Faster R-CNN-based Nuclei Segmentation of Cytological images using Localization

A thesis submitted in partial fulfilment of the requirement for the

Degree of Master of Computer Science and Engineering

of

Jadavpur University

By

Nilanjan Chatterjee

Registration Number: 154132 of 2020-21

Examination Roll Number: M4CSE22008

Under the Guidance of

Prof. Nibaran Das

Professor

Department of Computer Science and Engineering

Jadavpur University, Kolkata-700032

India

June, 2022

FACULTY OF ENGINEERING AND TECHNOLOGY

JADAVPUR UNIVERSITY

CERTIFICATE OF RECOMMENDATION

This is to certify that the thesis entitled "**Faster R-CNN based Nuclei Segmentation of Cytological images using Localization**" has been completed by Nilanjan Chatterjee (University Registration No.: 154132 of 2020-21, Examination Roll No.: M4CSE22008). It is a bonafide piece of work completed under my guidance and supervision and is accepted in partial fulfillment of the requirement for the award of the Degree of Master of Engineering, Department of Computer Science and Engineering, Faculty of Engineering and Technology, Jadavpur University, Kolkata

_____

Prof. Nibaran Das

(Thesis Supervisor)
Professor, Department of Computer Science and Engineering
Jadavpur University, Kolkata-700032

Countersigned

_____

Prof. Anupam Sinha
Head, Department of Computer Science and Engineering,
Jadavpur University, Kolkata-700032.

_____

Prof. Chandan Mazumdar
Dean, Faculty of Engineering and Technology,
Jadavpur University, Kolkata-700032.

FACULTY OF ENGINEERING AND TECHNOLOGY
JADAVPUR UNIVERSITY

CERTIFICATE OF APPROVAL

This is to certify that the thesis entitled "Faster R-CNN based Nuclei Segmentation of Cytological images using Localization" is a bonafide record of work carried out by Nilanjan Chatterjee in partial fulfillment of the requirements for the award of the degree of Master of Engineering, in the Department of Computer Science and Engineering, Jadavpur University during the period of January 2021 to July 2022. It is understood that by signing this approval, the signatories are just endorsing the thesis for the intended use and are not necessarily endorsing or applauding any statements made, opinions offered, or conclusions reached therein.

_____

Signature of Examiner

Date:

_____

Signature of Supervisor

Date:

FACULTY OF ENGINEERING AND TECHNOLOGY
JADAVPUR UNIVERSITY

DECLARATION OF ORIGINALITY AND COMPLIANCE OF ACADEMIC ETHICS

I hereby declare that this thesis entitled "Faster R-CNN based Nuclei Segmentation of Cytological images using Localization" contains literature survey and original research work by the undersigned candidate, as part of his Degree of Master of Engineering.

All information in this document has been obtained and presented in accordance with academic rules and ethical conduct.

I further affirm that, in accordance with these guidelines, I have properly attributed and referenced all information and findings that are not unique to this work.

Name: Nilanjan Chatterjee
University Registration No.:154132 of 2020-21

Examination RollNo.:M4CSE22008

Thesis Title: Faster R-CNN based Nuclei Segmentation of Cytological images using Localization.

Signature:

_____

Date:

_____

ACKNOWLEDGEMENT

I would like to express my extreme gratitude towards God for providing me the strength and wisdom for successfully completing the project.

I am extremely grateful to my thesis guide, **Prof. Nibaran Das**, Department of Computer Science and Engineering, Jadavpur University, Kolkata for providing me his invaluable guidance, constant encouragement and for inspiring me for the tenure of my dissertation.

I am highly indebted to **Jadavpur University** for providing me the opportunity and the required infrastructure to carry on my thesis.

I am also pleased to thank the **Center for Microprocessor Applications for Training Education and Research Lab, Computer Science &Engineering Department, Jadavpur University** for providing me with the proper laboratory facilities as and when required and also thanks **Theism Diagnostic Centre** for providing required dataset for accomplishing the task.

For all the teaching and non-teaching staff whose support has helped me to complete my project successfully

Nevertheless, I am glad to express my gratitude towards family members, classmates, seniors especially Mr. Soumyajyoti Dey and friends for giving me constant encouragement and mental support throughout my work.

_____

Nilanjan Chatterjee
University Registration No. : 154132 of 2020-21
Examination Roll No. :M4CSE22008
Master of Engineering
Department of Computer Science and Engineering
Jadavpur University

CONTENTS

FIGURES

TABLES

CHAPTER ONE

INTRODUCTION

Cancer is one of the deadliest diseases in recent times. According to the World Health Organization (WHO), 10 million people [1] around the world died from cancer in 2020. The WHO further predicts that there could be 24 million new cases and 14.4 million deaths due to cancer by 2035. [2]. As per the American Cancer Society, 70% of cancer cases occur in low-income countries. [3]. In India, cancer cases have risen at an average annual rate of 1%-1.2% in the last three decades [3]. The most common cancer types in men are lung, colorectal, liver, stomach, and prostate, while colorectal, lung, breast, thyroid, and cervical cancer are common in women[3].

Generally, the malignant cells divide themselves uncontrollably in a shortest time period and it can spread to other tissues. So, the chances of survival for a patient suffering from the last stage of cancer are extremely low. However, when a patient suffering from earlier stages of cancer is provided with proper medical treatment, there is a chance of improvement of the health of that patient. Therefore, cancer detection in the early stages is crucial to reducing the chance of mortality. When a patient sees lumps or tumors at any part of the body, he or she comes to the doctor, and the doctor suggests a biopsy test. Among all the different types of biopsy processes, Fine Needle Aspiration Cytology (FNAC) [4] is a less painful and easier diagnostic procedure. In the FNAC test, the fluid is collected from the tumor by a needle and after that, the slide is smeared by the staining process. Different types of stains like MGG, and PAP are used for this purpose. These slides are viewed under the microscope and the professional practitioners are decided it is cancer or not. However, manually annotating a large number of slides is time-consuming for the pathologist. So, researchers have been trying to develop many efficient methods for automated cancer detection from digital pathology images for several years. By analyzing digital cytological images,

CAD (Computer Aided Diagnosis) based systems can make the cancer diagnosis process more economical.

In recent years, deep learning has taken a dominant role in the CAD-based cancer diagnosis process. In the cytology image, the observation of two major objects: nuclei and cytoplasm, has taken an important role in diagnosing malignancy. So, automated nuclei detection and segmentation from cytology images is a challenging task for the computer scientist.

In computer vision problems, two major subdomains are localization [7] and segmentation [7]. Localization is used to detect all the objects present in a particular image. The position of the objects and their corresponding size and class can be retrieved using localization. On the other hand, semantic segmentation produces a pixel-wise mask image for the input image. The output mask consists of several regions where each region corresponds to a particular class. This provides the boundary information for each region. However, in semantic segmentation, two objects of the same class are identified as distinct instances. This problem is resolved in instance segmentation, where each object is uniquely identified. Several traditional and Convolutional Neural Networks (CNN) based algorithms exist that can easily segment the nuclei from the cytological images. CNN is a deep learning-based supervised algorithm that can extract features present in an image. It can be used to isolate a nucleus from its surrounding cytoplasm. We aim to do instance segmentation where each nucleus can be uniquely identified. A two-stage approach can be used, where, in the first stage; a localization algorithm can detect all the nuclei present in the cytological image. Then a semantic segmentation algorithm can be applied in the second stage, where each localized nuclei can be isolated from its surrounding cytoplasm and the adjacent cell.

1.1 MOTIVATION

In the present work, we have tried to develop a deep learning-based automated nuclei segmentation algorithm. The system should be able to extract all the structural features of all the nuclei present in the image. Though many traditional and deep learning-based segmentation methods exist, there are still lots of challenges in the proper segmentation of all nuclei from the cytology image.

a) The traditional deep learning algorithm uses handcrafted features based on contour, pixel gradient, and pixel intensity. They work well only for such datasets, where these features are much more prominent. While this algorithm can do semantic segmentation, it often fails to do instance segmentation, recognizing each individual successfully.
b) The more advanced supervised deep learning model uses labelled data on which the model is trained to obtain the state-of-the-art result. Since the annotation for the input image is created manually, there are several instances of incomplete annotations present in the object. The majority of deep learning models learn this in a negative way, resulting in a large number of detections. The cytological images consist of several overcrowded regions containing a dense cluster of cells in a small region. They are too closely spaced for a deep learning model to identify each cell uniquely.

Another problem with the cluster of nuclei is that the cells are highly overlapped, resulting in less boundary information between two nuclei. While this does not hinder semantic segmentation, instance segmentation often becomes quite challenging, especially preserving the boundary information. Extracting the correct boundary information for each of the nuclei is crucial. However, most semantic segmentation algorithm produces much coarser boundary failing to retrieve the exact smooth border. Since Improper results in the medical domain can be life-threatening. A system which could take care of the above problem along with the one taken care of by the previous

models is required to be developed. This motivates us to develop a segmentation technique with localization properties that will aid in cancer diagnosis. The two major contributions of the present thesis are given below

I. Deep Learning based automatic nuclei localization and segmentation from digital cytology images.
II. Segmentation of localized nuclei from Grabcut method.

CHAPTER TWO

LITERATURE SURVEY

For the past few decades, researchers have been trying to develop automated nuclei segmentation from digital pathology images.

Mouelhi et al. [9] used a modified geometric active contour on cytology image, followed by a method combining a concave vertex graph and a watershed algorithm for properly quantifying different stains. They further used some morphological operations. This algorithm's performance surpasses the previous algorithm's performance by 3-4%. Ronneberger *et al.* [10] *proposed* U-Net for biomedical image segmentation, which uses an encoder-decoder architecture with a contracting and expanding path with skip connections, connecting layers of the encoder to the decoder at the same level. The skip connections prevent the coarser information from getting lost in the downsampling step. This algorithm won the ISBI cell tracking challenge in 2015. However, this model suffers from its sensitivity to its predefined weights. Raza *et al.*[11] propose a unified model called Micro-Net, an advanced extension of U-Net. Its architecture allows the processing of the images at multiple resolutions, detecting nuclei of various sizes.

Oda *et al.* [12] proposed BES-Net, which has a similar structure to U-Net but uses two decoding paths to account for the problem of inaccuracies in the boundary of the small objects. While one detecting path enhances the edge of the image, another path focuses on the overall segmentation. It uses adaptive weighing of the loss function and skip connection to strengthen the boundary of the cells, similar to two U-Net. The detection rate of this method is 89.5% on HE-stained images approved by Nagoya University Hospital (Japan). Zhou et al. [13] suggested a new deep neural network called the Contour-aware Informative Aggregation Network (CIA-Net) to handle this. It made use of two task-specific decoders with a multilayer information aggregation module. Further abnormalities due to outliers are further reduced by truncation loss.

However, there are two disadvantages of the above approaches. The shared encoder network fails to properly differentiate between the semantic segmentation and the instance segmentation, creating problems in case of overlapping nuclei. It aggregated task-specific features in a bidirectional manner to account for the texture and spatial dependencies. This model outperforms other models in unseen images of the MoNuSeg dataset of the 2018 MICCAI challenge by an F1 score of 0.8458.Schmidt et al. [14] propose a method for adequately segmenting the borders of overcrowded nuclei regions using star–convex polygons to localize the cell nuclei. This method needs no shape refinement module, as the polygon better refines the shape than the bounding boxes. Finally, a convolutional neural network is used to predict a polygon for each pixel of the cell instance. However, in the case of a large-sized nucleus, contextual information is lost. The average precision of this model on the DSB2018 dataset is 68%-86%. For the prediction, only the features of the centre pixels are used.

Mask-RCNN [1] is a two-stage convolutional neural network model proposed by Gkioxari et al. [15] . It was built over the Faster-RCNN model. Faster-RCNN is a neural network where a feature region proposal network proposes several regions from the feature map obtained from the input image. These proposals are used to create bounding box offsets. Mask-RCNN produces instance segmentation output from these localized boxes.

SpaNet proposed by Koohbanani et al. [16], is a proposal free spatially aware network (SpaNet) that uses a multiscale approach to extract spatial information. The centroid detection maps and pixel-wise segmentation of nuclei are first predicted using a dual-head variant of the SpaNet. A single head SpaNet further processes their outputs to extract the positional information for each positional nuclei. This is followed by a spectral clustering method which determines the connected component using the nuclear mask and Gaussian-like detection mask. The F1-score of SpaNet is 0.8281 and 0.8451 on seen and unseen organs. Another variant of the two-stage detector is BRP-Net or Boundary-assisted Region Proposal Network proposed by Song *et al.*[17]. The author first proposed a Task-aware Feature Encoding (TAFE) network which aids in

extracting features for semantic segmentation and boundary information. This architecture had shown state of art results. The dice score of this dataset is 87.7 on the CPM17 dataset.

In opposition to this, Stacked U-Nets(SUNet) [18] is another two-stage architecture proposed by Kong et al. In the first stage, the model created a pixel-wise segmented image that was concatenated with the input image and fed to the second stage. The method used cross-entropy loss and focal loss in the first and the second stage respectively. However, they used a watershed algorithm for post-processing which is sensitive to noise.

Recently Feature Pyramid Network[19] has emerged as one of the enhanced architectures for instance segmentation. In this architecture, feature maps are extracted from the intermediate layers at multiple scales. It contains a bottom-up and top-down pathway with skip connections between them.

Hwang et al. propose Skipped-Hierarchical Feature Pyramid Networks[20] or SH-FPN to improve the imperfection in the nuclei segmentation of the original FPN architecture. The model further uses Stage residual connection and aggregation between reference and predicted feature map over FPN.

TSFD-Net [21] proposed by Ilyas et al. uses Tissue-Specific Feature Distillation(TSFD)[21]as the backbone. The model contains a bidirectional Feature Pyramid Network (BiFPN)[21] and combinational loss function for faster convergence and joint optimization of this model.

To overcome the limitations of all the current approaches, we propose a model that can efficiently extract all the information from a cytology image.

Chapter Three

PROPOSED METHODOLOGY

In this chapter, we have proposed a two step approach for automated nuclei segmentation in digital cytology images. In the first step, Faster R-CNN [22] is used for nuclei localization from the cytology images. Then, for each nucleus, the fragment of the image containing that nucleus is cropped using the localization properties obtained by Faster R-CNN. In the second stage, a systematic repetitive process is proposed, where for each cropped fragment, stain normalization with repeated dilation is applied, followed by a GabCut [23] method. Each subsequent iteration proceeds when the current iteration fails to segment.

3.1 DATA DESCRIPTION

In this experiment, we have used the JUCYT-V1 cytology dataset (ref: https://github.com/DVLP-CMATERJU/JUCYT_V1) that consists of 61 cytological images captured through a CMOS camera, attached to a trinocular microscope at 40x magnification. The dimension of each of these images is 960X 2048X 3 pixels (height X width X number of channels). The images in the dataset are categorized into two classes: Benign and Malignant. Figure 1 and Figure 2 show image samples for the Benign and malignant classes, respectively. This dataset is partitioned into training and validation sets in a 4:1 ratio. The number of images of Benign and malignant classes is shown in Table 1.

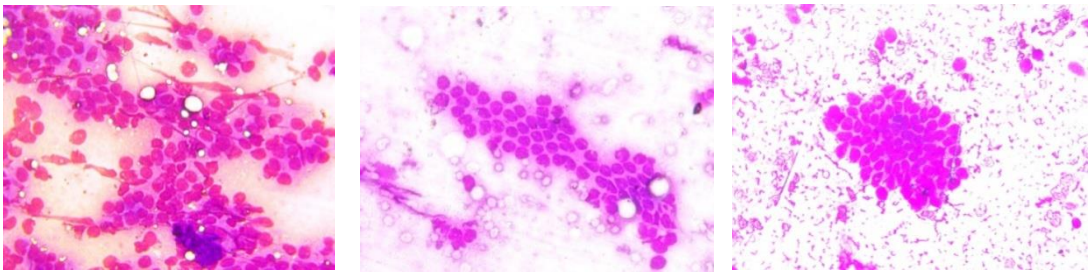


Figure 1: **Image samples for Benign tumor class**

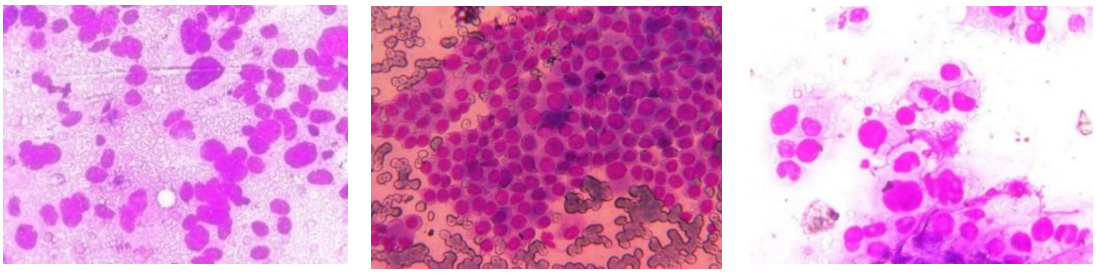


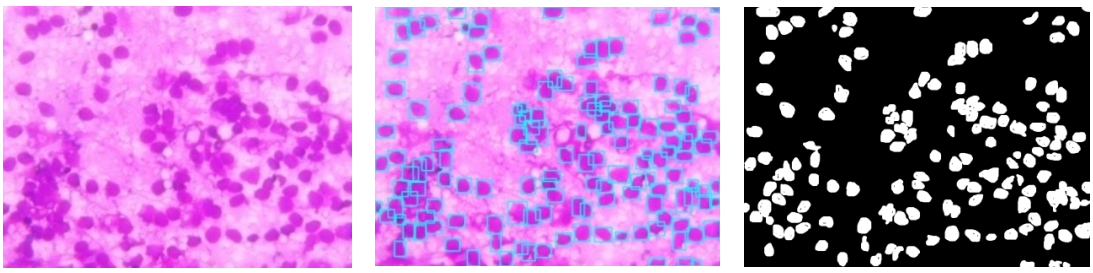
Figure 2:**Image samples for Malignant tumor class**

Table 1: Data distribution of the collected samples

Sub Dataset	Benign samples	Malignant samples	Total Samples
Training	33	16	49
Validation	6	6	12
Total	39	22	61

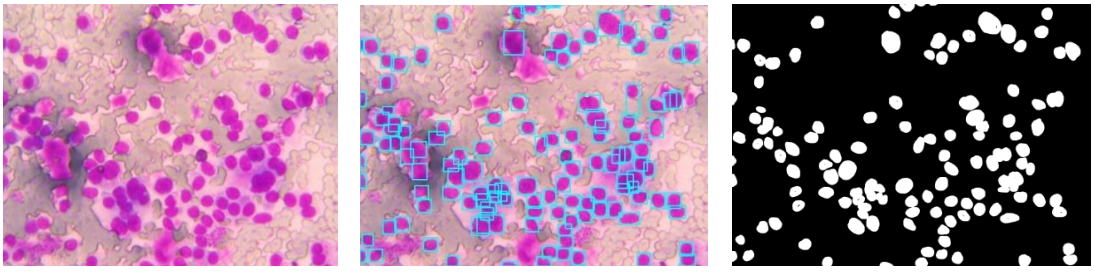
For each image in the dataset, there is a corresponding text file and a mask image containing the ground truth localization and segmentation details respectively. The text file contains several rows, where each row corresponds to a nucleus present in the images. Each row consists of 4 space-separate values. The first two values represent the x-coordinate and the y-coordinate of the uppermost left point of the nucleus in the image. The next two values represent the height and width of the nucleus.

The dimension of the segmentation masks is the same as that of their corresponding cytology images. These semantically segmented masks are binary images, where the foreground is the nucleus and the background is all other non-nuclei substances. Figure 3 and 4 show instances of Benign and Malignant class samples in the dataset.



(a) (b) (c)

Figure 3: **An instance of Benign class and its corresponding Localization and Segmentation Ground Truth: (a) Cytology image (b) Cytology image with localized nucleus (c) Segmented mask**



(a) (b) (c)

Figure 4: **An instance of Malignant class and its corresponding Localization and Segmentation Ground Truth: (a) Cytology image (b) Cytology image with localized nucleus (c) Segmented mask**

3.2 METHODOLOGY

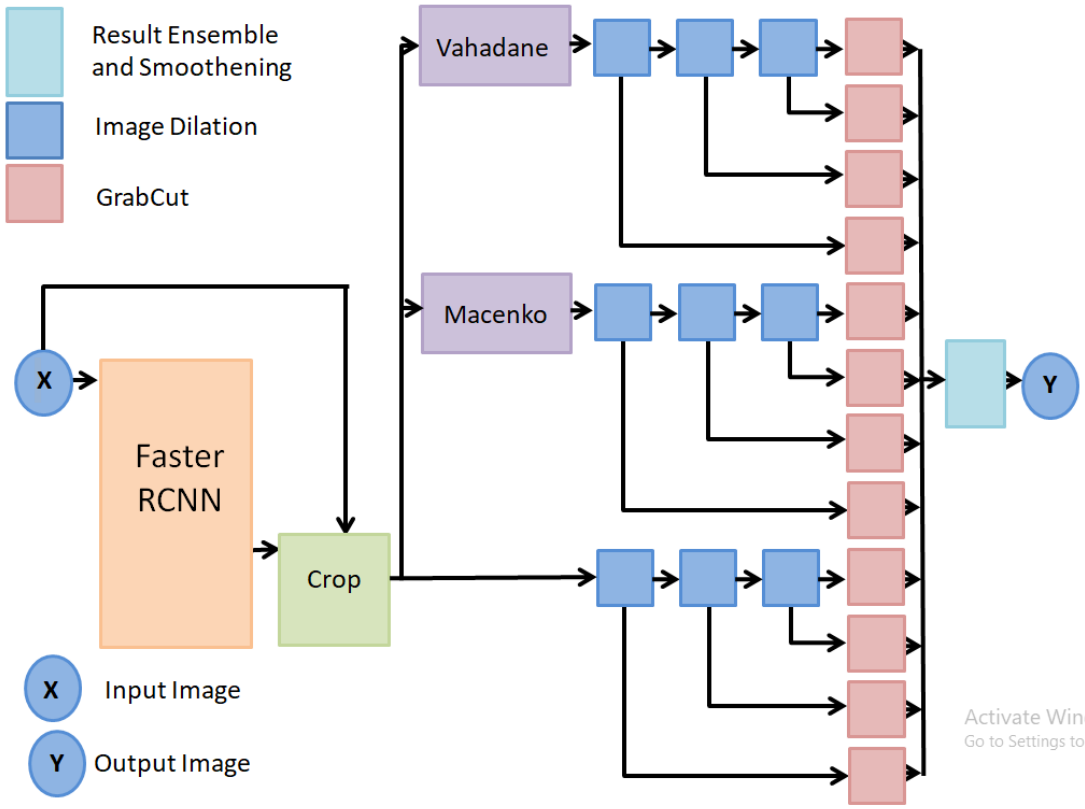


Figure 5: **Diagram for proposed segmentation method using localization model**

3.2.1 LOCALIZATION USING FASTER-RCNN

Faster RCNN[24], the third version of the R-CNN series (RCNN[25], Fast R-CNN[22] and Faster R-CNN[24]), is used in the proposed method for localization. In Faster R-CNN a region proposal network and a detection module produce a set of the bounding boxes and class offset for each object. The main component in the Faster R-CNN network is the base network, which significantly determines the accuracy. In the original Faster RCNN model, ZF [26] and VGG-16[27] models are used as the base models. This base model is being replaced by the Feature Pyramid Network, which is a more advanced architecture that makes the accuracy much better.

The Feature Pyramid Network (FPN) [19] is a convolutional neural network module that takes a single-scale image as input and creates a set of feature maps at multiple scales. FPN consists of mainly three modules that are as follows.-

- Bottom Up Pathway [19]: Bottom Up Pathway uses a feed-forward network that consists of convolution layers grouped into several blocks at each level. Feature maps are obtained from the last layer of each level. The deeper we go into the network, the more blocks produce feature maps of lower dimensions than the preceding blocks but have more semantic information.

- Top-down pathway [19]: In a top-down pathway, higher resolution feature maps are obtained by up-sampling, from the feature maps obtained from the previous level.

- Lateral connections[19]: Lateral connections merge feature maps of the same spatial dimensions obtained from the Top-Down and Bottom-Up pathways.

Feature Pyramid Network uses ResNet 50 as the backbone of our proposed method. For an image of dimensions (B, C, H, W) where, B=batch size, C= number of channels, H=Height of the image, W=Width of the image, the FPN produces feature maps of the following dimension:

- (B, 256, H/4,W/4)

- (B, 256, H/8,W/8)

- (B, 256, H/16,W/16)

- (B, 256, H/32,W/32)

- (B,256,H/64,W/64)

Using this set of feature maps, the Region Proposal Network (RPN) [24] makes 2000 region proposals with the help of the anchor boxes. The RPN consists of two heads—

the classification head and the regression head. The classification head is responsible for determining the objectness of the proposed region, and the regression head is responsible for producing the box offset based on the anchor boxes.

However, there might be more than one bounding box predicted for a specific object. Faster RCNN further uses a Non-Max Suppression [24] module which removes those bounding boxes which overlap with a higher priority bounding box when the ratio of intersection and union of two bounding boxes is greater than 0.5.

3.2.1.1 TRAINING PROCESS

The Faster R-CNN is trained on the training dataset of the JUCYT-v1 dataset. The training images and the text file containing the bounding boxes are fed to the model using a data loader. But for each offset of the bounding box, the top-left coordinate of the nucleus and the bottom-right coordinate are found.

Losses in each iteration are calculated for the Region Proposal Network (RPN) and ROI Head Detector. The RPN returns the loss obtained from both the classification and the regression part. For the regression part, smooth l1 loss is used, and for the classification loss, cross-entropy loss with logits is used.

Stochastic Gradient Descent [29] is used, as the optimizer with 0.9 as momentum. The learning rate is set to 0.005 with weight decay as 0.0005.

The model is trained for 10 epochs on the training dataset, with a batch size of 1. After every epoch, the performance of the model is evaluated on the validation dataset. F1 score of the validation set is calculated for every epoch. The model corresponding to which the validation set produces the best F1 score is retained. The Faster R-CNN is trained on the training dataset of the JUCYT-v1 dataset. The training images and the text file containing the bounding boxes are fed to the model using a data loader. But for each offset of the bounding box, the top-left coordinate of the nucleus and the bottom-right coordinate are calculated.

Losses in each iteration are calculated for the Region Proposal Network (RPN) and ROI Head Detector. The RPN returns the loss obtained from both the classification and the regression part. For the regression part, smooth l1 [28] loss is used, and for the classification loss, cross-entropy [29] loss with logits is used.

Stochastic Gradient Descent [30] is used, as the optimizer with 0.9 as momentum. The learning rate is set to 0.005 with weight decay set to 0.0005.

The model is trained for 10 epochs on the training dataset, with a batch size of 1. After every epoch, the performance of the model is evaluated on the validation dataset. F1 score of the validation set is calculated for every epoch. The model corresponding to the validation set that produces the best F1 score is retained.

3.2.2 SEGMENTATION USING GRABCUT METHOD

Grabcut [23] is a traditional segmentation algorithm that segments the foreground from the background using the rectangle specified by the user. The bounding box off-set from the Faster R-CNN model is used to crop the image fragment containing the corresponding nucleus. Semantic segmentation is performed on this cropped fragment using GrabCut.

3.2.2.1 GRAY SCALE IMAGE

Assume a grayscale image z= (z_1,z_2,…..,z_N) where N is the number of pixels. The segmented image is interpreted as an array of opacity value α= (α_1, α2,….., α_N) where α=0 represents the background and α=1 represents the foreground. Parameter Θ [23] is introduced and describes the image as a grey level distribution.

$$\Theta = \{\boldsymbol{h}(\boldsymbol{z};\boldsymbol{\alpha}), \boldsymbol{\alpha} = \mathbf{0}, \mathbf{1}\}$$

…..(1)

Histograms h(z;α) are normalized to sum to 1 over the grey-level range Θ consists of histogram grey level values.

3.2.2.2 ENERGY MINIMIZATION FOR SEGMENTATION

E is defined as an energy function in such a way that its minimum should corresponds to more reasonable segmentation in a way that guides both for foreground and back-ground pixels with coherent opacity reflecting a propensity to solidity. It takes the form of Gibb energy[23] as

$$E(\alpha, \theta, z) = U(\alpha, \theta, z) + V(\alpha, z)$$

…..(2)

The coherence of the z with α is defined as U[23],where

$$U(\alpha, \theta, z) = \sum n - logh(z_n; \alpha_n)$$

…..(3)

The smoothness is defined as

$$V(\alpha, z) = \gamma \sum_{(m,n)\in C} dis(m,n)^{-1} [\alpha_n \neq \alpha_m] exp - \beta(z_m - z_n)^2$$

…..(4)

Here C is the set of pairs of adjacent pixel and (n,m) is an instance of C. dis(.) is the distance of the adjacent pixel. The value of β is chosen as follows-

$$\beta = (2D(z_m - z_n)^2)^{-1}$$

…..(5)

Since the energy function is already defined the segmentation can be estimated as global minimum.

$$\hat{\alpha} = argmin_{\alpha} E(\alpha, \theta)$$

…..(6)

This method forms the basis of the GrabCut[23] algorithm.

3.2.2.3 COLOR DATA MODELING

For an RGB image it becomes unfeasible to construct a sufficient color space histogram and hence uses GMM[23]. A full-covariance Gaussian mixture with K components is taken for each GMM. An additional vector $k = \{k_1,..., k_n,..., k_N\}$ is inserted in the optimization framework to deal with the GMM tractably, with $k_n = \{1,..., K\}$ assigning a unique number to each pixel. The energy function E[23] and data term U[23] are modified as-

$$E(\alpha, k, \theta, z) = U(\alpha, k, \theta, z) + V(\alpha, z)$$

…..(7)

$$U(\alpha, k, \theta, z) = \sum_{n} D(\alpha_n n, k_n, \theta, z_n)$$

…..(8)

Here D[23] is defined as-

$$D(z_n, \alpha_n, k_n, zn) = -logp(z_n, \alpha_n, k_n, \theta) - log\pi(\alpha_n, k_n)$$

…..(9)

Here p(·) is a Gaussian probability distribution, and π(·) are mixture of weighting coefficients. So D becomes-

$$D(\alpha n, kn, \theta, zn = -\log\pi(\alpha_n, k_n) + \frac{1}{2}\log\det\Sigma(\alpha_n, k_n) + \frac{1}{2}[z_n - \mu(\alpha_n, k_n)]^T\Sigma(\alpha_n, k_n) - 1\,[z_n - \mu(\alpha_n, k_n)]$$

…..(10)

So the parameters become

$$\theta = \{\pi(\alpha, k), \mu(\alpha, k), \Sigma(\alpha, k), \alpha = 0,1, k = 1 \ldots K\}$$

…..(11)

This algorithm then uses iterative energy minimizing algorithm until the convergence criterion is met.

3.2.2.4 SEGMENTATION

Given an input image, the bounding box offset generated by the Faster R-CNN algorithm is used to crop the corresponding nucleus fragment present in the image.is

Two stain normalization methods, Vahadane [31] and Macenko [32], are applied to the cropped fragment for each nucleus to obtain the two corresponding stain normalized cropped image fragments. there are three cropped image fragments for each nucleus. Additionally, repeated dilation is performed on each fragment, due to which we have 12 cropped images for each nucleus. All the versions of the cropped nuclei in each image are shown in Table2 .

Table 2: Different version of preprocessing on input image

Index	Stain	Number of Dilations
1	Original	0
2	Original	1
3	Original	2
4	Original	3
5	Vahadane	0
6	Vahadane	1
7	Vahadane	2
8	Vahadane	3
9	Macenko	0
10	Macenko	1
11	Macenko	2
12	Macenko	3

For a particular nucleus, the cropped images are prioritized in the order in which they are indexed in table 2. So for a particular cropped image, if segmentation on the current version results in more than 20% of the pixels being classified as foreground, then that version is retained. Otherwise, the next version is segmented. This is done for all the 12 versions of the cropped image, in an iterative manner.

At the end, an image B is created which contains only background pixels. All the segmented cropped fragments are super imposed on B to obtain the final input image.

3.2.1 ALGORITHM

Algorithm 1: FasterR-CNN based Nuclei Segmentation

Input: *A*

Output: *C*

1. From the A, obtain N bounding boxes of all the regions of interest using FasterRCNN with Res-Net FPN backbone.

2. Using the set of bounding, crop the M fragments from the input image X image, containing the nucleus.

3. For every cropped fragment I=1 to N

4. Apply stain normalization J and K using Vahadane and Macenko respectively on each cropped fragment.

5. Apply repeated dilation on I,J, and K to obtain I1,I2,I3,J1,J2,J3,K1,K2,K3.

6. Let L={I,J,K,I1,I2,I3,J1,J2,J3,K1,K2,K3}

7. for P in L

8. if(1/5th of segmented image > foreground)

9. Set Q_I=P and break

10. end if

11. end for

12. end for

13. Create a background image B

14. Superimpose Q={Q_1,Q_2,…..Q_M) on B to obtain C

CHAPTER FOUR

RESULT AND DISCUSSIONS

The output obtained by the proposed model is a binary segmented image as shown in Figure 8(c) and Figure 9(c). Figure 8(b) and Figure 9(b) show the output obtained in the intermediate localization step.

For both the localization and the segmentation tasks, a model is evaluated on the validation set and is compared with the other state-of-the-art models. The evaluation metrics used by the model are Mean IOU[33], precision, recall and F1 score. Mean IOU for an object is defined as the intersection of the actual and predicted object and the total area occupied by the actual and predicted object.

$$Mean\ IOU = \frac{Intersection}{Union} \quad \text{.....(12)}$$

Precision[34], Recall[34]and F1-Score[34] are calculates as shown in equation 12,13, 14 and 15 respectively.

$$Precision = \frac{TruePositive}{TruePositive+FalsePositive} \quad \text{.....(13)}$$

$$Recall = \frac{TruePositive}{TruePositive+FalseNegative} \quad \text{.....(14)}$$

$$F1Score = \frac{2\times Recall\times Precision}{Recall+Precision} \quad \text{.....(15)}$$

In localization, true positive detections refer to those predicted bounding boxes which overlap with the actual bounding box with an IOU of 0.5. All the other predicted and actual bounding boxes are False Positive and False Negative samples respectively. The comparative study in Table 3 depicts that the Faster R-CNN with ResNet FPN backbone outperforms all the other models.

Table 2 shows that all the evaluation metrics for YOLOv3 [35] are less than 0.5 in the current dataset. The architecture of the YOLOv3 uses one forward propagation module of several convolutional layers, where each layer extracts certain features of the input image. The deeper layers extract more complex layers. While Yolov3 is good at detecting objects with complex structural features, it performs poorly in the current dataset that consists of highly overlapping nuclei.

However, Faster-RCNN uses a region proposal network (RPN) to propose regions from which the bounding box offsets are determined. However, the performance of the Faster R-CNN algorithm also varies based on the backbone model used for the region proposal. Table3 shows that the F1-score of the Faster R-CNN model with Res-Net [26] Feature Pyramid Network as the backbone is 20% higher than Faster R-CNN that uses MobileNet [36] Feature Pyramid Network as the backbone. This is because the main feature of ResNet is the skip connections that reduce the gradient descent problem, thereby increasing the efficiency of the model. On the other hand, MobileNet uses a combination of depthwise and pointwise convolution to increase the processing speed.

The Faster R-CNN with Resnet FPN backbone has shown better results than the other localization models. However, it should be noted that, in some training images, some nuclei are not labelled. Those nuclei are properly detected using the FasterR-CNN algorithm with Res-Net FPN backbone, thereby increasing the number of false positives and hence the low recall.

The F1-score for the Faster-CNN model is 0.8496 on the validation dataset. This result is obtained after the 8th epoch of the model training process. The statistical metrics for training the model are shown in Figure 6.

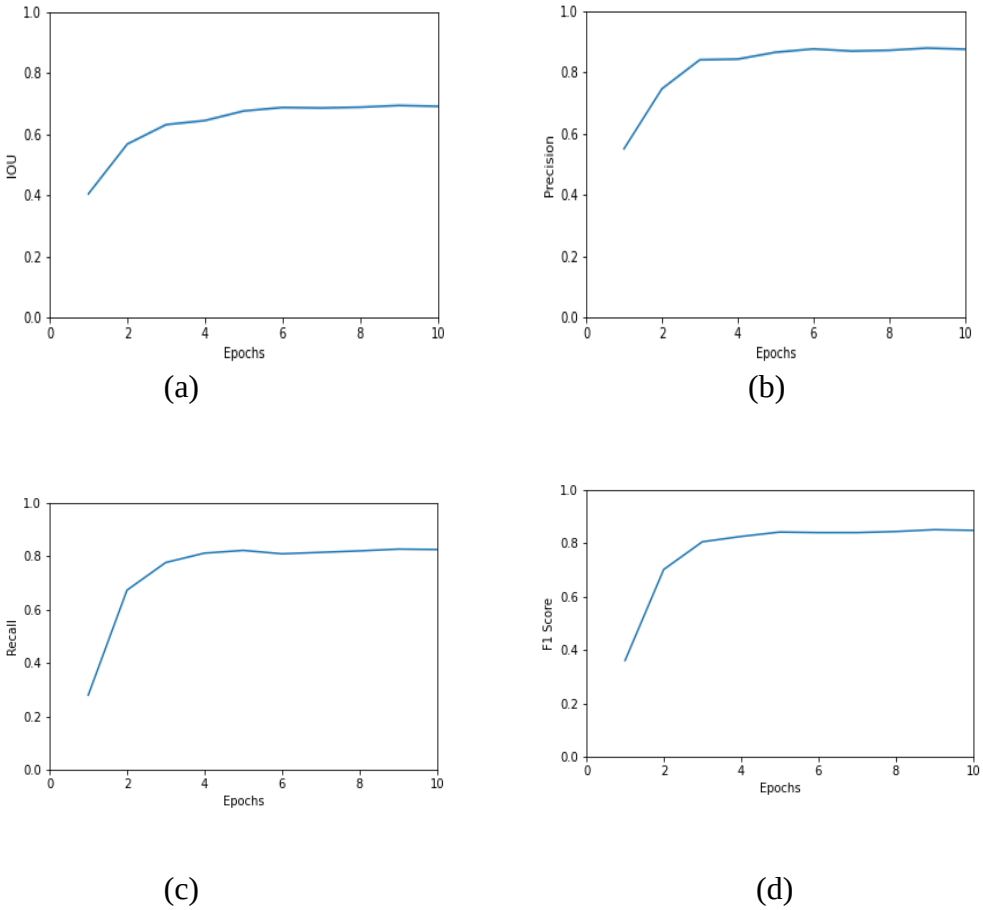


Figure 6: **Statistical metric for (a) IOU,(b)Precision,(c)Recall and (d) F1 score with epoch**

Table 3:Comparative study of state-of –art localization technique

Model	Mean IOU	Precision	Recall	F1 Score
Yolov3[35]	0.2455	0.3662	0.3212	0.3431
FasterRCNN with Mobilenet FPN[19]	0.5844	0.6777	0.6365	0.6492
FasterRCNN with Resnet FPN[19]	**0.6930**	**0.8780**	**0.8255**	**0.8496**

The bounding box information obtained from the Faster R-CNN with ResNet FPN backbone is used for cropping the fragment from the image containing the nucleus, for segmentation.

True Positive, False Positive and False Negative in segmentation use the idea of pixel-wise overlapping. Table4 shows the comparison of the efficiency of some segmentation models. U-Net [10] has the lowest F1 score compared to the other two segmentation models. It is first trained on the individual nucleus image fragments in a supervised manner. However, in a cropped image fragment of a particular nucleus, parts of the adjacent nuclei may be present. Hence, U-Net fails to properly learn nuclei segmentation. As a result, it produces a nearly circular shape for all the nuclei.

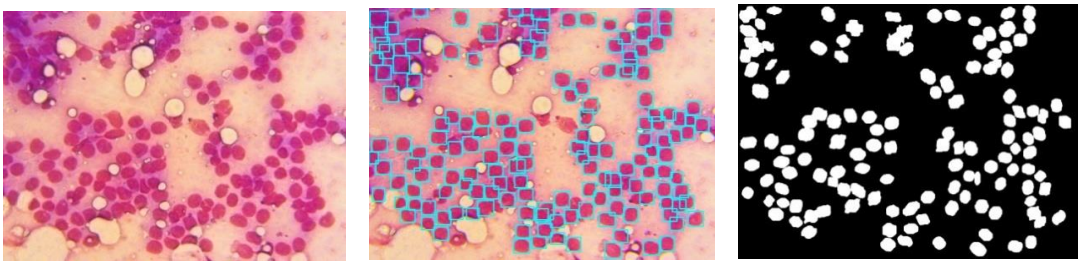
Thresholding [36] preserves better boundary information than U-Net. Thresholding is an unsupervised algorithm that uses a threshold pixel value to segment the image into foreground and background. But as the cytology images have variations in color gradient, the thresholding technique creates very disruptive edges of segmented nuclei in the output image. This loss of boundary information makes it impractical for isolating the nucleus from the image.Table 4 shows that the GrabCut [23] method has performed better than U-Net and Thresholding. The F1 score for the GrabCut is 0.7931.The output for localization and segmentation for the benign and malignant tumour classes is shown in Figure 8 and Figure 9 respectively.

Table 4:Comparative study of state of art Segmentation work

Model	Mean IOU	Precision	Recall	F1 Score
Unet[10]	0.5122	0.6344	0.6422	0.6383
Thresholding[36]	0.6241	0.7138	0.7343	0.7233
Proposed Methodology	**0.6582**	**0.7857**	**0.8075**	**0.7931**

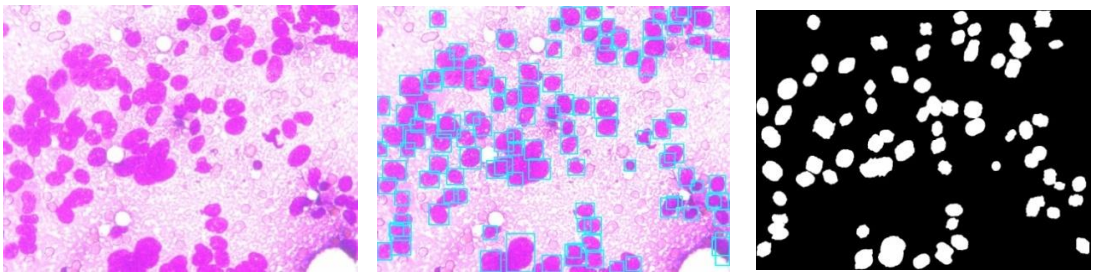
The overall performance of the segmentation and the intermediate localization model is shown in Table 4. The final segmentation result depends on both the performance

of the Faster R-CNN and the GrabCut algorithm. But the localization result depends only on the Faster R-CNN algorithm. Hence, the F1 score of the segmentation is less than the localization F1 score.



(a) (b) (c)

Figure 7: **An example of Segmentation of the Benign tumor class using the proposed method (a) Original image and its corresponding (b) Localized output and (c) Segmented output**



(a) (b) (c)

Figure 8:**An example of Segmentation of the Malignant tumor class using the proposed method (a) Original image, (b) Localized output and (c) Segmented output**

Table 5:Result of Proposed model

Dataset	Localization				Segmentation			
	IOU	Precision	Recall	F1 Score	IOU	Precision	Recall	F1 Score
JUCYT-V1	0.6930	0.8780	0.8255	0.8496	0.6582	0.7857	0.8075	0.7931

CHAPTER FIVE

CONCLUSION

In this work, we have proposed a two-step approach-based nuclei segmentation technique from cytology images. Here, nuclei are first localized by Faster RCNN with backbone model ResNet-50 FPN and after that, they are segmented by the GrabCut method. In this two-step approach, we have found a good F1 score of 79%, which are better than other state-of-art approaches. There are lots of challenges for the segmentation task, like overlapping nuclei and prominent labelling. Also, it is observed that the model could uniquely segment most of the nuclei, but some highly overlapped nuclei regions are interpreted as a single nucleus. Moreover, there are numerous nuclei present in the image, and this uses GrabCut to individually segment each nucleus separately. On top of that, it creates several versions of the same nucleus image fragment to enhance segmentation. Consequently, a convolutional neural network model capable of segmenting the input image in a single step can be proposed.

In future, we would like to explore unsupervised localization techniques due to the lack of labelled medical data. In the GrabCut-based segmentation process, we have loosed many informative regions of the nuclei structures. So here, we can use the Conditional Random Field model to fine-tune the nuclei segmentation. Also, we would like to improve the localization performances by implementing a new deep learning-based localization model.

REFERENCES

[1] "Cancer." https://www.who.int/news-room/fact-sheets/detail/cancer (accessed Jun. 26, 2022).

[2] "Cancer: A global threat - BBC News." https://www.bbc.com/news/health-26031748 (accessed Jun. 26, 2022).

[3] P. Mathur *et al.*, "Cancer Statistics, 2020: Report From National Cancer Registry Programme, India," *JCO Glob. Oncol.*, no. 6, pp. 1063–1075, Nov. 2020, doi: 10.1200/GO.20.00122.

[4] Y. M. George, B. M. Bagoury, H. H. Zayed, and M. I. Roushdy, "Automated cell nuclei segmentation for breast fine needle aspiration cytology," *Signal Processing*, vol. 93, no. 10, pp. 2804–2816, 2013, doi: 10.1016/j.sigpro.2012.07.034.

[5] S. Mitra, S. Dey, N. Das, S. Chakrabarty, M. Nasipuri, and M. K. Naskar, *Identification of Benign and Malignant Cells from Cytological Images Using Superpixel Based Segmentation Approach*, vol. 836. Springer Singapore, 2018. doi: 10.1007/978-981-13-1343-1_24.

[6] "What is Artificial Intelligence (AI)? - India | IBM." https://www.ibm.com/in-en/cloud/learn/what-is-artificial-intelligence (accessed Jun. 26, 2022).

[7] "Image Classification vs. Object Detection vs. Image Segmentation | by Pulkit Sharma | Analytics Vidhya | Medium." https://medium.com/analytics-vidhya/image-classification-vs-object-detection-vs-image-segmentation-f36db85fe81 (accessed Jun. 26, 2022).

[8] "Convolutional Neural Network (CNN) in Machine Learning - GeeksforGeeks." https://www.geeksforgeeks.org/convolutional-neural-network-cnn-in-machine-learning/ (accessed Jun. 26, 2022).

[9] A. Mouelhi, M. Sayadi, F. Fnaiech, K. Mrad, and K. Ben Romdhane, "Automatic image segmentation of nuclear stained breast tissue sections using color active contour model and an improved watershed method," *Biomed. Signal Process. Control*, vol. 8, no. 5, pp. 421–436, 2013, doi: 10.1016/j.bspc.2013.04.003.

[10] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," pp. 1–8.

[11] S. E. A. Raza *et al.*, "Micro-Net: A unified model for segmentation of various objects in microscopy images," *Med. Image Anal.*, vol. 52, pp. 160–173, 2019, doi: 10.1016/j.media.2018.12.003.

[12] H. Oda *et al.*, "BESNet: Boundary-enhanced segmentation of cells in histopathological images," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11071 LNCS, pp. 228–236, 2018, doi: 10.1007/978-3-030-00934-2_26.

[13] Y. Zhou, O. F. Onder, Q. Dou, E. Tsougenis, H. Chen, and P. A. Heng, "CIA-Net: Robust Nuclei Instance Segmentation with Contour-Aware Information Aggregation," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11492 LNCS, pp. 682–693, 2019, doi: 10.1007/978-3-030-20351-1_53.

[14] U. Schmidt, M. Weigert, C. Broaddus, and G. Myers, "Cell detection with star-convex polygons," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11071 LNCS, pp. 265–273, 2018, doi: 10.1007/978-3-030-00934-2_30.

[15] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 386–397, 2020, doi: 10.1109/TPAMI.2018.2844175.

[16] N. A. Koohbanani, M. Jahanifar, and A. Gooya, "Nuclear Instance Segmentation using a Proposal-Free Spatially Aware Deep Learning Framework," pp. 1–9.

[17] S. Chen, C. Ding, and D. Tao, "Boundary-Assisted Region Proposal Networks for Nucleus Segmentation," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 12265 LNCS, pp. 279–288, 2020, doi: 10.1007/978-3-030-59722-1_27.

[18] Y. Kong, G. Z. Genchev, X. Wang, H. Zhao, and H. Lu, "Nuclear Segmentation in Histopathological Images Using Two-Stage Stacked U-Nets With Attention Mechanism," *Front. Bioeng. Biotechnol.*, vol. 8, no. October, pp. 1–7, 2020, doi: 10.3389/fbioe.2020.573866.

[19] X. Li, T. Lai, S. Wang, Q. Chen, C. Yang, and R. Chen, "Weighted feature pyramid networks for object detection," *Proc. - 2019 IEEE Intl Conf Parallel Distrib. Process. with Appl. Big Data Cloud Comput. Sustain. Comput. Commun. Soc. Comput. Networking, ISPA/BDCloud/SustainCom/SocialCom 2019*, pp. 1500–1504, 2019, doi: 10.1109/ISPA-BDCloud-SustainCom-SocialCom48970.2019.00217.

[20] H. Hwang, T. D. Bui, S. Il Ahn, and J. Shin, "Skipped-Hierarchical Feature Pyramid Networks for Nuclei Instance Segmentation," *2018 Asia-Pacific Signal Inf. Process. Assoc. Annu. Summit Conf. APSIPA ASC 2018 - Proc.*, no. November, pp. 689–693, 2019, doi: 10.23919/APSIPA.2018.8659795.

[21] T. Ilyas, Z. I. Mannan, A. Khan, S. Azam, H. Kim, and F. De Boer, "TSFD-Net: Tissue specific feature distillation network for nuclei segmentation and classification," *Neural Networks*, vol. 151, pp. 1–15, 2022, doi:

10.1016/j.neunet.2022.02.020.

[22] R. Girshick, “Fast R-CNN,” *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2015 Inter, pp. 1440–1448, 2015, doi: 10.1109/ICCV.2015.169.

[23] C. Rother, V. Kolmogorov, and A. Blake, “‘GrabCut,’” *ACM Trans. Graph.*, vol. 23, no. 3, pp. 309–314, 2004, doi: 10.1145/1015706.1015720.

[24] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, 2017, doi: 10.1109/TPAMI.2016.2577031.

[25] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 580–587, 2014, doi: 10.1109/CVPR.2014.81.

[26] A. M. C. Antioquia, D. S. Tan, A. Azcarraga, W. H. Cheng, and K. L. Hua, “ZipNet: ZFNet-level Accuracy with 48× Fewer Parameters,” *VCIP 2018 - IEEE Int. Conf. Vis. Commun. Image Process.*, vol. 2013, no. Ilsvrc 2013, pp. 1–4, 2018, doi: 10.1109/VCIP.2018.8698672.

[27] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, pp. 1–14, 2015.

[28] “SmoothL1Loss — PyTorch 1.11.0 documentation.” https://pytorch.org/docs/stable/generated/torch.nn.SmoothL1Loss.html (accessed Jun. 24, 2022).

[29] “Cross-Entropy Loss Function. A loss function used in most… | by Kiprono Elijah Koech | Towards Data Science.” https://towardsdatascience.com/cross-

entropy-loss-function-f38c4ec8643e (accessed Jun. 24, 2022).

[30] S. Ruder, “An overview of gradient descent optimization algorithms,” pp. 1–14, 2016, [Online]. Available: http://arxiv.org/abs/1609.04747

[31] A. Vahadane *et al.*, “Structure-Preserving Color Normalization and Sparse Stain Separation for Histological Images,” *IEEE Trans. Med. Imaging*, vol. 35, no. 8, pp. 1962–1971, 2016, doi: 10.1109/TMI.2016.2529665.

[32] M. Macenko *et al.*, “A method for normalizing histology slides for quantitative analysis,” *Proc. - 2009 IEEE Int. Symp. Biomed. Imaging From Nano to Macro, ISBI 2009*, pp. 1107–1110, 2009, doi: 10.1109/ISBI.2009.5193250.

[33] “Intersection over Union (IoU) for object detection - PyImageSearch.” https://pyimagesearch.com/2016/11/07/intersection-over-union-iou-for-object-detection/ (accessed Jun. 24, 2022).

[34] “Understanding Confusion Matrix | by Sarang Narkhede | Towards Data Science.” https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62 (accessed Jun. 24, 2022).

[35] J. Redmon, A. Farhadi, and C. Ap, “YOLOv3 : An Incremental Improvement”.

[36] P. K. Sahoo, S. Soltani, and A. K. C. Wong, “A survey of thresholding techniques,” *Comput. Vision, Graph. Image Process.*, vol. 41, no. 2, pp. 233–260, 1988, doi: 10.1016/0734-189X(88)90022-9.