Islamic University of Technology (IUT)

Nuclei Instance Segmentation of Cryosectioned H&E Stained Histological Images using Deep Learning

Authors

Zarif Ahmed, 170041056

Chowdhury Nur e Alam Siddiqi, 170041057

Fardifa Fathmiul Alam, 170041060

Supervisor

Tareque Mohmud Chowdhury

Assistant Professor

Dept. of CSE, IUT

Co-Supervisor

Tasnim Ahmed

Lecturer

Dept. of CSE, IUT

A thesis submitted in partial fulfilment of the requirements

for the degree of B. Sc. Engineering in Computer Science and Engineering

Academic Year: 2020-2021

Department of Computer Science and Engineering (CSE)

Islamic University of Technology (IUT)

A Subsidiary Organ of the Organization of Islamic Cooperation (OIC)

Dhaka, Bangladesh

May 12, 2022

Declaration of Authorship

This is to certify that the work presented in this thesis is the outcome of the analysis and experiments carried out by under the supervision of Tareque Mohmud Chowdhury, Assistant Professor of the Department of Computer Science and Engineering (CSE), Islamic University of Technology (IUT), Dhaka, Bangladesh. It is also declared that neither of this thesis nor any part of this thesis has been submitted anywhere else for any degree or diploma. Information derived from the published and unpublished work of others has been acknowledged in the text and a list of references is given.

Authors:

Zarif Ahmed

Student ID - 170041056

Chowdhury Nur e Alam Siddiqi

Student ID - 170041057

Fardifa Fathmiul Alam

Student ID - 170041060

Approved By:

Supervisor:

Tareque Mohmud Chowdhury

Assistant Professor

Department of Computer Science and Engineering (CSE)

Islamic University of Technology (IUT), OIC

Co-Supervisor:

Tasnim Ahmed

Lecturer

Department of Computer Science and Engineering (CSE)

Islamic University of Technology (IUT), OIC

Acknowledgement

We would like to express our grateful appreciation for **Assistant Professor Tareque Mohmud Chowdhury**, Department of Computer Science & Engineering, IUT for being our adviser and mentor. His motivation, suggestions and insights for this research have been invaluable. Without his support and proper guidance this research would never have been possible. His valuable opinion, time and input provided throughout the thesis work, from first phase of thesis topics introduction, subject selection, proposing algorithm, modification till the project implementation and finalization which helped us to do our thesis work in proper way. We are really grateful to him.

We are also grateful to **Tasnim Ahmed**, Lecturer, Department of Computer Science & Engineering, IUT for his valuable inspection and suggestions on our proposal of Biomedical Image Segmentation.

Abstract

Nuclei instance segmentation is an important step for oncological diagnosis and pathology research of cancer. HE stained images are considered the gold standard for medical diagnosis. But before being used for segmentation, it is required to preprocess them. There are two principle methods to preprocess them,formalin-fixed paraffin-embedded samples (FFPE) and frozen tissue samples (FS). Even though FFPE is widely used , it is a time consuming process whereas FS samples can be processed very quickly. But analysis of FS-derived HE stained images can be more difficult as rapid preparation, staining, and scanning of FS sections results in degradation of image quality. Therefore, in this thesis, we explored various state of the art segmentation architectures to create a model that will segment nuclei of FS-derived HE stained images with a high quality feature extraction.

Here, we have been working on a novel dataset called CryoNuSeg that contains 30 FS-sectioned images of 10 human organs.It has a benchline score of DICE 80.3 $\pm$ 4.3, AJI 52.5 $\pm$ 5.0, PQ 47.7 $\pm$ 6.1. U-Net is the first and most prominent architecture for biomedical image segmentation. We are exploring various U-Net architectures. We have trained Triple U-NET on the dataset using binary masks in place of U-NET keeping all other parts of the instance segmentation algorithm same such as Gaussian Filtering and Watershed Post processing.The results using Triple U-NET crossed all the benchline scores. The triple U-Net architecture gives a score of DICE 80.33, AJI 67.41 and PQ 50.56.

We have developed a deep learning model that performs highly accurate nuclei segmentation of FS sections despite degraded image quality for fast oncological diagnosis.

Contents

1 Introduction

1.1 Overview

Cancer is one of the most common cause of death in the whole world, causing the death of nearly 10 million people in 2020 alone [1]. Between 30 and 50% of cancer cases can be prevented by quickly diagnosing and implementing existing evidence-based prevention strategies. The rate of cancer can also be reduced through early detection of cancer. Many cases of cancer patients have a high chance of recovery if diagnosed early and treated appropriately. Sometimes, a surgeon must know the abnormal cell condition of a patient in the middle of an operation. That's why it is vital to diagnose cancer very fast so that treatment can start as soon as possible.

The domain of Computer Science most relevant to our research is Bioinformatics. Bioinformatics is the study of the process of extraction of biological data, converting it into meaningful forms and finally the analysis of said data. A significant portion of Bioinformatics research is focused on cancer diagnosis. Nuclei segmentation of HE stained images is monumental to understand the morphology, shape and count of nuclei which gives us important insights into abnormal cell patterns.

1.2 What is Nuclei Segmentation

A cell consists of three parts: the cell membrane, the nucleus, and, between the two, the cytoplasm. The nucleus contains the majority of the genetic materials. In the studies of cancer, a significant portion of research has usually been focused on this "gene-centric view" of the nucleus. The whole process starting from the identification of oncogenes and tumor-suppressor genes to the establishment of the multiple cell mutations concept is now commonly accepted as a standard requirement for cancer initiation and progression [2]

Nuclei segmentation is the process by which nuclei regions are segmented and extracted from cell tissue images. It is performed on Hematoxylin and Eosin(HE)

stained images. HE is the most popularly used stain in pathological diagnosis [3]. In this method, the nuclei is stained in the color of purplish blue due to the effect of hematoxylin while the extracellular matrix and cytoplasm is stained in the color of pink by Eosin , with other parts of the tissue cell taking on different shades, hues, and combinations of blue and pink.

Examination and monitoring of Hematoxylin and Eosin (HE)-stained tissue sections reveals critical information about each cell and its functional situation [4]. That's why, analysis of these images remains the "gold standard" [3] in diagnosing almost all types of cancer. Nuclei morphology, shape, type, count, and density are the key indicators of evaluation of HE-stained tissue cell images. To extract these indicator features automatically with a digitized method, nuclei segmentation is required.

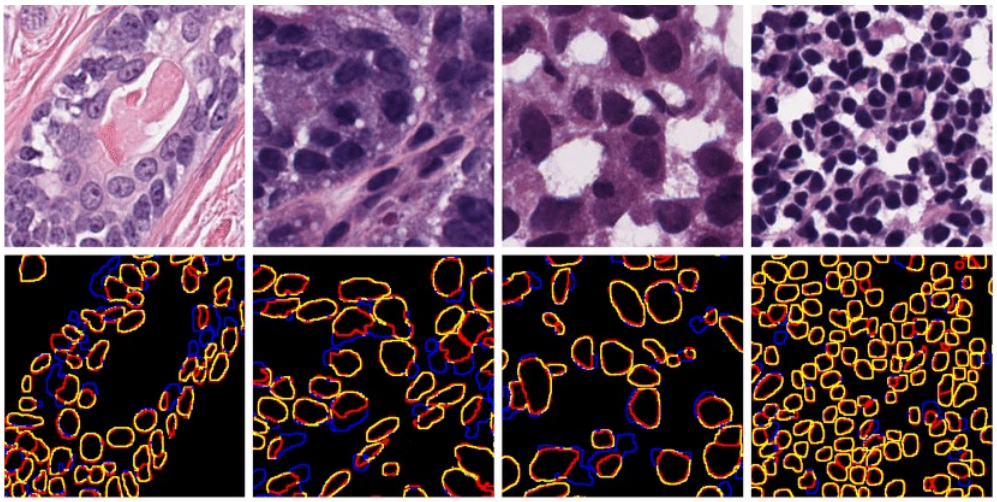

Figure 1: Nuclei Segmentation of HE Stained Images

1.3 Tissue Processing Protocols

There are two types of protocols to process tissues before they are being stained , namely formalin-fixed paraffin-embedded samples (FFPE) and frozen tissue samples (FS). HE staining of FFPE samples is more widely commonly used. Here, samples are processed with formalin and embedded in blocks of paraffin wax and

hence, the name. Thin slices of these blocks are then stained and used for diagnostic purposes. This procedure has better image quality so nuclei can be detected in an easier way but the entire procedure is very time-consuming, taking hours to days [**?**]. On the other hand, the preparation procedure of FS images is very smooth, easy and time efficient . The entire procedure can be completed in between 5 and 20 minutes [5]. The samples are then rapidly frozen for preservation of cell structure since paraffin was is not used here. Then a cryo-microtome is used to slice the frozen tissues.

Even though FFPE-derived HE stained tissue parts are traditionally the most commonly used samples, FS samples derived from cryosectioned HE stained tissue cells are more efficient for intra-operative surgical sessions as it can be performed fast. During the surgery itself, the oncosurgeon might need a rapid diagnosis which is not possible using FFPE as it takes days to be prepared. But a major problem is that,since these protocols are extremely different from each other in terms of processing, the resultant images will be widely different. Particularly, the nuclei quality might be drastically different in the Whole Slide Images (WSIs) . Analysis of FS-derived HE stained images can be more difficult as rapid preparation, staining, and scanning of FS sections may lead to degradation in image quality. The nuclei and the other cell parts might not be clearly visible.

1.4 Motivation and Problem Statement

For rapid diagnosis of cancer, usage of FS stained images is revolutionary as it decreases diagnosis time by multiple folds, bringing it down from days to a few minutes. Therefore, its nuclei segmentation can be performed in the middle of an operation too, where the surgeon might need crucial information about the nuclei morphology of the abnormal cells.

The main challenge of using FS stained images is its deteriorated image quality. That's why, we propose to develop a deep learning model that performs highly accurate nuclei segmentation of FS sections despite deteriorated image quality for fast oncological diagnosis.

2 Literature Review

In computer vision, image segmentation is the process of dividing the pixels of images into various areas and labeling each area into a certain part. Each area is called a segment.Each area has certain shared characteristics that differentiates them from other areas.This is a significant procedure because it extracts regions of interests and helps to analyze those regions. There are mainly two kinds of approaches to perform segmentation:

1. Classical Non AI Based Approaches
2. AI Based Approaches

2.1 Classical Approaches to Image Segmentation

Variousimage segmentation algorithms have been developed in scientific research from a very long time ago. Most renowned methods include thresholding [6], histogram-based bundling, regiongrowing [7], k-means clustering [8], watersheds [9], to more advanced algorithms such as active contours [10], graph cuts [11], conditional and Markov random fields [12], and sparsity based [13]- [14]methods. But they had very low accuracy and couldn't perform multi object detection.

2.2 Deep Learning Approaches to Image Segmentation

CNN is of the most prominent architectures in the research and applications of Computer Vision. The first CNN architecture was proposed by Fukushima in his paper on the "Neocognitron" [16].

Long et al. [17] projected one amongst the primary deep learning works for semantic image segmentation, employing a fully convolutional network (FCN). An FCN includes solely cnn layers, in order that it will take a picture of random size and manufacture a segmentation map of identical size. The authors changed existing CNN architectures, such as VGG16 and GoogLeNet by by exchange all fully-

connected layers with the fully-convolutional layers.In this manner, they managed variable sized input and output. As a result, the design shows spatial segmentation map and not classification scores. This work is taken into account a milestone in image segmentation, demonstrating that deep networks are often trained for semantic segmentation variable sized pictures.But it had some limitations:

1. Time consuming
2. Can't understand semantic meaning and contextual relationship
3. Doesn't work on 3D Images

Also, traditional CNN architecture doesn't work on medical images due to discrepancy in domain. Medical images required specialized highly precise segmentations.

2.3 Medical Image Segmentation

U-Net [**?**]is the first ever architecture proposed for biomedical image segmentation.It was a strong variation of a typical convolutional network. It focuses on precise localization of objects e.g. pixels with small number of data instead of the normal classification task of CNN.

The main plan is to feature consecutive layers to a traditional contracting network, wherever pooling operations area unit replaced by upsampling operations. Hence, these changed increase the output resolution. This high resolution contains the output from the contracting path area unit side to the upsampled output. Bupported this info of options, a consecutive convolution layer will learn to supply a a lot of precise output at the end. One vital modification is that a large range of feature channels area unit contained by the upsampling layer, which permit the network to send contextual data to higher resolution layers. Consequently, the expanding path is bilaterally symmetric to the contracting path, and yields a u-shaped network form and therefore the name.

2.4 Dataset

This [18] paper introduces CryoNuSeg, the first fully annotated FS-derived cryosectioned and HE-stained nuclei instance segmentation dataset. Nuclei segmentation is the process by which nuclei regions are segmented and extracted from cell tissue images. It is performed on Hematoxylin and Eosin(HE) stained images. HE is the most popularly used stain in pathological diagnosis. In this method, the nuclei is stained in the color of purplish blue due to the effect of hematoxylin while the extracellular matrix and cytoplasm is stained in the color of pink by Eosin , with other parts of the tissue cell having colors of various shades, hues, and combinations of blue and pink. But before being employed for segmentation, it is needed to pre-process them. There are 2 principle ways to preprocess them,formalin-fixed paraffin-embedded samples (FFPE) and frozen tissue samples (FS). albeit FFPE is wide used , it's a time intense method whereas FS samples may be processed terribly quickly. However analysis of FS-derived HE stained pictures may be tougher as fast preparation, staining, and scanning of FS sections ends up in decrease of image quality.

The dataset was made by selecting 30 WSIs from 10 different human organs. There were 3 WSIs per organ namely the lymph node,pancreas, pleura,,adrenal gland, skin, testis,larynx, mediastinum, thymus, and thyroid gland. Samples were taken from sampling centers with patients of various sexes, ages and races. Manual nuclei segmentation was performed by two trained annotators, a biologist (Annotator 1) and a bioinformatician (Annotator 2). This helps U.S.A. to look at inter-observer variability. Moreover, one among the annotators re-labelled the complete dataset in order that intra-observer variability may be measured.

U-Net architecture is enforced to separate foreground and background, whereas distance U-Net is enforced to seek out out the distance maps for all nuclei. After that a Gaussian smoothing filter is applied on these maps so that wrong local maxima are not detected. The size of kernel of the filter is detected from the U-Net results supported the average expected nuclei size. Finally, a a watershed-

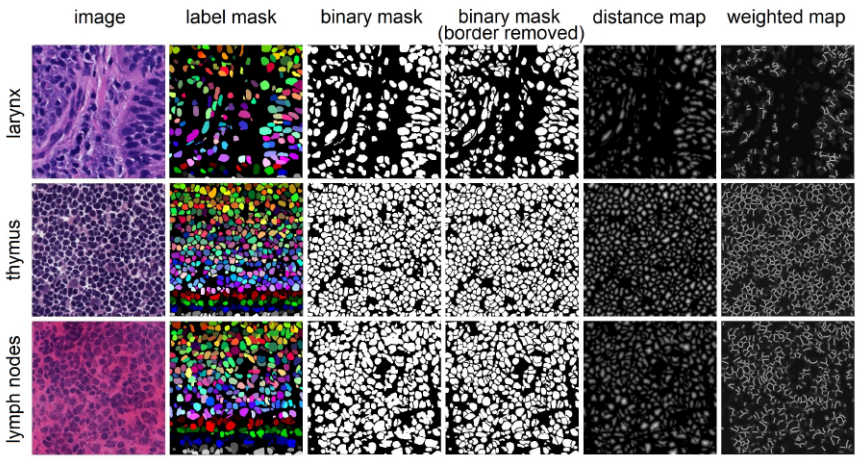

Figure 2: Sample FS-stained images from three human organs and their corresponding segmentation masks (from Annotator 1). For each sample, we show the raw image, the manually labelled nuclei, the binary segmentation mask, the binary mask with touching borders removed, the distance map, and the weighted map.

based rule is employed to merge the results from the 2 phases.

To assess the segmentation performance of the architecture on nuclei images, three evaluation indexes were used, the *Dice score, aggregate Jaccard index (AJI), and panc optic quality (PQ) score*. Because DICE score evaluates semantic results while AJI and PQ measures instance segmentation results.

2.4.1 Results

Several experiments were performed using MonuSeg-FS and MonuSeg-FFPE as training images respectively and CryoNuSEg as test. The results are shown in the table below.

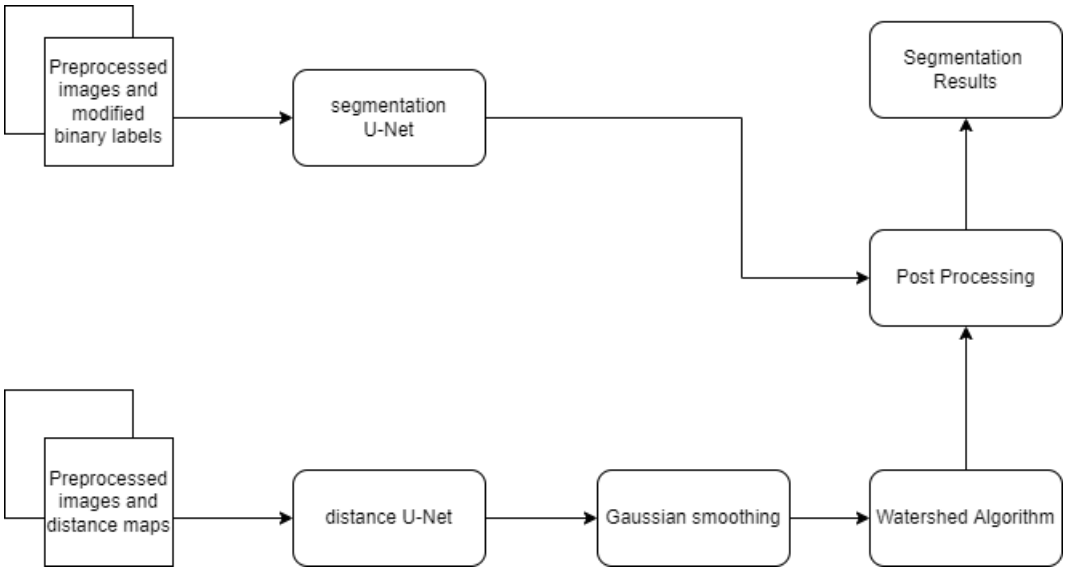

Figure 3: Flowchart of the employed instance segmentation algorithm

test organ	Dice score (%)		AJI (%)		PQ score (%)	
	MoNuSeg-FS	MoNuSeg-FFPE	MoNuSeg-FS	MoNuSeg-FFPE	MoNuSeg-FS	MoNuSeg-FFPE
adrenal gland	78.31	80.80	51.64	57.11	47.01	52.08
larynx	80.54	80.80	57.61	57.68	54.56	54.65
lymph node	79.71	80.00	49.62	49.90	49.07	49.18
mediastinum	83.24	83.34	50.56	50.84	47.19	47.85
pancreas	67.43	66.06	38.04	37.18	32.43	31.91
pleura	71.74	69.20	42.96	41.74	37.31	36.22
skin	72.96	72.33	45.82	45.73	40.29	40.40
testis	81.08	82.20	48.54	50.21	44.20	44.96
thymus	84.50	84.66	53.36	53.76	48.35	48.97
thyroid gland	80.11	80.46	55.80	56.83	48.19	51.62
average	78.0 ± 5.5	78.0 ± 6.4	49.4 ± 5.9	50.1 ± 6.8	44.9 ± 6.5	45.8 ± 7.4
p**-value**	0.6800		0.1529		0.0719	

In the next experiment, 10 cross-fold validation is implemented on the CryoNuSeg dataset to obtain results. The results are given below.

test organ	Dice (%)	AJI (%)	PQ (%)
adrenal gland	78.17	53.49	48.30
larynx	81.65	59.70	54.50
lymph node	81.56	53.54	50.79
mediastinum	84.87	54.10	50.73
pancreas	74.31	44.84	37.75
pleura	75.49	46.49	40.02
skin	74.75	47.84	40.78
testis	84.27	50.49	47.51
thymus	85.90	56.46	52.83
thyroid gland	81.81	58.20	53.48
average	80.3 ± 4.3	52.5 ± 5.0	47.7 ± 6.1

In the next experiment, the inter-observer variability was found out. For this,

manual segmentation masks of Annotator 2 were compared to those from Annotator 1 (first round). Annotator 1 is considered the ground truth.

3 Proposed Methodology

In the proposed paper of the dataset the author introduced a U-NET based algorithm They used a recently published SOTA Deep Learning instance segmentation algorithm as the baseline segmentation model. Below the diagram shows a visual walkthrough of the whole algorithm . The segmentation model consists of two sub-models, which are trained independently. A U-Net based model (segmentation U-Net) is used to perform basic semantic segmentation, while a Unet model focusing on the distance between centers of each nuclei is used to predict the distance maps for all nuclei instances. Both sub-models contain an upsampling part with five blocks of convolutional layers, dropout layers and max-pooling layers and a downsampling module with five blocks of convolutions layers, dropout layers and transposed convolutional layers. Similar to the original U-Net implementation, both sub-models use skip connections,which connect different blocks of the encoder path to the decoder path. The main differences between the two sub-models are the last activation layer and the utilised loss functions. For the segmentation U-Net, a sigmoid activation is used in the last layer while for the distance U-Net, a linear activation is applied. They use a combination of Dice loss and binary cross-entropy to train the segmentation U-Net, whereas mean squared error loss function is used to train the distance U-Net.After training the two sub-models, their results are merged to form the instance segmentation masks.To obtain the final instance segmentation masks, they first apply a Gaussian smoothing filter on the predicted distance maps to prevent false local maxima detection, with the kernel size of the filter determined from the segmentation U-Net results based on the average predicted nuclei size. Then, the local maxima are derived from the smoothed distance maps and are used as points for a watershed algorithm. They use the results from the segmentation U-Net as the labels for the watershed method to determine all background pixels. To segment the nuclei with

higher accuracy we propose to use better and latest versions of U-NET in place of traditional U-NET and would also experiment with other different instance segmentation methods if needed.We propose to use Triple U-NET in place of U-NET in the proposed algorithm for instance segmentation.

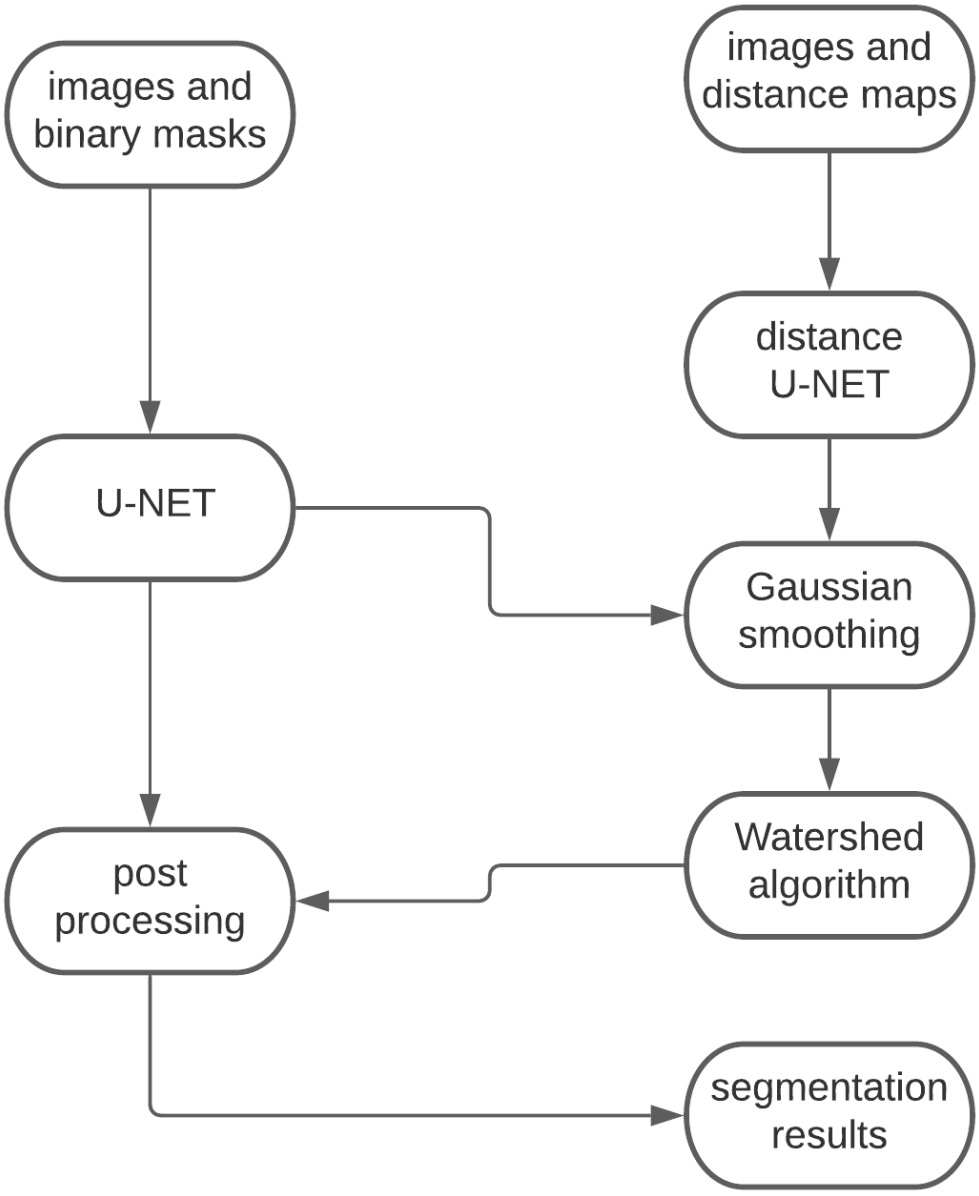

Figure 6- Author's proposed architecture

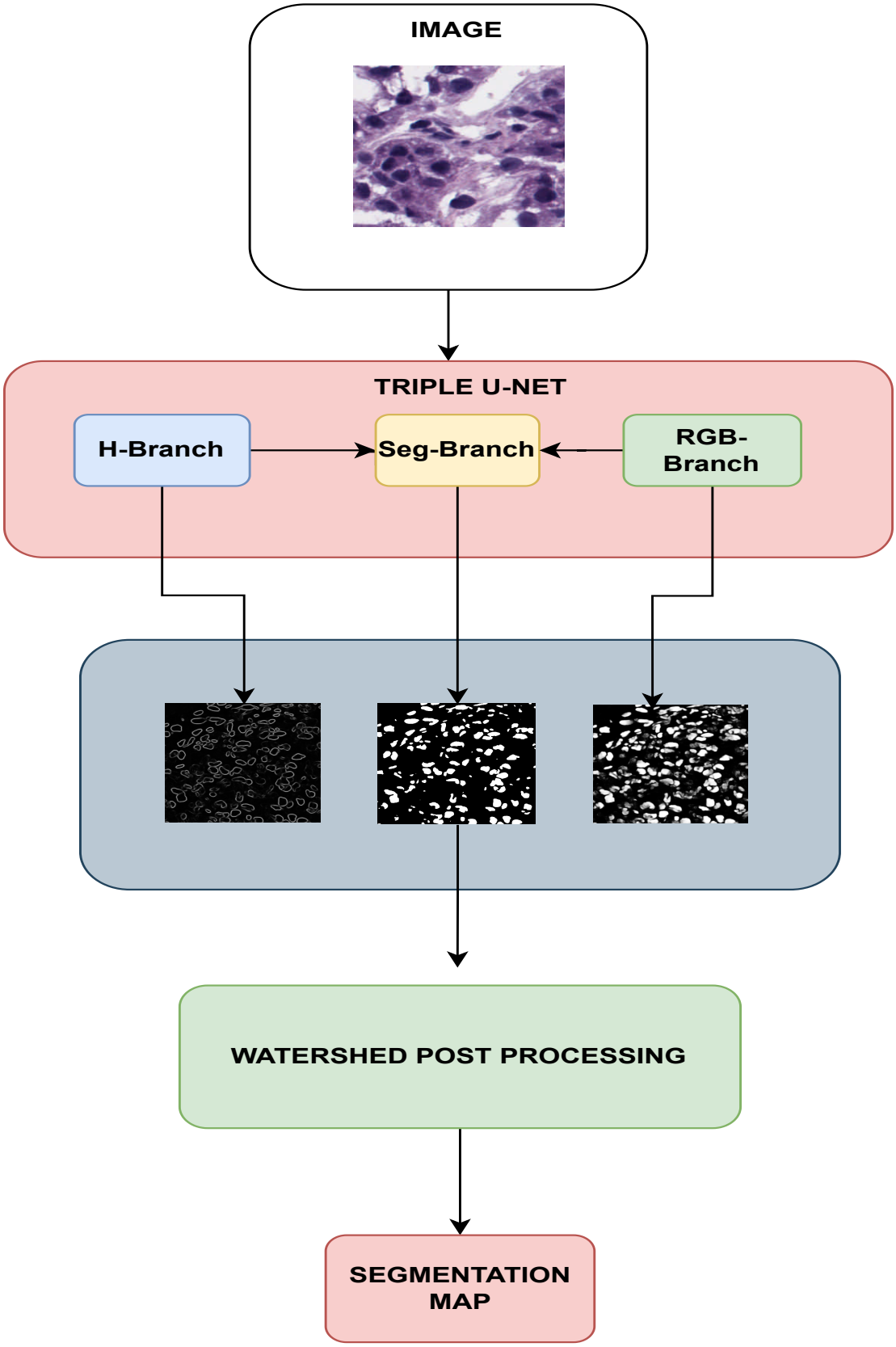

Figure 7- Our proposed architecture

3.1 Convolutional Neural Networks(CNN)

With the advent of image recognition tasks the need for better image suitable neural networks were required. Traditional Feed Forward neural Networks could not solve this problem with satisfatory accuracy. Thus the Convolutional Neural Network came into existense through thorough research.Convolutional Neural Networks as the name suggests uses Convolution Operation on input data.This helps in preserving the spatial information along with the contextual information and also the temporal information. Thus it can be used for higher dimensional data such as images which are generally 2-dimensional.The basic components of a CNN are as follows: "filter" which have the same dimension as the input, it performs the convolution operation on the input.Then there are "Pooling" Layers that works towards downsampling the image so that the computational burden decreases.Further there are regularization methods such as Dropout,Batch Normalization.Other important terminologies are "strides" which is the number of steps the filter will move across the input matrix,"padding" which is extra cells added to the border of the input matrix,"kernel" which is the name given for the filter.The filters have learnable parameters that are optimized during training through backpropagation,thus the feature extraction step of the model is also learned through training. At the last layer the final feature map is flattened to form a 1-dimensional vector and is passed through a Densely connected FeedForward Neural Network.This last layer gives the final classification output for the required task.

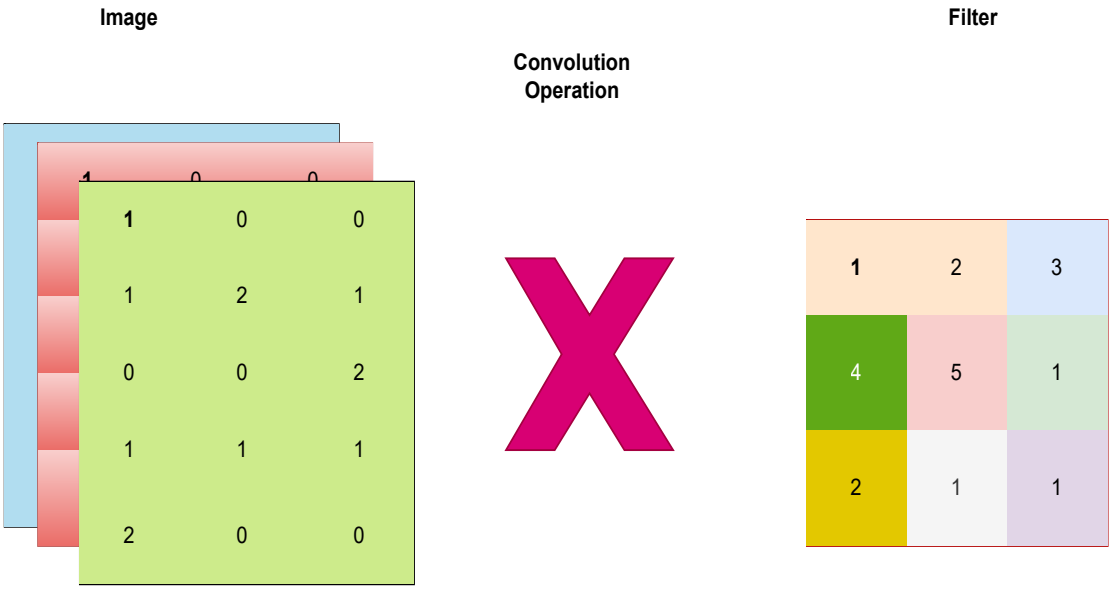

Figure-8 Image and Convolution Filter

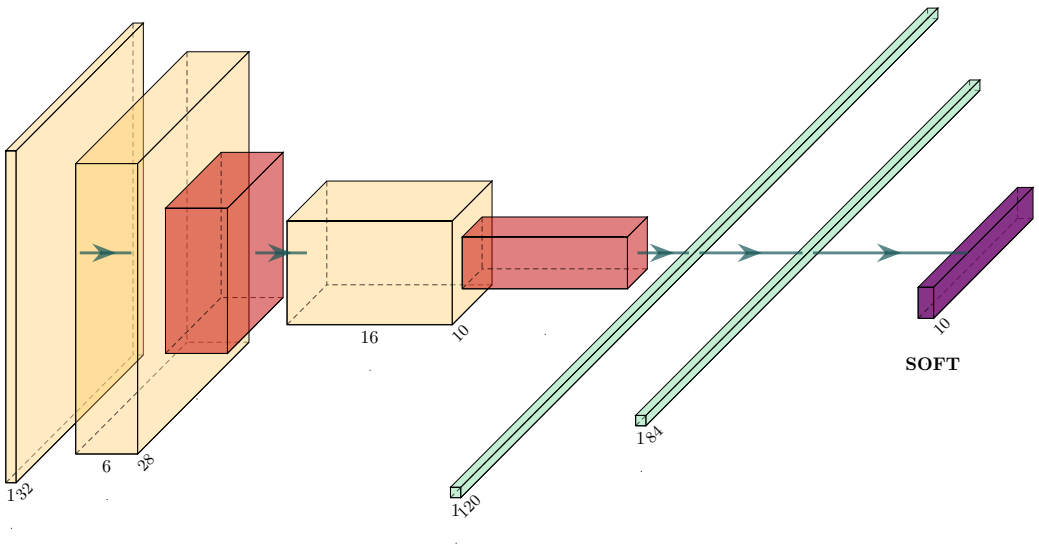

Figure-9 Convolutional Neural Network

3.2 Fully Convolution Neural network(FCN)

Traditional Convolutional Neural Network architecture uses convolution operations an also pooling operations to down sample the image and learn a condensed features map of the image given as input. This is done mainly to make the whole algorithm to be computationally feasible to be implemented. This works fine and is quite efficient for simple image classification tasks where we are not concerned with the location of the objects and information about its presence is enough to accomplish our task.However for segmentation task localization information is necessary to locate the object in the image but the down-sampling does not conserve the localization information.To mitigate this problem and make image segmentation feasible without making it computationally expensive a new design architecture was proposed called "Fully Convolutional Neural Network".FCN further adds an Up-sampling part after the down-sampling part of traditional CNN,this allows for using location information of objects in images along with the context.And allows us to perform semantic segmentation on images efficiently.

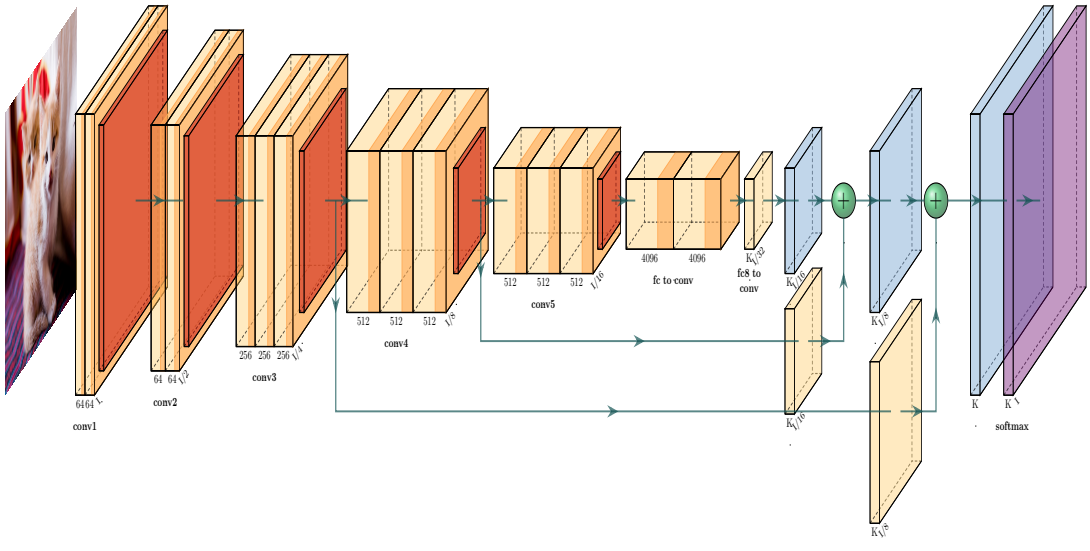

Figure-10 Fully Convolutional Neural Network

3.3 U-NET

U-net is one of the most popular models for biomedical image segmentation.U-net is built upon the architectural design of Fully Convolution Neural Networks(FCN).U-net consists of an encoder and also decoder.The encoder usually down-samples the image like any other CNN architecture and produces a condensed feature map of the image.The condensed feature map is then passed onto the decoder which then up-samples the image back again.The whole down-sampling is done using only 3x3 filters and 2x2 pooling layers.The feature maps are increased two folds at each layer.The up-sampling or decoder network works in the fundamental principle of transpose convolution which does the opposite task of convolution and increases the sample size. The unique technique used in the U-net that sets it apart from traditional FCN is the concatenation of feature map from encoder network to the decoder network.The concatenation process provides the network to better understand the representation during the training process. The activation function used for the network is Rectified Linear Unit which is quite famous across all neural network architecture and has proven its efficiency throughout multiple research papers.

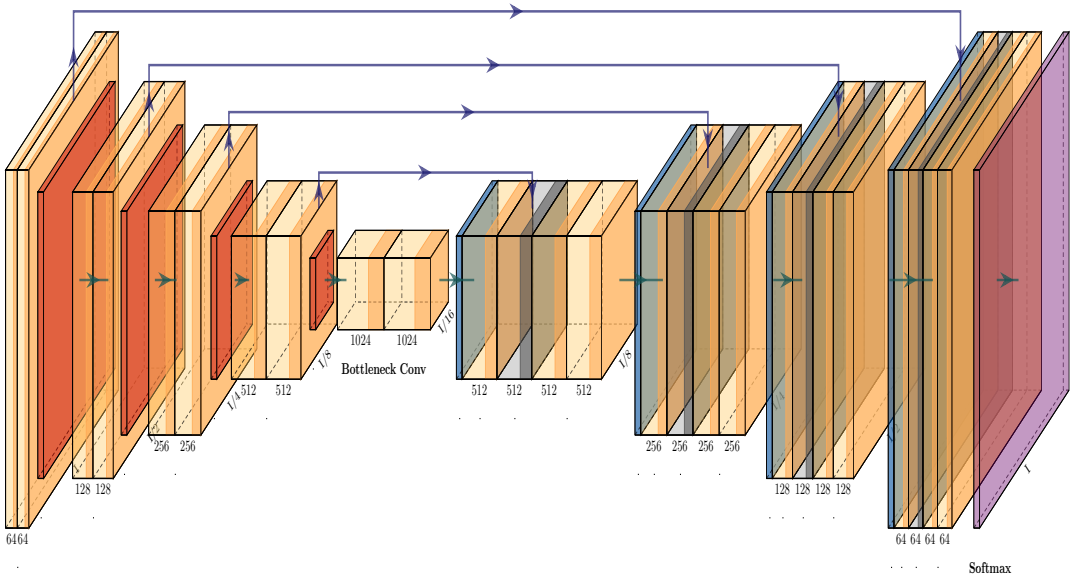

Figure-11 U-NET

3.4 Triple U-NET

This network contains three U-net branches that serve different purposes. The RGB branch focuses on the raw features of the segmentation task. The H branch focuses on the Hematoxylin-aware feature extraction of the nucleus which aids in detecting the edges of the nuclei in the images and enhances the diffrentiating ability of the network interms of overlapping regions in the images and helps in getting more accurate segmentation maskd of the cryosectioned smaples of whole slide images. The hematoxylin module is stable interms of inconsistent color of the H and E stains so applying normalization is not necessary and was not done as it may cause unnecessary information loss. The segmentation branch fuses the RGB raw features and the Hematoxylin-aware contour features and then predicts the final segmentation results.In this architecture, the authors present a Hematoxylin-aware Triple U-net for nuclei segmentation from WSI images, this is quite a new approach used by the authors to exploit the properties of HE stained images.The Concept of Beer Lamberts law has been used here, relating to light absorption property of objects to extract the Hematoxylin component, the proposed model is much more stable to color inconsistency, hence the color normalization is no longer necessary.The Progressive dense feature aggregation module also allows the model to learn features through merging them effectively. Through ablation studies and robust experiments across three data-sets this architecture has proved to be very effective in segmenting nuclei in whole slide images. Triple U-Net has shown promising results across a lot of standard data sets.

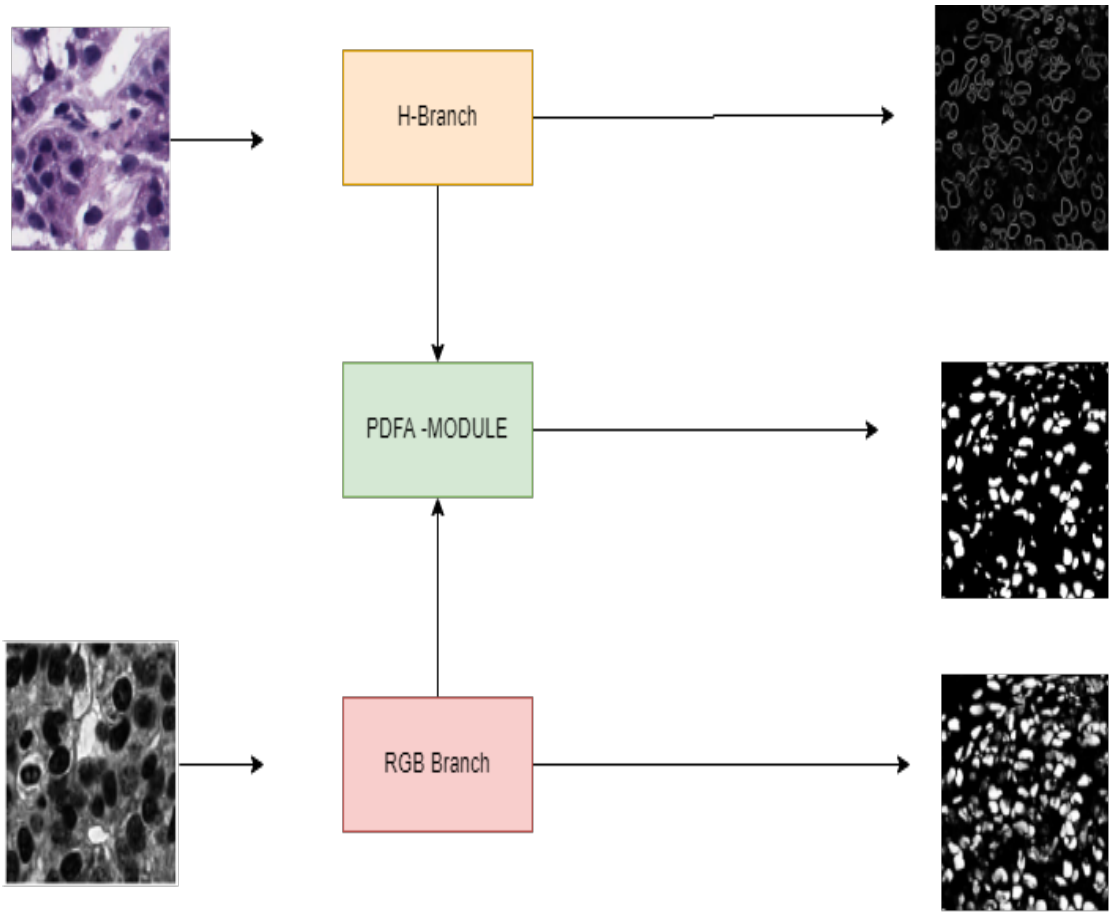

Figure-12 Triple U-Net

3.4.1 Progressive Dense Feature Aggregation

Progressive Dense Feature Aggregation(PDFA).Tradition Unet Architecture usually concatenate features directly,but features from previous layer are progressively combined with the later features This feature fusion technique is quite reasonable,as it improves feature propagation and reuses the features effectively.Here the encoding layer contains three layers in the PDFA module and the decoder layer has four layers.

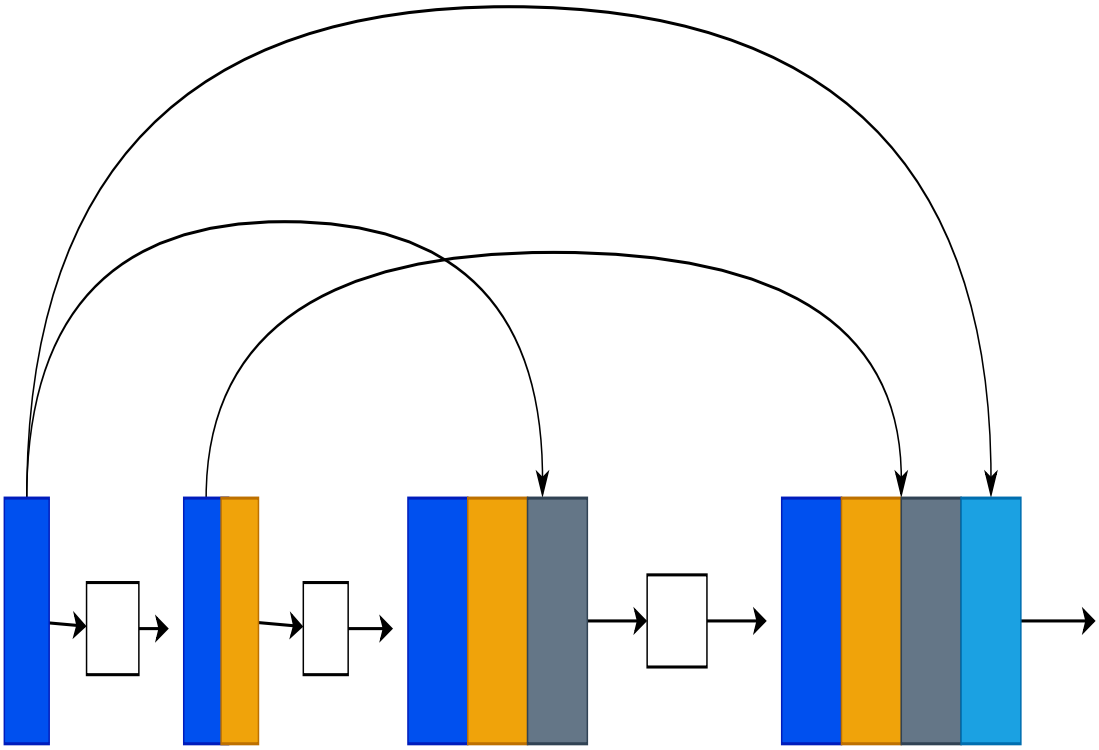

Figure-13 PDFA Module

3.4.2 Hematoxylin Component Extraction

Hematoxylin and eosin stains have been used for a long time for differentiating cellular contents in slides.Nucleus takes up blue colour whereas other cytoplasmic contents are coloured pink when these stains are applied.This property has been used to guide the network for the segmentation tasks.

3.5 Watershed Algorithm

Watershed is very popular image processing technique widely used for segmentation purposes.It is very useful for differentiating between overlapping region in segmentation tasks which is quite an important task for getting accurate segmentation maps of images.The algorithm requires user-defined markers for it to function. User defined markers can be defined using techniques such as thresholding or morphological operations. The idea is to identify valleys and peak region in an image.A peak is a region of high intensity and a valley is a region of low intensity.Considering each valley region we fill it with water of some colour and

put barriers at the point it may merge with other valleys.This barrier acts as the separation between objects and the algorithm succeeds in separating different objects.

3.6 Gaussian Filter

Gaussian Filtering is a technique used in image processing to remove noise from images.The Gaussian filter uses the Gaussian Weighted function to perform the smoothing operation.The technique requires convolving the image with a gaussian function.It is a type of low-pass filter that reduces the high frquency components in the images. Below is the formula for gaussian filters:

$$P(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-x^2}{2\sigma^2}}$$

$$P(x,y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-x^2+y^2}{2\sigma^2}}$$

When the formula is applied on an image it produces contours which forms concentric circles from the gaussian distribution.We obtain a convolution matrix using the values from the distribution that is used to convolve with the images. For each pixel a weighted average is calculated and then placed as the pixels values. The further the pixel the lower the weight it gets.Gaussian blur or filtering provides benefits of removing high frequency components and removes sharp edges

3.7 Implementation and training details

The whole experiment was performed using Google COLAB.The Deep Learning Framework used for the experiments was Pytorch.Gpu used was Nvidia Tesla K80.During training process the input images were augmented using various transformations techniques such as elastic transformation,random cropping,mirror rotation and flipping.The original images were resized to 256x256 pixels. During training learning rate was varied from 0.001-0.002 and learning rate schedulers were tried such as Exponential LR reduce on Plateau,Cosine annealing,Cosine annealing restarts.

3.8 Evaluation Metrics

We used three types of evaluation metrics- AJI(Average Jaccard Index), DICE, PQ.

AJI refers to Average Jaccard Index. The Jaccard index is also known as intersection over union. The Jaccard similarity coefficient is a statistic for measuring the similarity and differences among samples. As the equation shows, Jaccard Similarity depends on two sets. It basically calculates how much overlap there is between the two sets. Many sophisticated machine learning or deep learning tasks can use Jaccard Index- mainly for image or object classification or segmantation tasks.

The dice coefficient is pretty much similar to Average Jaccard Index. Similar to Average Jaccard index, it also considers area of overlap, but the difference is here, the overlap is multiplied by two.n Then this result of multiplying the overlap by two is further divided by the union of the two sets. The union of two sets can also be thought of as the total number of pixels in the two region, since we are dealing with images in this paper. Dice coefficient is very similar to Intersection over Union. Dice coefficient is positively correlated to Intersection over Union. This means if Intersection over Union decides that a certain model is better than the other then at segmenting an image then dice score will make the same decision. Similar to intersection over union, dice score also ranges from 0 to 1, where 1 denotes the greatest similarity between predicted region and the ground truth region.

The PQ (Panoptic Quality) is used to evaluate the performance of medels in Panoptic Segmentation tasks. In the equation of Panoptic Quality, the numerator sums up all the intersection over union ratios for all true positive values that have been detected. The denominator is calculated by summing true positives, false positives and false negatives. However, the false positive and false negative values are halved. The PQ score can be imagined as having two parts- a Segmentation Qality part, that evaluates how closely our segments matched with the ground

truths. When the SQ value comes closer to 1, it means that the true positive values are more closely matching with ground truths. But, no bad predictions are taken into account.

$$\mathrm{PQ} = \underbrace{\frac{\sum_{(p,g)\in TP} \mathrm{IoU}(p,g)}{|TP|}}_{\text{segmentation quality (SQ)}} \times \underbrace{\frac{|TP|}{|TP| + \frac{1}{2}|FP| + \frac{1}{2}|FN|}}_{\text{recognition quality (RQ)}}$$

Our proposed model works significantly better than the baseline model, in terms of all kinds of evaluation metrics. Among them, the AJI score has increased substantially. The results are shown in following tables.

AJI Score		
Fold	Baseline	Our Model
Thyroid	0.5349	0.6682
Thyroid	0.5349	0.668186
Thymus	0.5970	0.720366
Testes	0.5354	0.72661
Skin	0.5410	0.639007
Pleura	0.4484	0.637411
Pancreas	0.4649	0.62711
Mediastinum	0.4784	0.707169
Lymph node	0.5049	0.684007
Larynx	0.5646	0.657913
Adrenal Gland	0.582	0.673637
AVERAGE	**0.525**	**0.6741416**

Table-3 Comparison of AJI scores

From the table above it can be observed that the proposed model shows significant improvement in terms of AJI score. The proposed model shows improvement in results for the images of all the organs involved in the experiment.

PQ Score		
Fold	Baseline	Our Model
Thyroid	0.550355	0.4830
Thyroid	0.4830	0.550355
Thymus	0.5450	0.543708
Testes	0.5079	0.499412
Skin	0.5073	0.480857
Pleura	0.3775	0.480179
Pancreas	0.4002	0.457219
Mediastinum	0.4078	0.512735
Lymph node	0.4751	0.477618
Larynx	0.5283	0.4807
Adrenal gland	0.5348	0.572905
AVERAGE	**0.477**	**0.5055688**

Table-3 Comparison of PQ scores

The proposed model also shows promising results. There is an overall improvement of about 3 percent. AJI score sometimes over penalizes the overlapping region. To avoid this problem Panoptic Quality (PQ) is used for evaluating the segmentation performance.

Dice Score		
Fold	Baseline	Our Model
Thyroid	0.7817	0.796866
Thymus	0.8165	0.83687
Testes	0.8156	0.841115
Skin	0.8487	0.779416
Pleura	0.7431	0.778141
Pancreas	0.7549	0.770692
Mediastinum	0.7475	0.828457
lymph node	0.8427	0.807039
larynx	0.8590	0.792724
Adrenal gland	0.8181	0.801589
AVERAGE	**0.803**	**0.8033**

Table-3 Comparison of DICE scores

Our proposed model shows competitive performance compared to the original model in terms of DICE score. Even though the performance is competitive, our proposed model performed slightly better than the original model.

3.8.1 Comparing AJI and DICE

As stated in the previous section our proposed model performs significantly better than the original model in terms of AJI score as compared to the DICE score.

In this section we will provide some comprehensive analysis of AJI and DICE scores in terms of the ways in which these scores are calculated. And eventually reach a conclusion stating which method evaluates a model more strictly.

The equation of AJI is given below:

$$AJI = \frac{\sum_{i=1}^{n} G_i \cap P_j}{\sum_{i=1}^{n} G_i \cup P_j + \sum_k P_k}$$

The equation of PQ is given below:

$$PQ = \frac{2\times|G_i \cap P_j|}{|G|+|P|}$$

The equations of AJI and PQ scores have been diagrammatically represented for better understanding.

Below the visual representation of AJI score has been displayed.

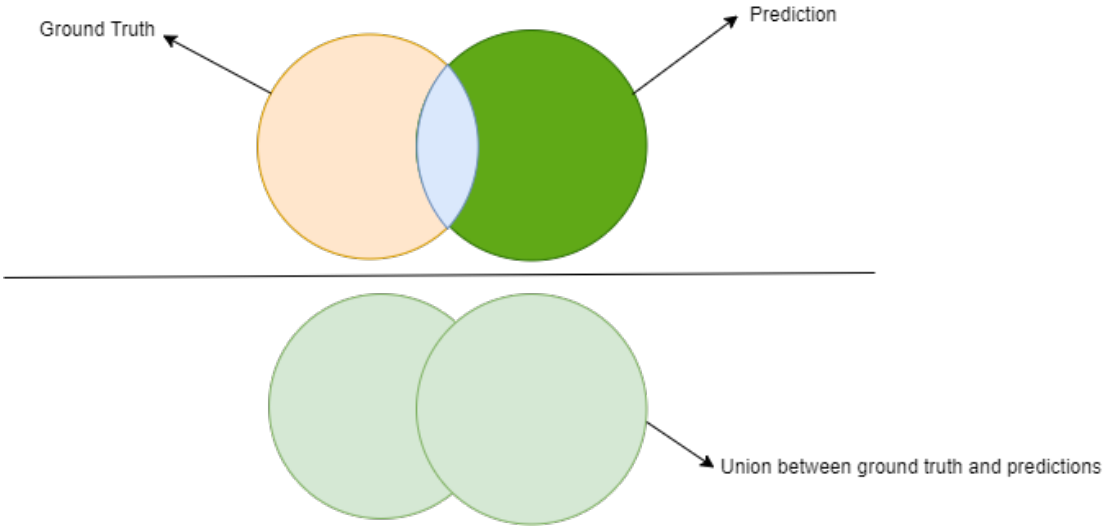

Figure -14 Visual representation of AJI score equation

The yellow region represents the ground truth and the dark green region represents the predicted pixels. Below same color represents union between the ground truth and the predicted region. The intersection is represented by blue color. AJI divides the intersection region with the union of the pixels.

Below the visual representation of DICE score has been displayed.

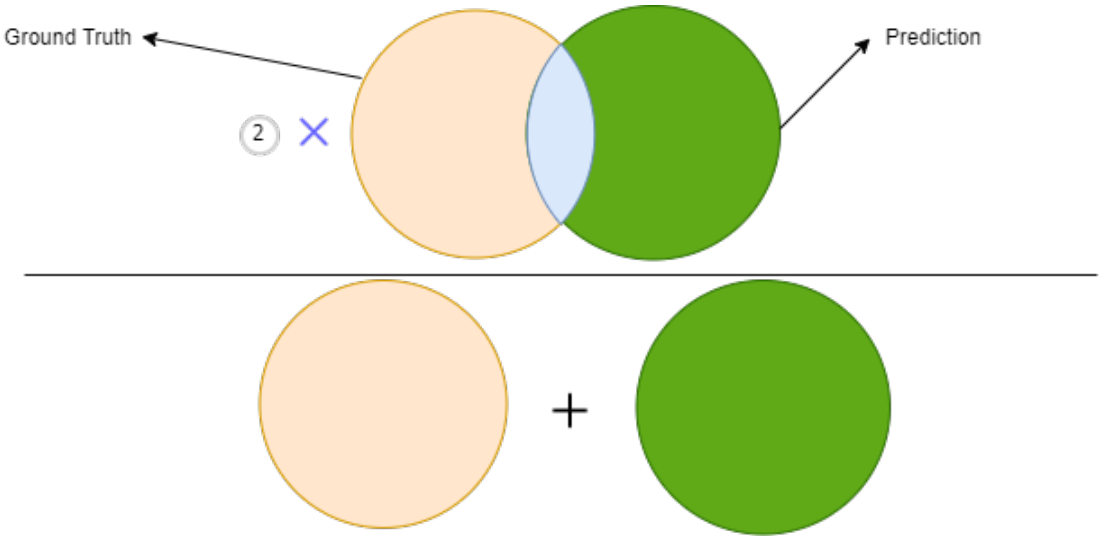

Figure -15 Visual representation of DICE score equation

In the figure above the yellow region represents the ground truth and the dark green region represents the predicted pixels.The light blue region represents the intersection between the predicted and the ground truth pixels, which is then multiplied by 2. Below, the total number of pixels are added, i.e the summation of ground truth pixels and the predicted pixels are taken. Finally the doubled intersection is divided by the summation of ground truth pixels and the predicted pixels.

3.8.2 Why is AJI more strict metric than DICE

In dice score we are multiplying the correct classification in numerator of the equation by 2 and then dividing it by the sum of prediction results and ground truths. AJI is similar to dice except for the fact that we are not multiplying the correct classification by 2. The effect of this is that the rate at which/the magnitude by which AJI penalizes wrong predictions is more drastic than dice as dice score awards correct prediction more than it penalizes incorrect predictions. Therefore, AJI has the tendency of over penalizing, and thus can be considered a more strict scoring method. The AJI score in our implementation out performs the baseline model by a large margin.

4 Results and Discussion

Authors from Cryonuseg dataset set the baseline segmentation benchmark using the dataset. For the first step we implemented the author's algorithm using the same dataset by coding it in python and obtaining the following results.

All the sub-models have been trained for 30 epochs. When trained, the starting learning rate was 0.01. It is to note that the training was done using a learning rate scheduler. As training progressed, the learning rate was halved after 8 epochs. Size of each batch was set to 8 and many different types of augmentation techniques were applied in the while training which includes flipping horizontally and vertically randomly and with a probability 0.5. There were rotations of 90 degrees, 180 degrees and 270 degrees in a random manner again having probability of 0.5. Shifting of brightness and contrast were done with contract shift and brightness limits of 0.15 having a probability 0.40. We also did histogram equalization, which was random and adaptive and images were also cropped randomly having a size of 512 $\times$ 512 pixels. Finally, during inference stage, we apply two post processing steps on the final segmentation results which removing instances from the segmentation masks, by small instances we mean objects with areas less than 20pixels. We then fill the holes within the objects detected.

Some of our experimentations using authors' weights are given below.

fold	PQ	DICE	AJI
Thyroid	0.515696	0.795297	0.666902
Thymus	0.51138	0.815816	0.691043
Testes	0.442313	0.777187	0.637739
skin	0.496141	0.801158	0.669015
pleura	0.493488	0.818558	0.692998
pancreas	0.459	0.767039	0.622113
mediastinum	0.4834	0.809779	0.680553
lymph node	0.487128	0.796483	0.66873
larynx	0.462777	0.756264	0.610017
adrenal gland	0.590263	0.822044	0.700428
AVERAGE	**0.4941586**	**0.7959625**	**0.6639538**

Table-4 Segmentation Results with Metrics: epochs=15,lr=0.0001 scheduler=StepLR where training is performed with triple unet authors weights)

fold	PQ	DICE	AJI
Thyroid			
Thymus	0.539165	0.856029	0.748777
Testes	0.453209	0.780814	0.640752
skin	0.498672	0.785637	0.647387
pleura	0.421398	0.733676	0.580131
pancreas	0.476078	0.768886	0.624723
mediastinum	0.476596	0.816135	0.689519
lymph node	0.461268	0.778346	0.643527
larynx	0.48074	0.801207	0.670525
adrenal gland	0.566073	0.790254	0.658103
AVERAGE	**0.485911**	**0.7901093333**	**0.6559382222**

Table-5 Segmentation Results with Metrics:
epochs=15,lr=2e-6,Scheduler=Cosine anealing Restarts where training is performed with triple unet authors weights)

fold	PQ	DICE	AJI
0	0.550355	0.796866	0.668186
1	0.543708	0.83687	0.720366
2	0.499412	0.841115	0.72661
3	0.480857	0.779416	0.639007
4	0.480179	0.778141	0.637411
5	0.457219	0.770692	0.62711
6	0.512735	0.828457	0.707169
7	0.477618	0.807039	0.684007
8	0.4807	0.792724	0.657913
9	0.572905	0.801589	0.673637
AVERAGE	**0.5055688**	**0.8032909**	**0.6741416**

Table-6 Segmentation Results with Metrics: epochs=20,lr=0.0001,Scheduler=StepLR where training is performed with triple unet authors weights)

By using the same dataset and running it on the Triple U- architecture we implemented , we managed to improve the baseline results.

Our accuracy scores on Triple U-net were *Dice 80.33%, AJI 67.41% and PQ of 50.56%.*

4.1 Performance Analysis

Paper	Model	AJI	DICE	PQ
Ronneberger et al.	U-Net	0.5250	0.4770	0.8030
Our Work	**Triple U-Net + watershed algorithm**	**0.6741**	**0.5056**	**0.8033**

Table-4 Performance Comparison of Proposed Architecture with the Baseline Model

Our proposed architecture outperforms the baseline model by all evaluation criterions. Amongst them , AJI has the highest improvement with over with almost 30 percent increase in score.

Triple U-net predicts precise nucleus boundary which a simple U-Net architecture is not good enough to do. It performs edge detection with high accuracy. It also predicts contour-aware boundaries more accurately. The PDFA allows the network to learn features and merge their information from various domains of the nuclei. Moreover, this model does not need any kind of color normalization. Thus, it saves computational power too.

Therefore, our proposed architecture provides a state-of-the-art solution to high precision nuclei instance segmentation of frozen sectioned tissues.

4.2 Error Analysis

Among all the organs,our model gave relatively lower dice scores for the Skin,Larynx,Lymph Node samples.This is partially due to the small sample size for each organs available for training. These organs have complex tissue cells, the nuclei are harder to detect because of their smaller size. Also, cancer in these organs are quite rare as well. Thus, after it is frozen, the image quality degrades even more. That's why, more samples are required for these particular organs to get better results.

5 Conclusion

We proposed a state of the art architecture for high precision nuclei instance segmentation of cryosectioned H E stained frozen sectioned tissues.This will help to analyze nuclei morphology, size, density and detect any kind of abnormality. Therefore, it will help to diagnose cancer in a patient in a very short time. Early diagnosis can help to save a patient's life. It will also help surgeons take intra-operative decisions with confidence.

We further conducted an in-depth error analysis on the performance of our proposed architecture,which can be pioneering for future research work in this domain. In future, we want to extend this work by incorporating our experimental setup with a larger dataset with the latest augmentation techniques and modification of the proposed Triple U-Net architecture.

References

[1] Ferlay J, Ervik M, Lam F, Colombet M, Mery L, Piñeros M, et al. Global Cancer Observatory: Cancer Today. Lyon: International Agency for Research on Cancer; 2020 (https://gco.iarc.fr/today, accessed February 2021).

[2] Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144(5):646–674. doi: 10.1016/j.cell.2011.02.013. [PubMed] [CrossRef] [Google Scholar]

[3] Titford, M. (2005). "The long history of hematoxylin". Biotechnic Histochemistry. 80 (2): 73–80. doi:10.1080/10520290500138372. PMID 16195172. S2CID 20338201.

[4] John K. C. Chan, "The wonderful colors of the hematoxylin-eosin stain in diagnostic surgical pathology," International Journal of Surgical Pathology, vol. 22, no. 1, pp. 12–32, 2014.

[5] Hariharan, B., Arbelez, P., Girshick, R., Malik, J.: Hypercolumns for object segmentation and fine-grained localization (2014), arXiv:1411.5752 [cs.CV] Prentice Hall Professional Technical Reference, 2002.

[6] N. Otsu, "A threshold selection method from gray-level histograms," IEEE transactions on systems, man, and cybernetics, vol. 9, no. 1, pp. 62–66, 1979.

[7] R. Nock and F. Nielsen, "Statistical region merging," IEEE Transactions on pattern analysis and machine intelligence, vol. 26, no. 11, pp. 1452–1458, 2004.

[8] N. Dhanachandra, K. Manglem, and Y. J. Chanu, "Image segmentation using k-means clustering algorithm and subtractive clustering algorithm," Procedia Computer Science, vol. 54, pp. 764–771, 2015.

[9] L. Najman and M. Schmitt, "Watershed of a continuous function," Signal Processing, vol. 38, no. 1, pp. 99–112, 1994.

[10] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," International journal of computer vision, vol. 1, no. 4, pp. 321–331, 1988.

[11] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," IEEE Transactions on pattern analysis and machine intelligence, vol. 23, no. 11, pp. 1222–1239, 2001.

[12] N. Plath, M. Toussaint, and S. Nakajima, "Multi-class image segmentation using conditional random fields and global classification," in Proceedings of the 26th Annual International Conference on Machine Learning. ACM, 2009, pp. 817–824.

[13] J.-L. Starck, M. Elad, and D. L. Donoho, "Image decomposition via the combination of sparse representations and a variational approach," IEEE transactions on image processing, vol. 14, no. 10, pp. 1570–1582, 2005.

[14] S. Minaee and Y. Wang, "An admm approach to masked signal decomposition using subspace representation," IEEE Transactions on Image Processing, vol. 28, no. 7, pp. 3192–3204, 2019.

[15] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," arXiv preprint arXiv:1706.05587, 2017.

[16] K. Fukushima, "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position," Biological cybernetics, vol. 36, no. 4, pp. 193–202, 1980

[17] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 3431– 3440.

[18] O Ronneberger, P Fischer, T Brox, "U-net: Convolutional networks for biomedical image segmentation", Medical Image Computing and Computer-Assisted Intervention (MICCAI), Springer, LNCS, Vol.9351: 234–241, 2015

[19] DA Novis and RJ Zarbo, "Interinstitutional comparison of frozen section turnaround time. a college of american pathologists q-probes study of 32868 frozen sections in 700 hospitals," Archives of Pathology Amp; Laboratory Medicine, vol. 121, no. 6, pp. 559–567, June 1997.

[20] Jevgenij Gamper, Navid Alemi Koohbanani, Simon Graham, Mostafa Jahanifar, Syed Ali Khurram, Ayesha Azam, Katherine Hewitt, and Nasir Rajpoot, "PanNuke dataset extension, insights and baselines," arXiv preprint arXiv:2003.10778, 2020.

[21] Ruchika Verma, Neeraj Kumar, Abhijeet Patil, Nikhil Cherian Kurian, Swapnil Rane, and Amit Sethi, "Multi-organ nuclei segmentation and classification challenge 2020," .

[22] Simon Graham, Quoc Dang Vu, Shan E Ahmed Raza, Ayesha Azam, Yee Wah Tsang, Jin Tae Kwak, and Nasir Rajpoot, "Hover-Net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images," Medical Image Analysis, vol. 58, pp. 101563, 2019.

[23] Ciresan, D.C., Gambardella, L.M., Giusti, A., Schmidhuber, J.: Deep neural networks segment neuronal membranes in electron microscopy images. In: NIPS. pp. 2852–2860 (2012)

[24] Dosovitskiy, A., Springenberg, J.T., Riedmiller, M., Brox, T.: Discriminative unsupervised feature learning with convolutional neural networks. In: NIPS (2014)

[25] Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2014)