent machines that can artificially mimic the sense of interpretation showing cognitive capabilities, the way humans and other animals can do. Living organisms use five kinds of senses to grasp the information from their surroundings and the nervous system monitor the whole body conditions by integrating the collected information. The ability of an eye to see and the ability of brain neurons to interpret what is seen, is a natural phenomenon called vision. Similarly, the ability of intelligent machines to sense and comprehend the visual world by their own algorithms, is called Computer Vision. Computer vision comes under the umbrella of Artificial Intelligence. The machine acquires digital images as a source of visual data, are trained by using machine learning and deep learning models and take intelligent decisions based on what they have seen.

Deep learning is a vast collections of supervised, semi-supervised and unsupervised learning models based on Artificial Neural Networks. ANNs are the chains of mathematical connections made up of thousands of artificial neuron, intend to learn from examples. An artificial neuron, inspired by the idea of neurons of mammals' nervous system, is a simple mathematical function that receives input from previous neuron, weight it according to the learning mechanism being followed and pass on to the neuron of next level. Now-a-days, deep learning algorithms are so capable to precisely accomplish all of the image processing tasks necessary for a comprehensive visual insight. Image processing comprises of four main tasks that are Image Classification, Object detection, Image description and Object segmentation shown detailed in Figure 1.1.

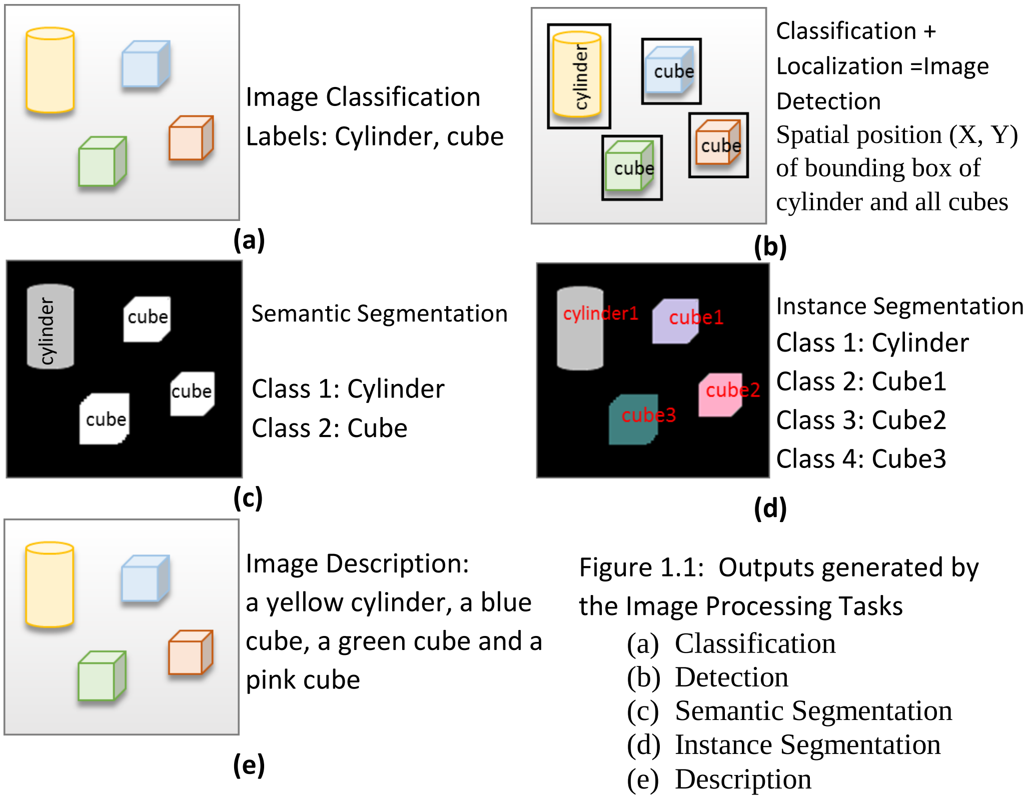


Figure 1.1: Outputs generated by the Image Processing Tasks

(a) Classification
(b) Detection
(c) Semantic Segmentation
(d) Instance Segmentation
(e) Description

Image classification task involves the prediction of objects within an image. Labels describing the object classes are produced as output in this model. In image classification's advanced form, the objects along with their exact spatial locations in the image are predicted, known as Image Detection. Then comes the Image description task, in which sentence based descriptions are depicted by understanding the objects of image. Image segmentation is the per pixel generation of class wise labels. It is also known as dense classification in which class based masks of the spatial resolution similar to that of input images are generated. Image segmentation task is further classified into two types. Separate labels for distinct classes are generated against every pixel called semantic segmentation whereas Similar classes with different instances generate separable labels, known as Instance segmentation.

Among the above discussed tasks of digital image processing, this research is keenly based on semantic segmentation. Instead of producing labels or categories against input images, it involves clustering various parts of image together that belong to specific object class. By retaining the location information exactly similar, every pixel is bounded to a certain class of objects. Semantic segmentation can be a binary or multi-class problem. If the task is to detect pixels of only one object class, it is a binary semantic segmentation task. The concerning regions would be labelled as 1, whereas the rest of the spatial space would be considered as background given the labels of 0, hence generating a binary mask of ones and zeros, similar in size to that of input image. Whereas if greater than two classes are needed to be semantically segmented against each image, it is a multi-class problem. In it, the predicted mask would be grey scale with similar height and width and with label count equal to the number of classes starting from 0 as background. Most common benchmark datasets to train and test a semantic segmentation technique are PASCAL VOC, ADE20K and Cityscapes. Mean Intersection-over-Union and Pixel Accuracy are two standard evaluation matrices used for evaluation of results.

1.2 Computational Pathology

Computational Pathology is an emerging approach to diagnosis that uses computational models at sub-cellular, individual and population levels, utilizing clinically important information extracted from pathological and radiological imaging, laboratory data and electronic medical records, producing improved diagnostic implications and predictions, empowering health care system and physicians to make promising medical decisions. Since 17th century, pathologists work integrally on histopathology and till now, the process is mainly based on manual brightfield microscopic examination by professional of the field. However, this process is time taking, tedious and imprecise due to the involvement of human error. Advancement in technology led the digitization of microscopic slides and right now, we are living in the era of Artificial Intelligence and Machine learning. Hence moving from traditional pathology towards computational pathology.

Traditionally, workflow of the digital pathology starts with the histopathology data preparation that includes several steps. First of all, patient's sample is acquired in the form of tissues (i.e. Biopsy) which are fixated onto glass slides. These prepared slides are stained with multiple chemicals to enhance their features visualization by improving color contrasts. The most common type of discoloration is the pigment staining, commonly known as Haematoxylin-Eosin(H&E) where Hematoxylin stains the DNA portion of cell (nucleus) bluish and Eosin turns rest of the sub-cellular structures (cytoplasm) pinkish in color. The task is also done by using antibodies, for example Immunohistochemistry (IHC) and immune-fluorescence labeling that use antibodies binding principle for selectively isolating any classified proteins or antigens present in tissues. Afterwards, these slides are digitally scanned at very high resolution by Whole slide scanners resulting in the creation of WSIs. Digital WSIs are most commonly stored in the TIFF, JPEG 2000, Leica SVS file formats with file size of typically more than 1GB. The histopathology images undergo labeling of Regions of interest (ROIs) called ground truths and processed

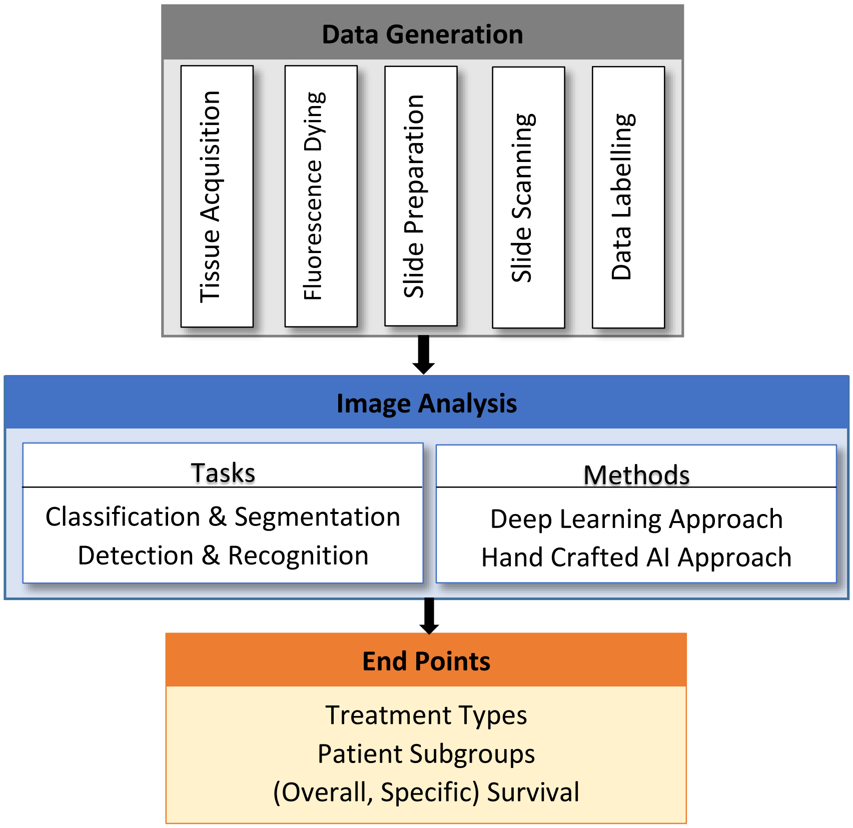


Figure 1.2 Computational Pathology Workflow

by using Hand crafted approach or deep learning approach [1]. The computational pathology workflow is shown in Figure1.2.

Deep learning (supervised or unsupervised leaning) algorithms are being used now-a-days to carry out numerous tasks like tumor grading, invasive versus in-situ classification, mitotic count, nuclei count, molecular localization within image, and segmentation of various pathological structures. The end results can be used to study specific to overall survival rates, to classify patient into subgroups and to adopt treatment types.

1.3 Areas of Application:

Computational pathology is impacting all fields of medicine resulting in

- Faster detection of disease and its current phase
- Precision Medicine
- Personalized treatment
- Therapy Acceleration or Deceleration
- Aftermaths of treatment

1.4 Predictive Cancer Analytics

Cancer, a term referred to a large group of diseases described as the uncontrollable and abnormal cell development that penetrates and kills normal cells of body. It can proliferate into other parts of body and often recurs again after the treatment or removal from body. Cancer is one of the five leading lethal diseases around the world. There are more than 100 types of cancer, also known as malignancy. Due to the seriousness of disease, researchers over the past few years are striving to lessen its impact on mankind.

One of the key directions to control cancer is to halt its growth to other body parts by subsiding the means tumor cells adopt to spread. The mico-environment of tumor comprises stroma, abnormal cells, immune system cells, lymphatic and blood vessles and nerves. Among these bodies, last two are most important prognostic factors when studied about tumor aggressiveness and disease recurrence. Studies have shown that micro-vessel density (MVD) is considered an important indication of Neoangiogenesis, metastasis formation and chemotherapy effects [2]. Neoangiogenesis is the characteristic of tumor cells that ensures the supply of food and oxygen by creating new blood micro-vessels. The abnormal cells can also leap into the blood circulatory system through these blood or lymphatic vessels called Lymphovascular invasion (LVI) [3] and hence penetrate somewhere else. Another way metastasis can develop is by invading nearby nerves, called Perineural Invasion (PNI), also result in difficult surgery-based tumor removal and recurrence of disease [4]. Hence PNI, LVI and MVD identification is worth doing to normalize cancerous cells. Till now, Nerves and micro-vessel identification is done manually but the process is time-consuming and subjective.

The method we intend to choose for semantic segmentation is deep learning based using the supervised learning technique. Supervised machine learning technique uses ground truths at the time of training to learn and predict in future. The semantic segmentation general pipeline includes each block that is responsible for compression of spatial resolution and progressively increasing the dimensionality (encoding of features). The other block receives the discriminative features over low resolution and semantically maps onto the original high resolution pixel space (decoding of features). A general depiction of a semantic segmentation pipeline is shown in Figure 1.3.

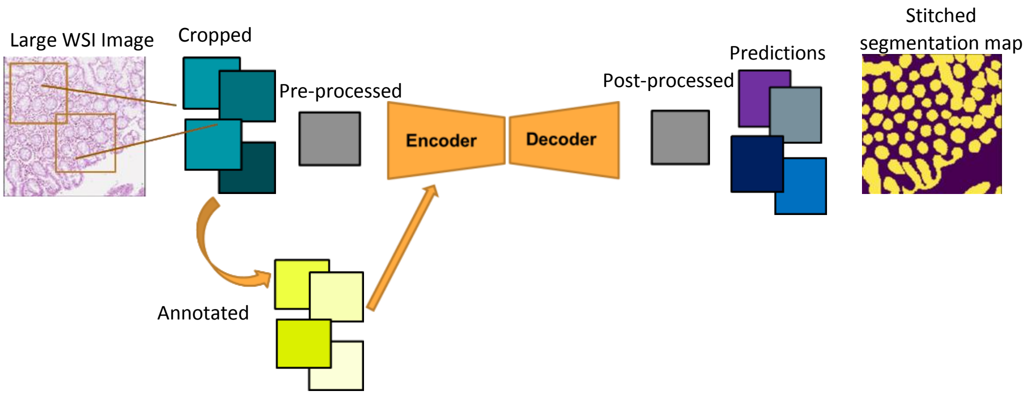


Figure 1.3: Deep learning based general segmentation pipeline

1.5 Challenges

- Deep learning based model accuracy is highly dependent on what they take on as example to learn from. That is why, an adequate amount of data along with precise ground truths is a must to get accurate predictions. In our case, the dataset of H&E stained WSIs for micro-vessels and nerves is very rare.

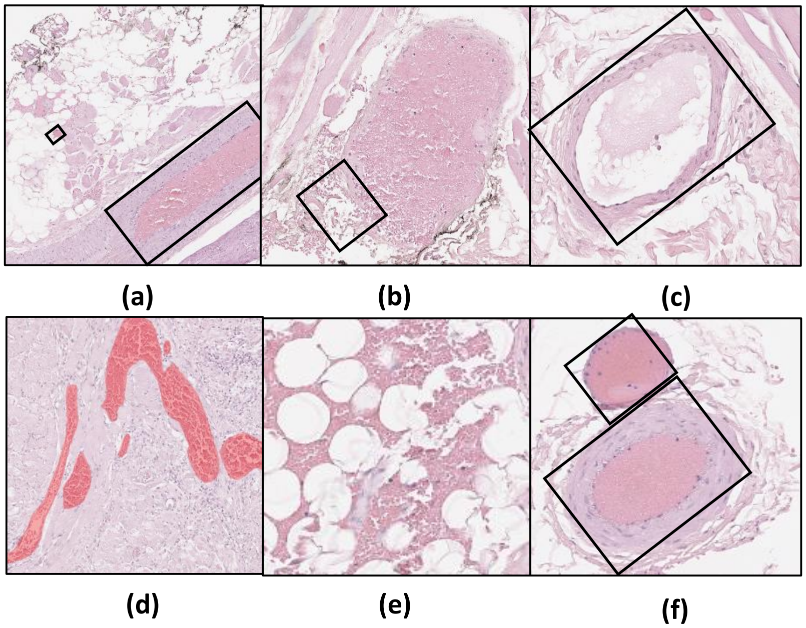


Figure 1.4: Challenging examples for segmentation task

Moreover, there is no publically available dataset for simultaneous semantic segmentation of nerves and micro-vessels.

- The occurrence density of micro-vessels and especially nerves is low limiting the instances of structures. The manual annotation generation is extremely laborious, time consuming and accuracy oriented.
- Vessels in histopathology images are characterized by the presence of Red Blood Cells (RBCs) and model would consider the presence of RBCs an essential feature to classify vessels. But sometimes, RBCs can peep out of vessel boundaries during slide preparation or due to some other injury, resulting in confusing the model to classify them as vessels too. The case is shown in Figure 1.4 (e).Moreover, empty vessels that occurs with strong boundaries but have no RBCs are clinically unimportant and must be ignored by the model, Figure 1.4 (c).
- Nerves and Micro-vessels vary in their size and appearance a lot. The structure can be as small as penetrating within few pixels to as large as occupying the whole sample image as shown in figure 1.4 (a) and (d). The

arteries are characterized as structures with thick prominent boundary walls whereas Veins have thin walls, Figure 1.4 (f). Even the shapes of micro-vessels does not appear alike.

- The intensity variation among multiple structures in routine H&E histology images is low hence color based segregation of classes become impractical.
- Other tissue structures like tumor keratinization looks very similar to RBCs, making the accurate segmentation harder.

1.6 Thesis contributions

This research intends to take part in the improvement of the field of Computational Pathology. The key contributions are:

- A novel encoder decoder based architecture has been proposed for image semantic segmentation task.
- Nerve and micro-vessel semantic segmentation dataset (NVSSD) has been prepared using histo-pathological WSIs of oral cell carcinoma tissues.
- Current state-of-the-art architectures as well as the proposed approach has been evaluated on NVSSD. Evaluation measures of the method revealed a significant elevation in performance compared to the former ones.

1.7 Thesis Organization

This thesis write-up is organized chapter-wise. Chapter 2 is based on the current research done in the field of histology segmentation. Our proposed methodology is explained in chapter 3. Chapter 4 is all about steps and tools used for generation of NVSSD. The method training and evaluation results and ablation study on dataset is discussed in Chapter 5.Chapter 6 provides conclusion and future work directions.

CHAPTER 2

Literature Review

This chapter is based on the discussion of prominent research done in the classification and segmentation of various structures in histopathology images. Nuclei, Glandular bodies, tumor, stroma regions, Nerves and vessels are the worth considering structures present in WSIs, that could become the source for improvement in medical science [5]. A significant amount of research has been done over past few years based on Hand crafted and Deep learning Algorithms.

2.1 Hand Crafted feature based Algorithms

Hand crafted feature based algorithms are the task based custom algorithms used in Image processing. For example detection of corners and edges, or extraction of lines of specific length or objects with certain properties or texture. Such algorithms are designed by adding layers or steps according to the requirement. There outcomes are not versatile in nature and to design such algorithms, deep knowledge of task should be there. A simplest algorithm based on such technique is believed to be the edge detector where focus of framework would be the mathematical calculation and plotting of sudden intensity changes in the image.

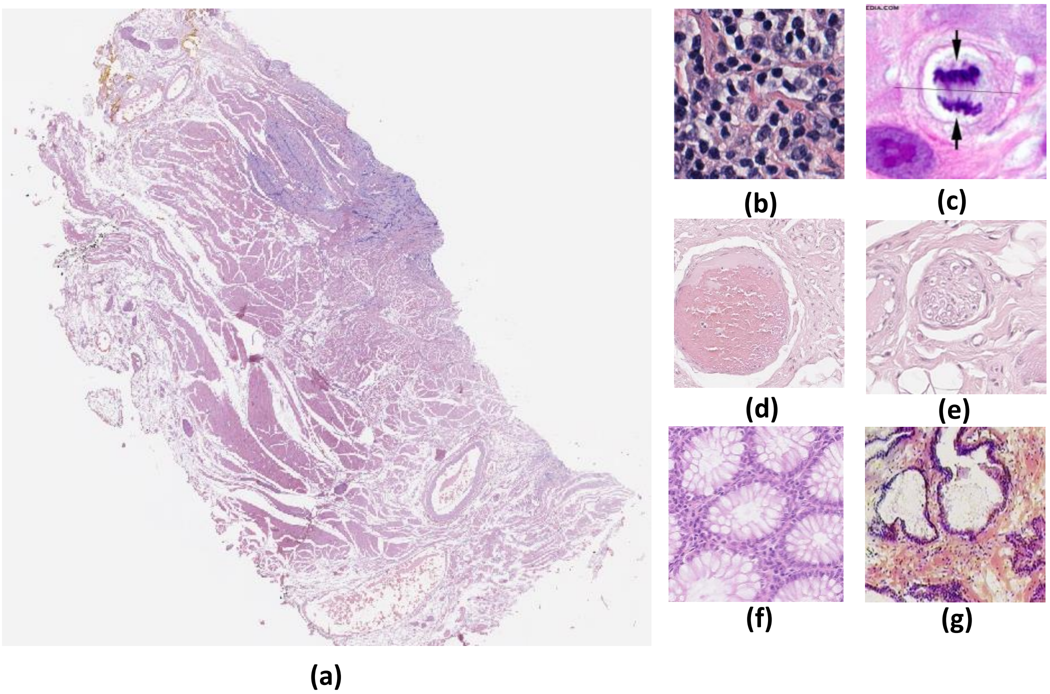


Figure 2.1: Showing a complete WSI (left) and it's micro-environment components (right) which are important in research point of view; (a) WSI (b) nuclei (c) mitosis (d) vessel (e) nerve (f) gland (g) stroma

Some of the segmentation of histopathology structure have done using hand-crafted algorithms. A novel marker-controlled watershed algorithm is proposed in [6] to successfully segment clustered cells by implying region-based approach. The same task of cell segmentation in WSI images is achieved by using morphological and thinning algorithms [7]. Entropy thresholding is a technique that uses the intensities histogram plot to choose the optimum threshold value for certain class of objects. Effective monocyte cells segmentation is achieved by using entropy thresholding in bipartite graph matching algorithm [8]. A multi-class-multi-scale-series-contextual-model is introduced in [9] for dense classification of cell nuclei in H & E stained images. In this approach, Supervised contextual based learning algorithm extracts multi-scale and multi-object information from the pixels of images. The similar several scale problem is addressed by using Gaussian filtering and graph-based binarization automatic segmentation algorithm [10]. Implementation of techniques like boundary based segmentation [11, 12] and active contour detection [13] has also been considered essential before the arrival of CNNs.

2.2 Deep learning based Algorithms

Deep convolution neural network is considered the most successful technique in machine perception and computer vision, including classification, recognition, object detection, semantic segmentation and description of images. Recently, machine learning and deep learning algorithms have widely been used for classification and detection tasks in computational pathology as well [14]. A number of bench marks have been achieved while classifying nuclei, tumor cells and glandular structures in H & E stained images as well as some IHC stained images are also being used for the task with similar methods [15]. A typical pipeline of feature extraction and map generation is used for Image semantic segmentation task. However, The methods of deep learning can be further categorized into encoding-decoding based, Dilation based or their hybrid is also sometimes practiced in literature for achieving precise results.

2.2.1 Encoder-Decoder Networks

This type usually form an encoder decoder based architecture where mostly the layers are CNNs, followed by pooling. The pooling layer plays vital part in encoder to reduce the spatial resolution of images. Encoder captures feature information across Dimensions which are then reduced and fused in decoder as the spatial resolution is again recovered by up-sampling. Pooling layer with specific kernel size and stride of 2 pixels can reduce the resolution and only retaining either maximum values or average of values within pooling kernel. This spatial reduction is sometimes also done by applying normal convolution layer with stride of 2 pixels.

U-net [16] is one of the earliest example of pooling driven network for the segmentation of Biomedical images. This fast light weight architecture uses 2 convolution and activation layers followed by pooling layer with stride 2 and repeating the same configuration four times form compression block. Up-Convolution is applied for symmetric expansion and a last layer of 1 x 1 x n generates masks where n is the number of classes. Up-convolution is the term used in this research paper that consist of an upsampling layer and a 3 x 3 convolution layer.

Fully convolutional neural network is also used to accomplish the semantic segmentation task of micro-vessels in the H & E stained images for lung adenocarcinoma (ADC) [17]. Pooling based encoder and Deconvolution based decoder with lateral connection from encoder makes the pipeline for segmentation. Another widely used network called segNet [18] proceeds by pooling in encoder. However no fully connected layers are used. Instead encoder pool indices are fed to each decoder stage and densified by trainable convolution filter banks.

Pooling layer works best when training resource constraints are there and the objects are multi-scaled. It can easily cut off large number of trainable parameters and passes only the important information to the next level. But when it comes to the accurate pixel wise classification of the whole image, it fails especially over the boundaries of objects because of the loss of exact locations.

2.2.2 Multi-path Networks

To address the issue of exact-location loss across the pipeline and to fetch features of variant sizes, Dilation based networks were proposed. Parallel convolution paths are adopted with different dilation rates called atrous convolutions. This arrangement can overcome the need of compression-expansion blocks, as the feature maps are never compressed. Dilation rate is basically the number of pixels each unit of kernel would ignore while doing convolution. This way, a smaller dilation rate would address smaller objects and bigger rate would catch up the features larger in scale, while retaining the original spatial resolution same throughout the pipeline. The outputs of these parallel pipelines are then fused together to create final masks.

Deeplab, a series of semantic segmentation models based on atrous convolutions is proposed by Google [19-22]. Deeplabv1 employed atrous convolution of certain rate and Conditional Random Field (CRF) to address the issues of feature resolution and localization accuracy. The need to insert rescaled variants of same image to capture multi-scaled features was tackled in Deeplabv2 by introducing parallel atrous convolutions at different rates and merging them together. In Deeplabv3, the same concept was used along with separable convolutions where a depth-wise convolution specifically perform conventional convolution over every channel independently and a point-wise convolution is used later that combines the features over the depth for each pixel. The latest version is Deeplabv3+ with encoder block similar to the previous version and employment of advanced decoder where the low level and high level feature maps undergo point-wise convolution and then joined together for dense classification.

Another approach for not using encoder and decoder is RefineNet [23]. In this method, multi-resolution feature maps are gained by applying Resnet over multiple paths, undergo element wise summation in fusion block. A chained pooling is applied afterwards and all the features are merged back using residual connection after every block of pooling. A final layer apply 1 x 1 convolution with the number of dimensions equal to the number of classes in image. This technique overcomes the issue of down-sampling and keeps the computation cheaper compared to atrous based convolutions.

2.2.3 Hybrid Networks

Hybrid networks use the characteristics of both the encoder-decoder framework along with atrous convolutions. FABnet, a recent framework has been proposed in [24] for simultaneous semantic segmentation of nerves and micro-vessles in histopathology images. This architecture uses encoder based on Xception residual blocks, further enhanced by incorporation of atrous spatial pyramid pooling. The decoder takes on as input the concatenated feature map generated by ASPP block. The skip connections between encoder and decoder are equipped by feature attention blocks to drive the model's attention towards prominent features, as the decoder propagates towards class-wise binary masks generation. Another methodology is employed using aligned Xception model along with pyramid pooling and a light weight decoder for semantic segmentation of micro-vessels in H&E stained WSIs [25].

2.3 Critical Analysis

A significant amount of research has been done in the digital world of pathology and a diverse range of approaches has been proposed. But a point worth mentioning here is, the task is relatively different from the classification of general objects and scenes of daily world. The main reason behind this is, first the color intensity variations among various structures in WSIs is quite low, making it difficult to classify on color basis. The second reason is the level of cropping WSIs undergo before integrating into machine is high, resulting in failure of the visualizing of each whole object within each test sample. Each approach designed, tested and out performing on general datasets like Pascal Voc and Cityscapes, is likely to fail performing the task with similar accuracy on histopathological images and vice versa.. Research on segmentation task for microvessels, besides it's great importance is limited and to the best of our knowledge, FABnet is the only approach for the simultaneous segmentation of nerves and micro-vessels in literature till now due to limited amount of datasets publically available for the task. That is why, a novel approach as well as a new dataset for simultaneous segmentation of nerves and micro-vessels have been prepared to help researchers explore new direction in the cancer prognosis in future.

CHAPTER 3

Methodology

This chapter explains the methodology proposed for image semantic segmentation task. The segmentation pipeline first needs to extract important features, which is done by encoder part that generally results in the reduction of height and width of image but increase in the number of feature maps as we go deeper in the layers. The second part of pipeline is required to bring the extracted features back to the original resolution of image called the decoder. Last layer calculates per pixel probability for each class and generates the segmentation mask against original image. Our proposed pipeline contains four blocks where output of each block becomes the input for the next one.

First of all, the original WSIs are preprocessed which is explained in chapter 4 in detail. RGB images at 10x resolution and with the size of 256 x 256 are inserted in the first block of architecture. Multi-Scale Feature Extraction Block is used as encoder to extract features at multiple resolutions. Then comes the first block of decoder part called feature fusion block. This block up-samples and concatenates

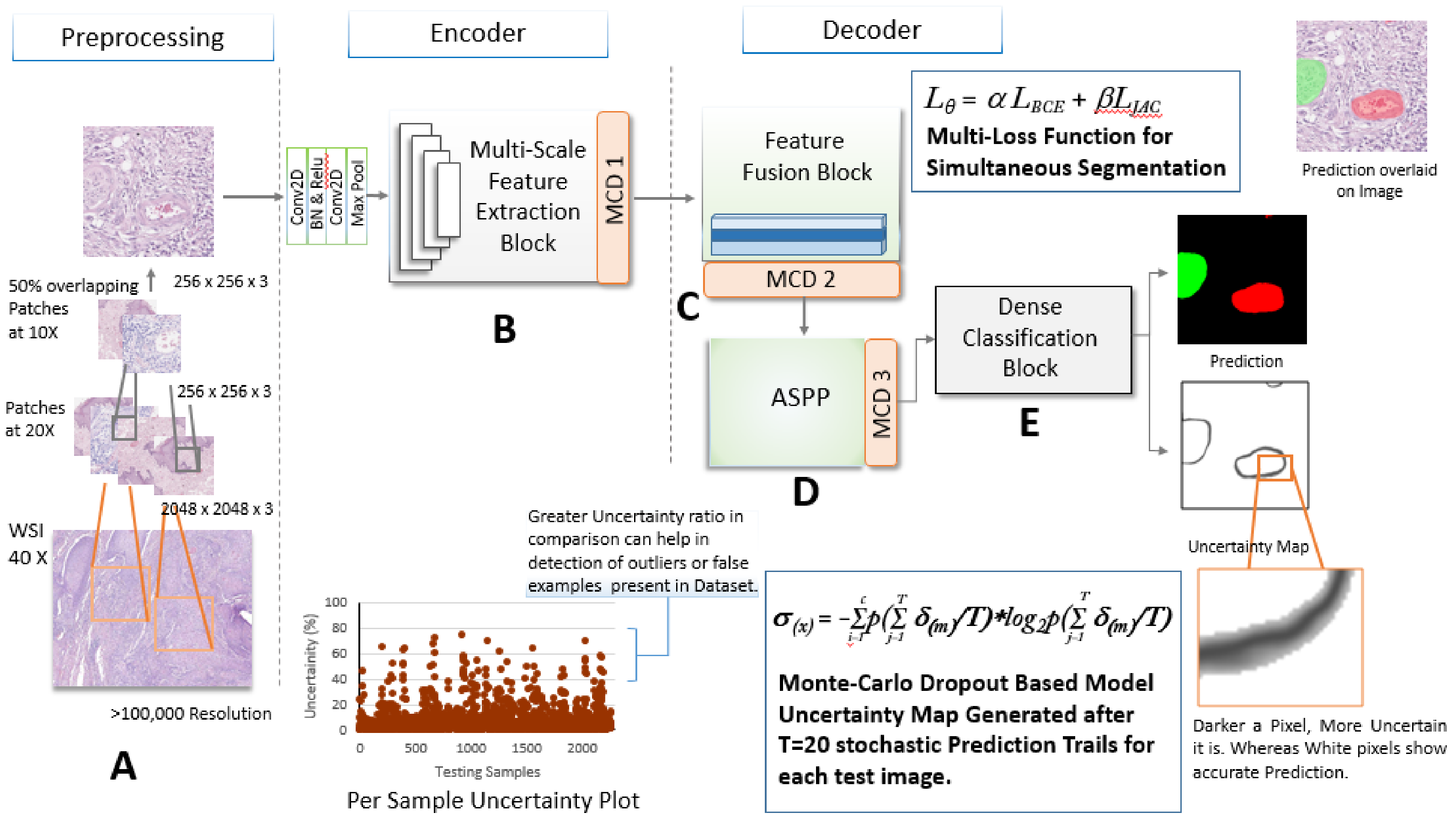

Preprocessing
Encoder
Decoder
50% overlapping Patches at 10X
256 x 256 x 3
Patches at 20X
256 x 256 x 3
2048 x 2048 x 3
WSI 40 X
>100,000 Resolution
A
Conv2D
BN & Relu
Conv2D
Max Pool
Multi-Scale Feature Extraction Block
MCD 1
B
Feature Fusion Block
MCD 2
C
ASPP
MCD 3
D
Dense Classification Block
E
$L_\theta = \alpha L_{BCE} + \beta L_{JAC}$
Multi-Loss Function for Simultaneous Segmentation
Prediction overlaid on Image
Prediction
Uncertainty Map
Darker a Pixel, More Uncertain it is. Whereas White pixels show accurate Prediction.
$\sigma_{(x)} = -\sum_{i=1}^{c} p(\sum_{j=1}^{T} \delta_{(m)}/T)*log_2 p(\sum_{j=1}^{T} \delta_{(m)}/T)$
Monte-Carlo Dropout Based Model Uncertainty Map Generated after T=20 stochastic Prediction Trails for each test image.
Greater Uncertainty ratio in comparison can help in detection of outliers or false examples present in Dataset.
Uncertainity (%)
100
80
60
40
20
0
0
500
1000
1500
2000
Testing Samples
Per Sample Uncertainty Plot

Figure 3.1: Layout of proposed methodology

the features of multiple scales into each unit and feed this to ASPP block. Atrous Spatial Pyramid Pooling is used here to mature the feature selection process by applying separable convolutions at multiple rates. Features from all scales would again be concatenated and passed through the final layers of architecture for the generation of segmentation masks. Model uncertainty is produced using Monte-Carlo dropout layers during test time which is later used to plot general certainty trend of model. The basic layout of pipeline is shown in Figure 3.1.

3.1 Multi-scale feature extraction Block

Feature Pyramid is a term used for stacked layers with similar resolution and specific number of feature maps at certain stage. Each pyramid level generates feature maps of alike resolution and dimensionality. The next pyramid would be half in resolution (X and Y) but carry semantically stronger features compared to the former stage. This way, as the network goes deeper and the number of pyramid stages increase, feature maps with lower resolution but vital features are achieved. The technique called Feature Pyramid Network (FPN) was first proposed by Facebook AI Research (FAIR) [26] and is incorporated as encoder in our proposed architecture. The number of pyramids used in Multi-scale feature extraction Block are four. The encoder starts with taking up image with the resolution 256 x 256 x 3. Every next pyramid starts with halving the resolution and doubling the number of feature maps. Each stage's last layer would form lateral connections in feature fusion block. The encoder ends up with a 2048 dimensional feature maps.

The stacked layers with in each stage comprises of convolutional layers with similar resolution. Inner layers use ResNeXt Blocks (RBs) inspired by [27], originally proposed as the extension of residual networks. RB partially utilizes the technique of grouped convolutions. Each layer of convolution is split into number of groups called Cardinality where each group is the similar version of the layer but with only limited number of feature maps. The split groups of feature maps under goes transformations like convolutions separately.

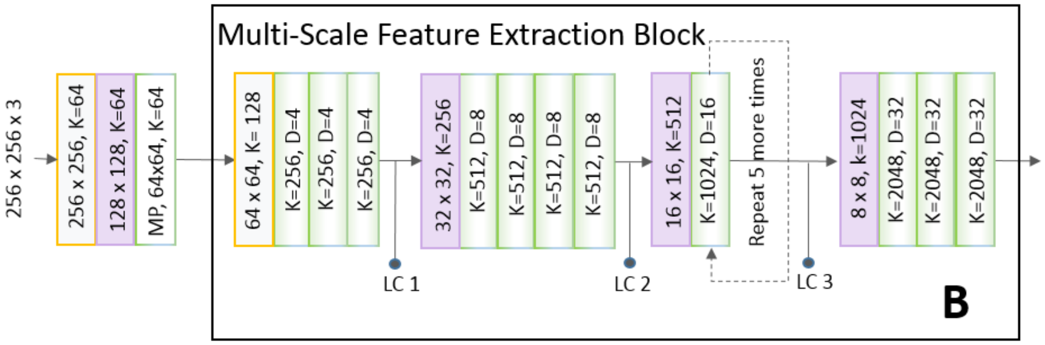


Figure 3.2: Encoder of architecture

Then all the groups are merged together to form single feature map. The original layer coming from the previous layer not only get split, transformed and merged back but also added to the concatenated map. Hence the effect of previous layer is also retained in the current layer and is being passed over to the next layer. This residual connection is helpful in the faster optimization of network.

The reason to use the ResNeXt blocks as backbone is to reduce the number of parameters and saving the resources while retaining the effects of normal convolutions. For example, let’s consider we want to perform a 3x3 convolution on a layer having 512 feature maps. This convolution would consume handsome amount of processing with 3 x 3 x 512 kernels. In contract to this, if the original layer with 512 features is reduced to C=32 groups of 16 feature maps each. The precious 3x3 convolution would take 32 times but with the kernel size of 3 x 3 x 16. It would result in faster convolution with reduced complexity. Moreover, it has been studied that when the convolutions are done in groups, every group learn features with different specializations resulting in more versatile collection of features generating stronger representation. The cardinality is considered the fourth dimension of layer and works like generating Network with Neurons and control the number of complex transformations.

In our network, resNeXt 50 configurations are being merged with FPN to acquire features of all scales with greater accuracy and less training time. This is the reason that our method converged within first 10 training epochs. The layers used are (3, 4, 6, 3) according to pyramids and cardinality is consistently kept 32. The feature maps during splitting is kept while considering D= K/(C *2). One pyramid is converted to next one by applying a 3 x 3 convolution with stride 2. Detailed arrangement of encoder layers and ResNeXt Block is shown in Figure 3.2 and Figure 3.3.

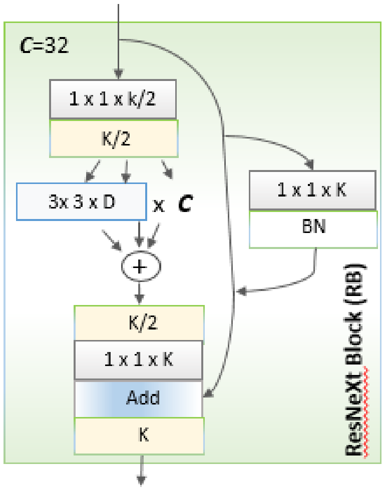


Figure 3.3: ResNeXt Block

3.2 Feature Fusion Block

This block is responsible for combining low resolution features containing fine-grained details and complex patterns, with the global boundaries over higher resolution, generating precise mid-level segmentation feature maps. Although this block final output resolution is less than the original image size, but still it has the capability to penetrate mega to minor details of the original image. This block initial layers work like bottom-up pathway as shown in Figure 3.4. The block starts with the last map of encoder block, it's channel dimensions are reduced to 256 by 1 x 1 convolution and Bilinear Interpolation with the factor of 2 is applied to up-sample it. The resultant map undergoes element-wise addition with LC3 layer. This way all the nodes of encoder i.e. LC1, LC2, LC3 perform reduction of feature map to

256 dimensions, added and up-sampled. Further, the 3 added layers separately perform 3 x 3 convolution , Batch Normalization and Relu and concatenated together to form a feature map of 64 x 64 x 1024.

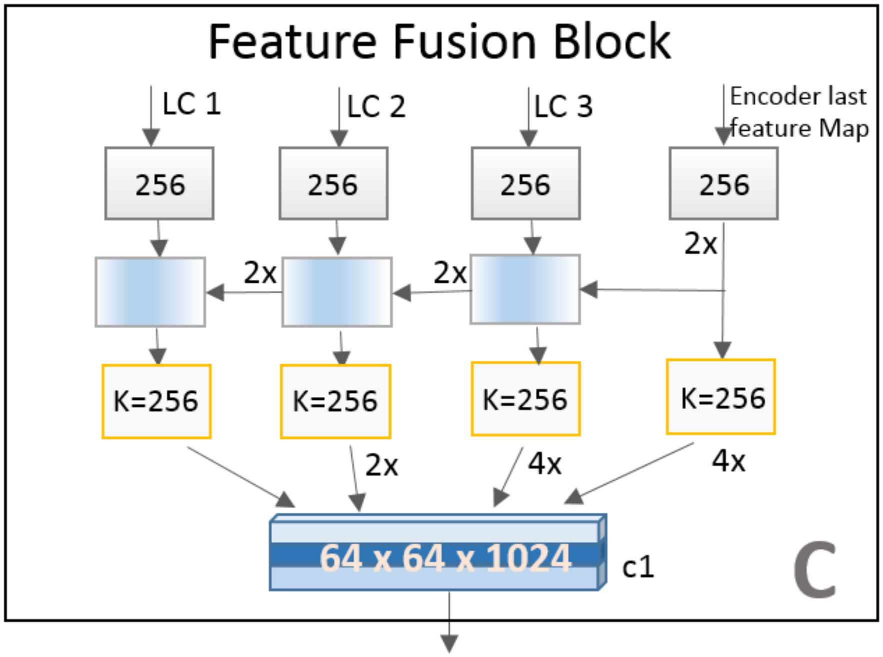


Figure 3.4: Feature Fusion Block

3.3 Atrous Spatial Pyramid Pooling (ASPP)

As discussed in the challenges section, the biggest problem of micro-vessel and nerve segmentation is that they vary in scales a lot. A structure can be a few pixels detectible to almost occupying whole sample image. This situation makes the leaning process tough and a need to deal with all scale features is obvious. ASPP [20] has the ability to trace multiple scales in parallel fashion. It applies 3 x 3 convolution at multiple dilation rates over similar feature map simultaneously, hence accounting for different object scales. The basic concept of ASPP is to apply parallel atrous convolutions at tune able rates and their fusion back to single layer. The block when used in decoder can enhance accuracy to much extent. Moreover, it can enhance the boundaries of objects by applying convolution over up-sampled features.

Input to the ASPP block is c1, the dilation rates applied are 4, 8, 12 (drawn by experiments) and followed by point wise convolution to merge features along all dimensions. The final point wise convolutions are used to reduce dimensionality of each block to 128. Global Average Pooling (GAP) is also a path of ASPP block. It is used to pick up most prominent features with higher pixel intensity across each tensors. A one dimensional array of size 1024 is generated where each entry represents each spatial tensor. It undergoes up-sampling and all the paths are merged and passed onto final layers of segmentation pipeline. Figure 3.5 shows layers used in ASPP and dense classification block and the legends explaining the layer convention id given in Figure 3.6.

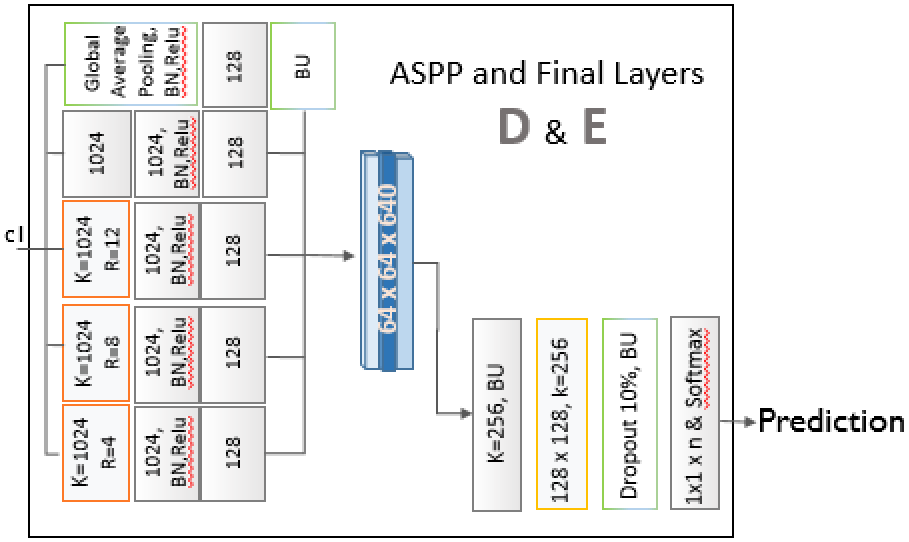


Figure 3.5: ASPP and final segmentation layers

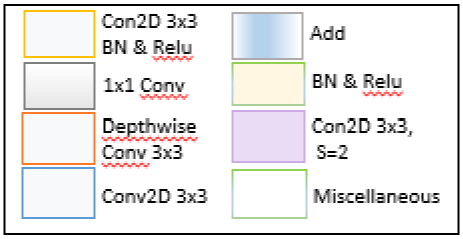


Figure 3.6: Legends

3.4 Dense Classification Block

Dense classification block is a symmetry of final layers including two Biliear up-sampling layers with 3 x 3 convolutions in between, a dropout layer and final 1 x 1 convolution layer with filters equal to the number of classes. Our problem comprises of 3 classes; micro-vessels, nerves and background class. Input to this section is the concatenated output feature map of ASPP block. A dropout of 10 % is applied to regularize the learned weights and avoiding over fitting. At the end, softmax classifier is applied to calculate probability for each class. A pixel space of 256 x 256 with each pixel comprising a class label either 0, 1 or 2 is gained displaying the segmentation.

3.5 Model Training and Optimization

NVSSD dataset is used for training, validation and testing of proposed model. The ratio of training, validation and testing data is kept 70 %, 15 % and 15% accordingly. Input shape is 256 x 256 x 3 and segmenting classes are 3.The model is trained for 25 epochs with batch size of 4. Optimization technique used is Adam with initial learning rate 0.001. The learning rate is decaying with a drop of 0.1 every 8th epoch. Binary Cross Entropy and Jaccard Loss based multi-Loss function is used for better optimization and 10 % dropout is used in CNN for regularization purpose.

Multi-loss function is calculated as

$$L_{\theta} = \alpha L_{BCE} + \beta L_{JAC} \tag{3.1}$$

Where α is the weight of binary cross entropy and β is the weight of jaccard loss. Here weight for both losses is kept 50 % each. The Binary Cross Entropy (BCE) formula is

$$L_{BCE} = -\frac{1}{N}\sum_{i=1}^{N} Y_i \left(\log (Y_p) + (1-Y_i) \log (1-Y_p)\right) \tag{3.2}$$

Here N is the number of pixels, Y_i is the ground truth and Y_p is predicted value. The other loss function Jaccard loss is given as

$$L_{JAC} = 1- \frac{2\sum_{i=1}^{N} Y_i Y_p}{\sum_{i=1}^{N} Y_i + \sum_{i=1}^{N} Y_p} \tag{3.3}$$

Where N is the number of pixels, Y_i is ground truth and Y_p indicates predicted values.

3.6 Model Uncertainty Estimation

Deep neural networks can make perfect predictions showing the level of accuracy with very well-calculated formulae. But with what confidence should one trust it's predictions? Functions like softmax cannot calculate the degree of uncertainty while throwing the probabilities for each class. Model uncertainty is a distinct task telling the analyzer and the model to either trust the predictions or not. It can assist decision-making by providing a visualization of where the model is uncertain and a threshold to accept or reject the segmentation especially in the situations when the testing sample is far from what the model has learned to forecast. Two steps have been incorporated into our methodology to address the improbabilities. One is Uncertainty map and the other is per sample uncertainty plot.

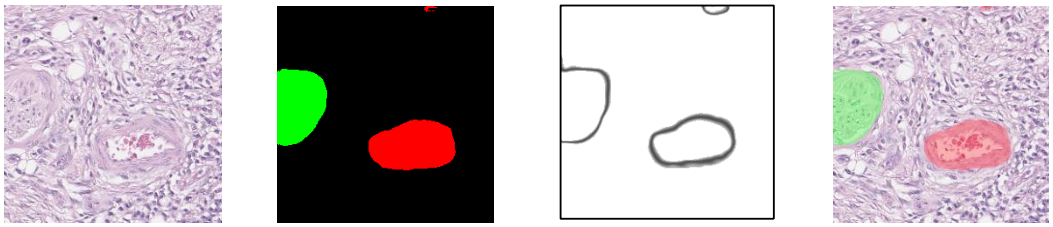
3.6.1 Uncertainty Map Generation

Generally the predictions are deterministic, that means with the exact configurations of architecture, the forecasting would be exactly similar no matter for how many times the similar data is passed through the pipeline. But a stochastic prediction would generate different results every time the similar data is tested with exact configurations.

Uncertainty map of individual example along with it's segmentation is generated during test time using Monte-Carlo Dropout originally proposed in [28]. Each layers of dropout inspired by Monte-Carlo Dropout is added after Encoder, Feature fusion block and ASPP with 50 % dropout to get stochastic-ness in architecture. During prediction, the testing data is divided into mini batches with each image in one batch. 20 stochastic predictions are produced for each testing image. There effects are accommodated and entropy is calculated for each class to get randomness in predictions. The model is separately trained and tested to get uncertainty maps. The formula for generating uncertainty maps is

$$\sigma_{(x)} = -\sum_{i=1}^{c} \left(\sum_{j=1}^{T} \delta_{(i)}/T\right) * \log_2\left(\sum_{j=1}^{T} \delta_{(i)}/T\right) \qquad (3.4)$$

X is the current image for which uncertainty map is being generated. T is the number of times stochastic predictions are generated and c depicts number of classes. In our case T is 20 and c=3. $\delta_{(i)}$ Shows the pixel wise predictions.



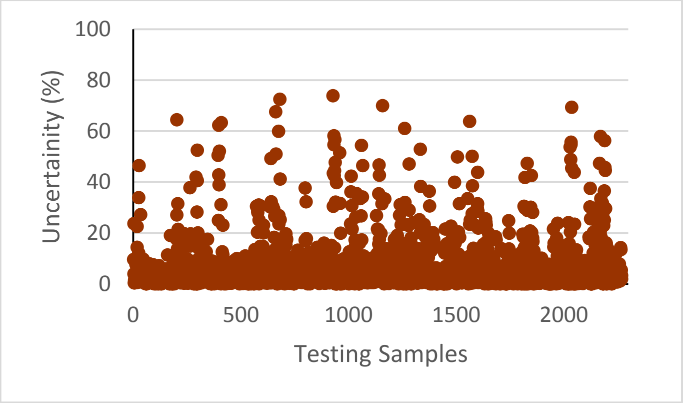
(a) (b) (c) (d)

Figure 3.7: (a) original image (b) predicted map
(c) uncertainty map (d) prediction overlaying original image

Figure 3.7 shows the uncertainty map generated against image. The white pixels in uncertainty maps show the accurate prediction whereas the grey pixels are the regions where the model gets confused. The darker the grey area, more uncertain it is and vice versa.

3.6.2 Per sample uncertainty plot

Uncertainty percentage for each image is calculated and plotted in the form of graph to visualize general trend of model being uncertain. The maximum number of testing sample uncertainty lies between 0-20 %. This can be the tolerance threshold for accepting or rejecting a predicted segmentation in the model. So this threshold can easily be guessed by simply having a look on the plot. Moreover, it shows overall model performance in one glance.



Graph 3.1: Per Sample Uncertainty Plot

A trend with majority certain examples and only a few lying out of the general trend is not indicating the incompetency of model. The information is also important and worth keeping. This is because the samples above general trend is showing the presence of outliers within dataset. The examples with uncertainty above 20 % are few in graph. By studying the behavior in depth, we came to know that these examples are either blur, with false ground truths or with some kind of color disorder, making them outliers, results in confusing the model.

CHAPTER 4

Materials

Semantic segmentation through supervised learning uses data in the form of X and Y as examples during training time to predict in future. Here X is the original RGB image of specific spatial resolution and Y is the grey scale mask of that image with similar spatial resolution indicating object class against every pixel. The procedure of such dataset generation in computational pathology starts by extracting patches of interest from original digitally scanned slides of tissues at certain resolution or level. These patches are saved in .jpg or .png file formats as individual 3 dimensional images and ground truths are drawn manually against them by using some annotating tools.

This chapter illustrates the steps followed and tools studied for generation of Nerve and micro-Vessel Semantic Segmentation Dataset (NVSSD). An introduction to histopathology images, dataset generation tools for visualization and annotation generation, and details of prepared dataset is being discussed in sub sections.

4.1 Whole Slide Imaging

Digital file creation by scanning a complete microscopic slide is known as whole slide imaging or digital microscopy. Such files can be used for storing, analyzing and sharing tissue slides by using technology. This technique is revolutionizing many field like pathology, genomics, proteomics and medical education. The advantage of digital slide over actual microscopic specimen is that it does not deteriorate over time, can be easily shared anywhere all over the world and comes with multiple resolution levels to let the analyzer choose any details rank.

Table 4.1: WSI resolution scales and levels

File Resolution	Tile Level	Down-sample Factor
40 x	0	Original / 2^0
20 x	1	Original / 2^1
10 x	2	Original / 2^2
5 x	3	Original / 2^3

A microscopic slide is scanned by using digital scanners like TissueScope and Vesalius. The scanning is usually done in small high resolution stripes called tiles which are then stitched together to form complete histology section image known as Whole Slide Image (WSI). The tiles are scanned and stored over multiple resolutions and hence each WSI can be visualized at different zoom level. Full resolution is known as level 0 and contain fine-grained details.

Table 4.2: Commonly used WSI file formats

Sr. No	File Formats	Extensions
1	TIFF, Nikon TIFF	.tif, .tiff
2	JPEG	.jpeg, .jpg
3	JPEG 2000	.jp2
4	Aperio / Leica SVS	.svs
5	Olympus VSI	.vsi
6	Hamamatsu NDPI	.ndpi
7	Objective Imaging (Glissando)	.sws

4.2 WSIs Visualizing tools

WSI files sizes can range between 1 GB up to 15 GBs and they are stored in special formats. WSI formats are specific and designed to store slide related all details in it. Common file formats for storing WSIs are given in Table 4.2. Due to this reason, such files cannot be visualized by ordinary digital viewers used for opening PNG like images. Most commonly used software for WSI files are listed below.

4.2.1 HiView

HiView is a GUI based image viewing and data exploring application for JPEG2000 file formats only. It is a simple and fast tool to visualize any JP2 file. After installing the software on machine, one need to drag, drop or use import option to visualize whole WSI in one go. It provide zooming option to increase the resolution. This tool is best to use by any layman because unlike other library based tools, it does not need exact code framework and adjustment of regions or levels. Hiview can also plot various histograms presenting various facts and can convert the selected regions into other formats.

4.2.2 OpenSlide

OpenSlide is a C library [29] that can be installed in Matlab or Pycharm environment. It provides interface to visualize whole slide images at multiple levels. The supported file formats are Leica(.scn), Aperio (.svs, .tif), Sakure (.svslide), Hamamatsu (.vmu, .ndpi, .vms), Philips (.tiff) and Generic tiled TIFF. Large scale of information can be extracted from WSIs using openslide functions.

- “Openslide_open” is a function to load whole WSI in the environment.
- “Openslide_get_level_count” can tell about the number of resolution levels the loaded image have.
- “Openslide_get_level_dimensions” is used to extract exact dimensions (X,Y) at particular level.
- “Openslide_read_region” can copy data from specific location (X, Y) and of given size from original WSI.
- “Openslide_close” is the function to close file from environment.

Other common visualizing tools that use OpenSlide library on backend are

- QueryPathology (QuPath)
- HistomicsTK

4.2.3 Automated Slide Analysis Platform (ASAP)

ASAP is a stand-alone multi-resolution histology image viewer as well as ground truth labeler. It provides simple interface where whole slide can be visualized and annotated using polygons, rectangles and dots.

4.2.3.1 “multiresolutionimageinterface” Library

Multiresolutionimageinterface is the python library binding comes with ASAP, providing multiple python functions for WSI visualization and regions extraction at certain level.

4.3 Ground truth generation tools

Most commonly used software for generating segmentation masks are manual and drawing based where regions of interest are segregated from rest image by drawing contours and filling the shapes with specific class intensity. These generated files are grey scale one dimensional PNG files. The other way to store different classes is by dedicating specific channels to certain class of objects. For example, Red channel storing pixels of micro-vessel class only, pixels indicating nerves are non-zero in Green channel only and the background class can occupy Blue channel, resulting in the formation of RGB masks.

4.3.1 GNU Image Manipulation Program (GIMP)

GIMP is an open source raster graphics editor launched in 1996 available for macOS, Windows and Linux operating system. It is a GUI based tool for image editing, retouching, free-form drawing, feature filtering, converting images into different formats and for other graphics related specialized tasks. GIMP native file format XCF can store all image related information produced in the tool. GIMP comes with a toolbox for the accessibility of image editing related tasks.

For annotating an image in GIMP, it is imported into the environment and a transparent layer is masked onto it. The regions for which we need to draw label are selected by drawing on object boundaries by free select tool. The selections are then locked and filled by Bucket fill tool using desired color or intensity. All these selections are stored in the additional layer which is then saved as independent PNG file or together with original image as XCF file.

4.3.2 Image Labeler

Image labeler is an application of MatLab designed specifically for labeling region of interest (ROI) on pixel, polyline or rectangle levels. It is a professional tool for creating ground truth data manually and by using automation algorithms. Intended Labels and sub-labels are firstly defined and then the objects are picked using rectangles or free selection. For semantic segmentation, these labels are assigned to every pixel of selected area.

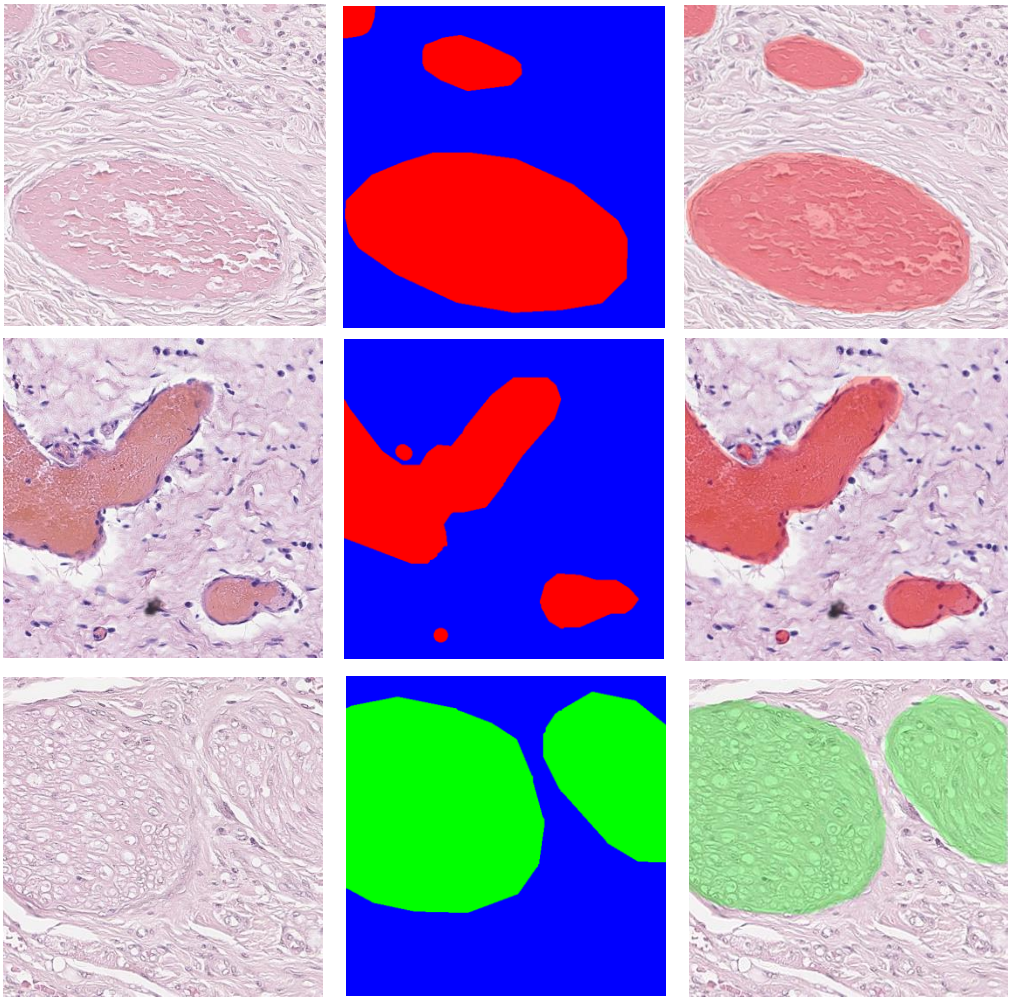
Table 4.3: Summary of NVSSD dataset

File Type	**No. of Images**	**Resolution**	**Level**
WSI	10	>100,000	0 (40 X)
Super-Patches	~450	2048 x 2048	1 (20 X)
Patches with 50% overlap	~ 18,000	256 x 256	2 (10 X)
Dataset after cleaning	~ 10,000	256 x 256	2 (10 X)
NVSSD-v2	**~ 10,000**	**256 x 256**	**2 (10 X)**
WSIs in NVSSD-v1	8	>100,000	0 (40 X)
NVSSD-v1	**~7,000**	**256 x 256**	**2 (10 X)**
NVSSD (v1 + v2)	**~ 17,000**	**256 x 256**	**2 (10 X)**

4.4 Summary of NVSSD

10 WSIs of oral cell carcinoma tissues are used for the generation of NVSSD-v2 dataset. Super-patches of 2048 x 2048 x 3 are extracted by cropping JP2 files at level 1 (20 x) using openSlide library. The extracted data is cleaned by removing unnecessary images i.e. those without Regions of interest. Grey scale annotations are drawn using GIMP tool from the cropped super-patches. Micro-vessels being labeled as class 1, nerves as class 2 and the rest of the area as class 0. The dataset ground truths have been verified by the pathologist. The super-patches along with their ground truth files are transformed to level 2 (10 x) and then again patches of 256 x 256 x 3 are taken with 50 % overlap. These cropped patches are saved as PNG files. NVSSD-v2 generated almost 10,000 number of samples.

NVSSD-v1 is the dataset of around 7000 images used originally to evaluate FABnet [24]. The dataset is extracted from 10 WSIs of H&E stained oral cell carcinoma tissues at 10x resolution and designed specifically for the semantic segmentation task of nerves and micro-vessels. Our generated dataset is merged into NVSSD-v1 to generate final number of samples. The summary of complete dataset used is given in Table 4.3. The complete NVSSD is used during training and testing of proposed architecture. Some examples of dataset are shown in Figure 4.1.



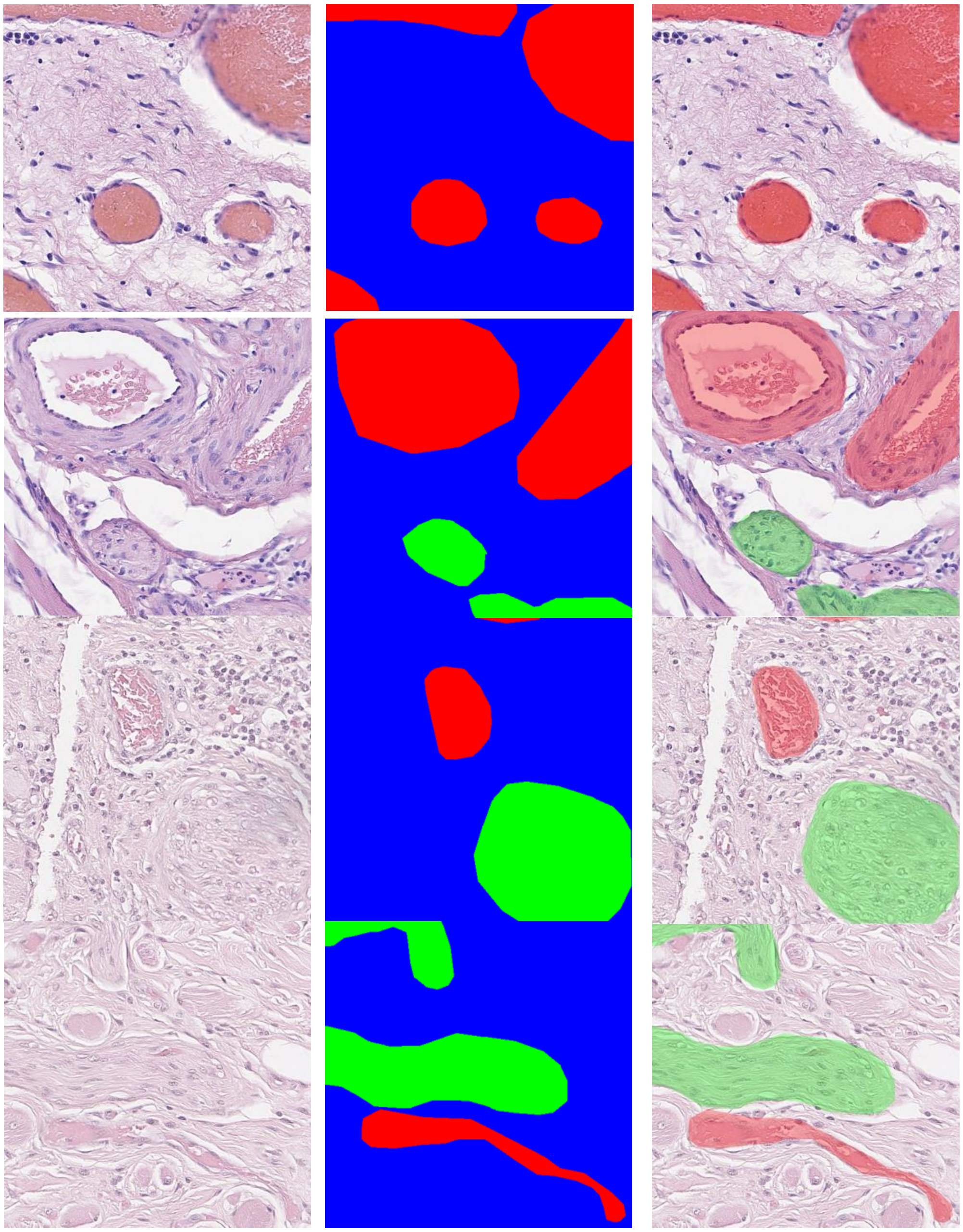


Figure 4.1: Examples of NVSSD images (left), annotation (middle) green color indicates nerve class, red indicates vessel class and blue showing background class and overlays (right)

CHAPTER 5

Experiment Results and Ablation Study

In this chapter, our proposed methodology is evaluated on the prepared Nerve and micro-Vessel Semantic Segmentation Dataset (NVSSD) and the results are compared with existing state-of-the art segmentation architectures. Our one-shot modular method is evaluated on standard semantic segmentation evaluation matrices used now-a-days. Table 5.1 shows the distribution of dataset used during training, validation and testing of architecture.

Table 5.1: Dataset division for machine training

Model	**No. of images**	**WSIs**
Training	12664 (70%)	12
Validation	2296 (15%)	3
Testing	2368 (15%)	3

5.1 Evaluation Matrices

We have used Jaccard Index, F1 score, Pixel accuracy, Precision, Recall and specificity for the evaluation purpose. All of these milieus are based on pixel wise and class wise correct and incorrect predictions. The exact terms TP, TN, FP and FN are discussed below followed by the details of evaluation matrices being used.

- True Positive (TP): The pixel belonging to c and predicted as c.
- True Negative (TN): The pixel does not belong to c and did not predicted as c.
- False Positive (FP): The pixel does not belong to c but predicted as c.
- False Negative (FN): The pixel belongs to c but didn’t predicted as c.

5.1.1 Jaccard Index

Jaccard Index, also known as Intersection of union is the most straightforward and frequently used matrix for segmentation task. For each class in an image, jaccard Index is calculated by dividing the overlapping area (Intersection) between the predicted map and the ground truth, with the area of union between predicted map and the ground truth. In other words, The TP pixel count for each class is divided by the sum of TP, FP and FN of respective class. The JI ranges from 0-1, it is 0 when segmentation is false, 1 when the prediction is exactly correct, and in between 0 and 1 otherwise, indicating the percentage of certainty when multiplied by 100.

The jaccard index will be calculated as

$$J(c) = \frac{Overlapping\ Area}{Area\ of\ Union} = \frac{|S_p \cap S_g|}{|S_p \cup S_g|} = \frac{TP}{TP+FP+FN} \qquad (5.1)$$

Where S_p is the predicted map and S_g is the ground truth and c is the class for which JI is calculated.

5.1.2 F1 (Dice Coefficient)

To calculate Dice coefficient for a particular class, Union between predicted mask and ground truth, is multiplied by two and divided by total number of pixels belonging to that class in predicted mask as well as in ground truth. Dice coefficient for every object class is calculated independently and there average is used to get mean F1 score of model. The formula of Dice coefficient is given below.

$$F1(c) = \frac{2*Overlapping\ Area}{Total\ no.of\ pixels} = \frac{2|S_p \cap S_g|}{|S_p|+|S_g|} = \frac{2TP}{2TP+FP+FN} \qquad (5.2)$$

Where S_p is the predicted map and S_g is the ground truth and c is the class of interest.

5.1.3 Pixel Accuracy

Another matrix for evaluating segmentation is pixel accuracy. Pixel accuracy is the percent of correctly classified pixels within each image. It can be calculated class wise separately and all classes globally can be used for reporting pixel accuracy. This matrix is not considered well for judging segmentation quality especially when the dataset is facing class imbalance issues. Accuracy will be calculated as:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \qquad (5.3)$$

5.1.4 Precision

Precision describes the percent of correctness in our predictions. The number of pixels belonging to each class that are correctly predicted relative to ground truth are divided by total number of pixels predicted for that class. It will be calculated as

$$\text{Precision}= \frac{TP}{TP+FP} \qquad (5.4)$$

5.1.5 Recall (sensitivity)

Recall addresses the question that out of all pixels belonging to each class, how many of them have we predicted correctly? Recall is gained by dividing True Positives with the sum of True positives (correctly predicted) and False Negative (incorrectly ignored). Recall is referred to the completeness and sensitivity of predictions. Sensitivity or recall can be calculated by given formula.

$$\text{Recall}= \frac{TP}{TP+FN} \qquad (5.5)$$

5.1.6 Specificity

Specificity is the matrix to quantitatively measure the accuracy of negative values predicted in a class. It illustrates that out of total pixels predicted as negative for each class, how many are actually negative relative to ground truth. It can be obtained by dividing TN with the sum of TN and FP as given in formula

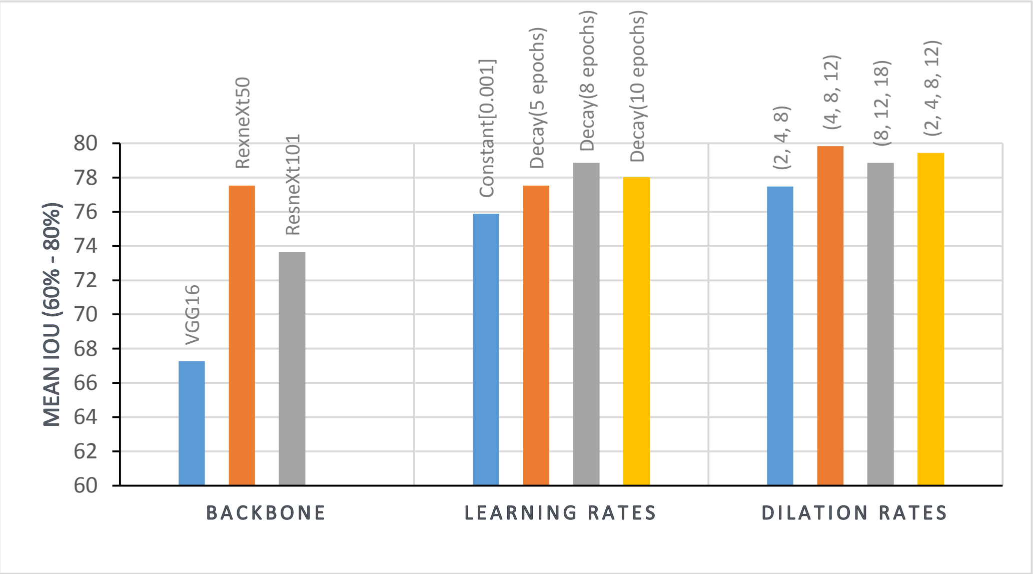
$$\text{Specificity}= \frac{TN}{TN+FP} \qquad (5.6)$$

Table 5.2: Ablation study

Hyper-parameters	**Instances Tuned**			
Backbone	VggNet	ResneXt50 ✓		ResneXt101
Learning Rate	Constant[0.001]	Decay (5 epochs)	Decay (8 epochs) ✓	Decay (10 epochs)
Epochs	15	25 ✓		50
Dilation Rates	(2, 4, 8)	(4, 8, 12) ✓	(8, 12, 18)	(2, 4, 8, 12)
Dropout	10 ✓	20		30
Image Normalization	Yes		No ✓	

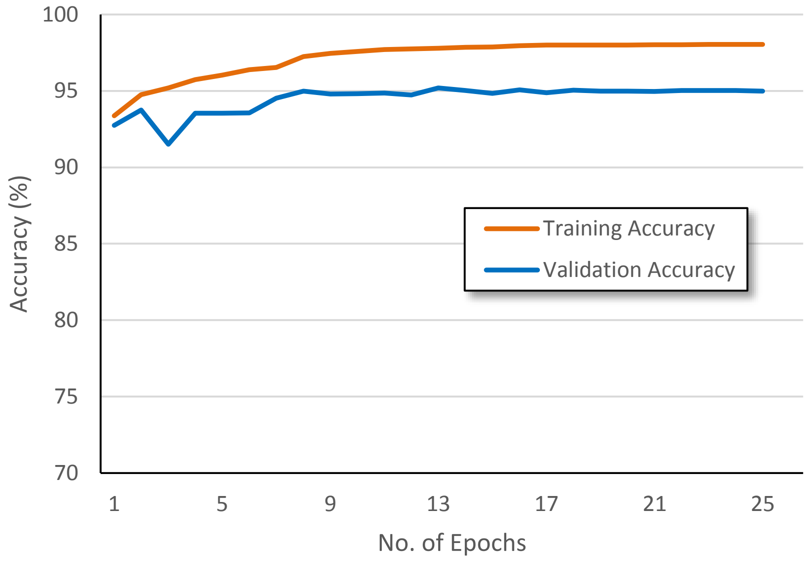
5.2 Ablation Study

Extensive experimentation is done on a number of hyper-parameters to achieve optimum results using proposed network. These parameters are number of training epochs, Backbone, Optimization Learning rate, Dilation rates in ASPP, Dropout and Normalization of training images. Each instance of a hyper-parameter is fluctuated keeping all other configurations constants and the behavior is analyzed. The dataset is trained and evaluated almost 20 times with different configurations. Multiple parameters tested along with best performing are indicated in Table 5.2.



Graph 5.1: Mean IOU (%) achieved against various instances of Backbone, Learning rates and Dilation Rates

Under-fitting and over-fitting are two practical phenomenon of convolutional neural networks, which are observed during the training phase of machine. They can directly affect the performance. Tuning some vital hyper-parameters by avoiding over or under-fitting can result in optimum convergence of the weights of CNNs. This selection can be network specific and dataset oriented. Table 5.2 depicts the best selection of instances for our network, and Graph 5.1 demonstrates three vital parameters, by fluctuating there values resulted in substantial shift in network outcomes.



Graph 5.2: Training and Validation accuracy VS no. of Epochs

5.3 Quantitative Results

The dataset is evaluated on proposed framework and the above evaluation matrices show a significant improvement in accuracy as compared to other architectures (Unet, Segnet, FCN, Deeplabv3+). Mean metrics comparison table (Table 5.3) and class-wise comparison table (Table 5.4) plotted depicts the capability of our network to perform segmentation task more precisely. Our architecture has achieved training, validation and testing accuracy of 98 %, 95% and 94 % respectively shown in accuracy graph 5.1.

Table 5.3: Comparison of mean quantitative with state-of-the-art techniques.

Algorithm	**JI**	**F1score**	**Accuracy**	**Precision**	**Recall**	**Specificity**
SEGNET	52.14	62.56	89.69	60.88	64.98	85.21
UNET	60.74	70.75	89.19	87.28	67.68	63.21
FCN8	72.30	82.83	95.01	88.20	78.59	90.35
Deeplabv3+	75.87	85.25	96.61	90.57	81.29	93.01
Proposed	**79.83**	**88.28**	**97.15**	**92.24**	**85.00**	**93.49**

Table 5.4: Comparison of class-wise quantitative results with state-of-the-art.

Class	Algorithm	JI	F1score	Accuracy	Precision	Recall	Specificity
Micro-vessel	SEGNET	51.19	67.26	92.20	67.70	66.74	95.77
	UNET	57.53	72.71	94.23	84.24	64.31	98.40
	FCN8	70.99	83.03	96.26	91.70	75.82	98.06
	Deeplabv3+	74.06	85.03	96.41	89.02	81.43	98.62
	Proposed	**78.82**	**88.15**	**97.13**	**91.75**	**84.83**	**98.90**
Nerve	SEGNET	23.22	30.72	92.48	19.03	41.50	95.71
	UNET	46.20	59.69	96.13	68.55	51.20	99.40
	FCN8	52.62	68.95	97.60	78.15	61.69	99.22
	Deeplabv3+	60.50	74.84	98.02	83.16	68.01	99.51
	Proposed	**65.58**	**79.21**	**98.57**	**88.44**	**71.73**	**99.63**
Background	SEGNET	83.20	90.79	84.82	92.34	89.37	62.40
	UNET	89.57	94.49	90.36	91.08	98.36	50.95
	FCN8	93.29	96.52	94.09	94.85	98.26	72.77
	Deeplabv3+	94.13	96.98	94.88	95.96	98.15	78.87
	Proposed	**95.09**	**97.49**	**95.75**	**96.53**	**98.45**	**81.95**

5.4 Qualitative Results

Our network has generated consistent predictions of the nerves and micro-vessels with clear and fine boundaries and has successfully avoided similar looking and unnecessary objects in images to much extent. By visualizing the sample-wise predictions generated by our network and that of existing segmentation architectures, one can easily figure out puzzling instances where the network was successful while other networks might got confused. Original images along with their ground truths, and their predictions generated by SegNet, Unet, FCN, Deeplabv3+ and proposed framework have shown in Figure 5.1.

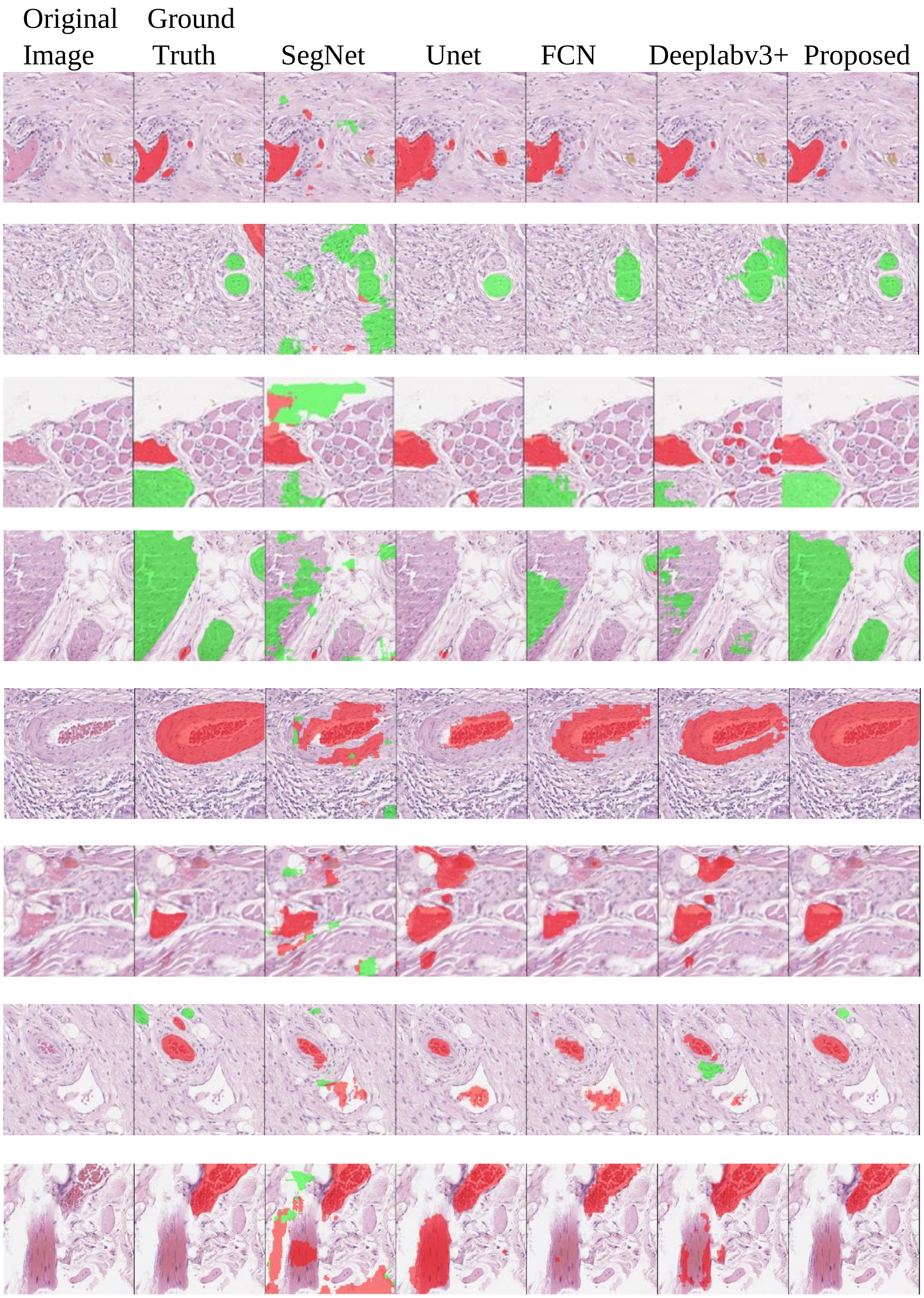


Figure 5.1: Comparison of qualitative results with State-of-the-art architectures

5.5 Discussion of Results

Above quantitative and qualitative results have shown an increase in the overall segmentation accuracy achieved by our network. It can generate consistent and more refined shapes of irregular dimensional objects as shown in the 1st ,3rd and 4th rows of Figure 5.1. Feature pyramids and Feature Fusion Block in our technique proved helpful for retaining exact location of multi-scaled objects avoiding holes or irregular prediction clusters. The convolution and normalization layers between decoder layers extends the feature learning process during spatial up-sampling. Hence, this configuration allowed the maturation and suppression of imperative features along with eradication of false accommodations. Our framework can also segment boundaries of objects with greater precision whereas other techniques specifically UNet and FCN has generated very poor boundaries of objects. The implementation of ASPP in our network addressed the problem of multi scaled Nerves and micro-Vessels as well.

Neither the empty vessels which are clinically unimportant nor the drained out red blood cells are classified as the regions of interest. These were the challenges we need to overcome during nerve and micro-vessel segmentation as discussed earlier in chapter 1. Taking in consideration the 6th and 7th example in Figure 5.1, it can be seen that our architecture has ignored empty micro-vessels and the red blood cells that are not enclosed in any vessel boundary and are residing in the stroma of tissue. But almost all other frameworks are somehow intermingling these areas with vessel class. Some of the relatively blurred images were also the part of testing dataset and they are predicted comparatively with greater details. As shown in 3rd and 8th sample of Figure 5.1, similar looking structures are often misclassified by other architectures but our network is able to accurately classify those structures as part of background class.

CHAPTER 6

Conclusion and Future Work

Computational pathology is an emerging field striving to push patient care to the next level. It's success is highly dependent on the accuracy of procedures being followed. Among the scope of digital pathology, cancer diagnosis and prognosis is an active research area now-a-days. Nerve and micro-vessel segmentation is a significant task for predictive cancer analytics by measuring MVD, LVI and PNI. A novel deep neural network based semantic segmentation architecture has been discussed for the histopathology structure segmentation task. FPN based multi-scale feature extraction block is used to encode discriminative features. The decoder comprises a feature fusion block to up-sample and concatenate the feature maps obtained from every pyramid of encoder. The feature selection process is further enhances by applying Atrous Spatial Pyramid Pooling Block. The whole framework is designed to specifically focus multi-scale objects of varying appearance. Model uncertainty is also studied by generating uncertainty maps and uncertainty percentage plot. Nerve and micro-vessel semantic segmentation dataset (NVSSD-v2) is generated from 10 WSIs of oral carcinoma tissue. The generated dataset is merged with already present NVSSD-v1 and is used to train and evaluate our proposed methodology. Evaluation matrices have shown a significant increase when compared to existing semantic segmentation algorithms like FCN, Unet, SegNET and Deeplab3+.

6.1 Future Work

Although the proposed framework has shown increased accuracy in segmentation task, some modifications can further enhance the performance.

- **Encoder-LSTM-Decoder Network:** Long Short Term Memory blocks can be added in the skip connections between general encoder and decoder of segmentation Network. It can prove helpful to retain only necessary information from encoded features and passing onto up-sampling layers for better predictions [30].
- **Boundaries and segmentation:** As the uncertainty maps have shown us, the predictions are not certain at boundaries of objects. A block to solely predict the boundaries of objects along with traditional segmentation framework can be added. It would let the machine to first predict the accurate boundaries and then retaining the segmented area between boundaries.
- **Uncertainty map during learning:** At training time, uncertainty maps can be generated and analyzed. It can let the model to learn where it’s not learning right. The samples with uncertainty greater than certain threshold can be focused to learn.

Bibliography

[1] T. J. Fuchs, J. M. Buhmann, “Computational pathology: Challenges and promises for tissue analysis,” in *Computerized Medical Imaging and Graphics,* 2011, vol. 35, no. 7–8, pp. 515–530

[2] S.P. Leon, R.D. Folkerth, P.M. Black, “Microvessel density is a prognostic indicator for patients with astroglial brain tumors” in *Cancer Interdisciplinary, International Journal of the American Cancer Society*, 1996, 77(2), 362

[3] N. Nishida, H. Yano, T. Nishida, T. Kamura, M. Kojiro, “Angiogenesis in cancer” in *Vascular health and risk management*, 2(3), 213, 2006.

[4] C. Liebig, G. Ayala, J.A. Wilks, D.H. Berger, D. Albo, “Perineural invasion is an independent predictor of outcome” in *colorectal cancer in Cancer: Interdisciplinary International Journal of the American Cancer Society*, 2009, 115(15), 3379.

[5] S.A. Hendry, R.H. Farnsworth, B. Solomon, et al, “The role of the tumor vasculature in the host immune response: implications for therapeutic strategies targeting the tumor microenvironment.” In *Frontiers in Immunology*, 2016;7:621.

[6] X. Yang, H. Li, X. Zhou, “Nuclei segmentation using marker-controlled watershed, tracking using mean-shift, and Kalman filter in time-lapse microscopy” in *IEEE Transactions on Circuits and Systems I: Regular Papers*, 2006, Pap. 53, pp: 2405–2414.

[7] A. Nedzved, S. Ablameyko, I. Pitas, “Morphological segmentation of histology cell images”, in *Proceedings 15th International Conference on Pattern Recognition*, 2000, Vol. 1,pp: 500–503.

[8] A. S. Chowdhury, R. Chatterjee, M. Ghosh and N. Ray, "Cell Tracking in Video Microscopy Using Bipartite Graph Matching," *2010 20th International Conference on Pattern Recognition (CVPR)*, Istanbul, 2010, pp. 2456-2459.

[9] M. Seyedhosseini, T. Tasdizen, “Multi-class multi-scale series contextual model for image segmentation,” in *IEEE Transactions on Image Processing*, 2013, 22, pp:4486–4496.

[10] Y. Al-Kofahi, W. Lassoued, W. Lee and B. Roysam, "Improved Automatic Detection and Segmentation of Cell Nuclei in Histopathology Images," in *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 4, pp. 841-852, April 2010.

[11] D. P. Mukherjee, N. Ray and S. T. Acton, "Level set analysis for leukocyte detection and tracking," in *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 562-572, April 2004.

[12] S.K. Nath, K. Palaniappan, F. Bunyak, “Cell segmentation using coupled level sets and graph-vertex coloring” *2006 Medical Image Computing and Computer-Assisted Intervention – MICCAI*, 2006, pp 101-108.

[13] K. Lee, W. Street, K.M. Lee, “A fast and robust approach for automated segmentation of breast cancer nuclei” in *Proceedings of the IASTED International Conference on Computer Graphics and Imaging, 1999*.

[14] G. Jiménez, D. Racoceanu, “Deep Learning for Semantic Segmentation vs. Classification in Computational Pathology: Application to Mitosis Analysis in Breast Cancer Grading,” *Frontiers in Bioengineering and Biotechnology* 2019; 7:145.

[15] C.L. Srinidhi, O. Ciga,, A.L. Martel, “Deep neural network models for computational histopathology: A survey”, 2019 Proceedings in *Image and Video Processing,* 2019.

[16] O. Ronneberger, P. Fischer, T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” in Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, vol 9351. Springer, Cham

[17] F. Yi, L. Yang, S. Wang, “Microvessel prediction in h&e stained pathology images using fully convolutional neural networks” in 2018 *BMC bioinformatics* 19(1), 64(2018).

[18] V. Badrinarayanan, A. Kendall and R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation,"

in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, vol. 39, no. 12, pp. 2481-2495.

[19] L.C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A.L. Yuille, “Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs” in *Transactions on Computer Vision and Pattern Recognition (CVPR)*, 2014.

[20] L. Chen, G. Papandreou, I. Kokkinos, K. Murphy and A. L. Yuille, "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, vol. 40, no. 4, pp. 834-848.

[21] L.C. Chen, G. Papandreou, F. Schroff, H. Adam, “Atrous Convolution for Semantic Image Segmentation” in *IEEE conference on computer vision and pattern recognition (CVPR)*, 2017, arXiv:1706.05587.

[22] L.C. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, “Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation” in *IEEE conference on computer vision and pattern recognition (CVPR)*, 2018.

[23] G. Lin, A. Milan, C. Shen, I. Reid, "RefineNet: Multi-path Refinement Networks for High-Resolution Semantic Segmentation," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, 2017, pp. 5168-5177.

[24] M.M. Fraz, S.A. Khurram, S. Graham, *et al.* “FABnet: feature attention-based network for simultaneous segmentation of microvessels and nerves in routine histology images of oral cancer” in *Neural Computing & Applications*, 2020, (32), pp: 9915–9928.

[25] M.M. Fraz, M. Shaban, S. Graham, S.A. Khurram, N.M. Rajpoot, “Uncertainty Driven Pooling Network for Micro-Vessel Segmentation in Routine Histology Images” in *Computational Pathology and Ophthalmic Medical Image Analysis, Springer International Publishing*, 2018, pp. 156-164.

[26] T. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan and S. Belongie, "Feature Pyramid Networks for Object Detection," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, 2017, pp. 936-944.

[27] S. Xie, R. Girshick, P. Dollár, Z. Tu and K. He, "Aggregated Residual Transformations for Deep Neural Networks," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, 2017, pp. 5987-5995.

[28] Y. Gal and Z. Ghahramani, “Dropout as a bayesian approximation: Representing model uncertainty in deep learning” in *International Conference on Machine Learning*, 2016, pages 1050–1059.

[29] Goode, Adam & Gilbert, Benjamin & Harkes, Jan & Jukic, Drazen & Satyanarayanan, Mahadev, “OpenSlide: A vendor-neutral software foundation for digital pathology,” in *Journal of pathology informatics*. 2013, 4. 27. 10.4103/2153-3539.119005

[30] F. Milletari, N. Rieke, M. Baust, M. Esposito, N. Navab, “CFCM: Segmentation via Coarse to Fine Context Memory” in *Medical Image Computing and Computer Assisted Intervention – MICCAI*, 2018, vol 11073. Springer, Cham.