International Journal of Computational Intelligence Systems
Vol. **14(1)**, 2021, *pp.* 122–131
DOI: https://doi.org/10.2991/ijcis.d.201030.001; ISSN: 1875-6891; eISSN: 1875-6883
https://www.atlantis-press.com/journals/ijcis/

Research Article

Pyramidal Nonlocal Network for Histopathological Image of Breast Lymph Node Segmentation

Zehra Bozdağ*, Fatih M. Talu

Computer Science Department, Inonu University, Malatya, 44280, Turkey

ARTICLE INFO

Article History
Received 08 Jun 2020
Accepted 20 Oct 2020

Keywords
Deep learning
Histopathological image segmentation
Nonlocal network
Machine learning

ABSTRACT

The convolutional neural networks (CNNs) are frequently used in the segmentation of histopathological whole slide image-(WSI) acquired breast lymph nodes. The first layers in deep network architectures generally encode the geometric and color properties of objects in the training set, while the last layers encode the distinctive and detailed properties between classes. Modern segmentation approaches (DeepLabV3+, SegNet, PSPNet) are realized by evaluating these layers together. However, having a high parameter space of all these networks increases the calculation costs and prevents the researchers from working more effectively. In this study, we present a new pyramid-structured segmentation network (NonLocalSeg). Although the proposed network has low parameter space, its segmentation performance is similar to current architectures. The integration of the Nonlocal Module (NLM-a form of attention mechanism) or Asymmetric Pyramid Nonlocal block (APNB) into classical pyramid-built architectures has led to the reduction of network depth and narrowing of the parameter space while enabling coding of low and high image features. These mechanisms suppressed the unfocused background image, emphasizing the focused foreground object. As a result of a series of ablation experiments carried out, it is seen that the NLM and APNL mechanisms give the succeeded results. Although the network architectures adapting these mechanisms contain fewer parameters than current networks, it is observed that they have a similar accuracy (mean intersection over union [IoU]) range.

1. INTRODUCTION

The digital pathology image is the digitization of slides which are tissues taken from the human body created by special processes using special scanners. These images are also called histopathological whole slide image (WSI). Pathological diagnosis is the examination of these images to determine whether there are diseases such as cancer. For the correct diagnosis of pathology, an accurate and reliable separation (segmentation) of tissues as normal or cancerous is essential.

With the proliferation of WSI data, many researches have worked on disease detection using machine learning methods [1–4]. As WSIs are large images, it is not possible to give a WSI directly into the machine learning model, considering GPU memory consumption. Commonly, when WSI is examined, small-sized images are analyzed independently of the whole. This method is called patch-based evaluation [5–7]. The image in Figure 1 shows patches from tumorous and normal tissues. Global content knowledge is lost in approaches that use only local patches, which hinders the detection of tumorous tissues.

Through segmentation, pixels with similar color and texture properties in the image are grouped. This process, which is assigned an independent label value to each group, is used in medical images such as ultrasound, histopathological images, and so on. In recent years, convolutional neural networks (CNNs) have been used in segmentation studies and many special architectures have been developed [8–11] reports that the use of deep learning algorithms, especially CNNs, in breast histopathological image analysis is quite common. Although there are intensive studies regarding segmentation in histopathological images, the success of tumorous tissue segmentation in medical images is still not at the desired level. This can be explained by a number of factors. Firstly, the variety of

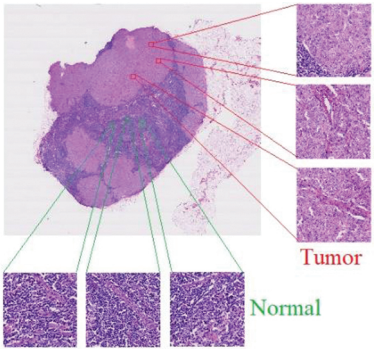

Figure 1 | Tumor and normal tissue sample patches in a whole slide image (WSI).

* *Corresponding author. Email: zbozdag@harran.edu.tr*

tissues in medical images is extremely high. Figure 2 shows the tumor and normal tissue patches obtained in the histopathological WSI of breast lymph nodes. The color and texture differences are observed even among the images belonging to the same class. Secondly, the diversity of devices used in tissue preparation, lighting, and scanning increase the tissue variety. Even external factors such as dyeing and folding can also distort image quality. Figure 2 shows the texture folding (overlap) and stains. The mentioned factors make the tumorous tissue segmentation process difficult. As a result of these factors, an improved segmentation architecture is needed, which will not only explain the correlation between patches in the same class but also distinguish the differences between textures.

In pathological images, two different deep learning architectures are used to determine long-range dependencies between objects with different tissue structures and geometric shapes—stroma, cells, epithelium, tumor area. The first is deep convolutional networks with small kernel sizes; the second is shallow networks with large kernel sizes. In both convolutional networks, the number of hyperparameters and computational complexity are high.

A module called Global Convolution Network (GCN) can improve the aforementioned flaws in shallow networks with large kernel sizes. Using a more efficient method and without increasing computational costs, this module with large kernel sizes captures global contexts.

In addition, the use of attention mechanism is envisaged to increase the segmentation capability. It appears that the nonlocal network (NLN) module is used to accomplish this [12–14]. Wang *et al.* have developed the nonlocal module (NLM) combining CNN with traditional nonlocal means, which is used to reduce noise in the image. This module has been used to model spatial-temporal dependencies in video sequences and has increased the performance of existing networks. It is also used in the literature for image generation and segmentation problems [13].

Zhu *et al.* have demonstrated that the NLM has high GPU consumption and computing cost. They have developed the Asymmetric Pyramid Nonlocal Block (APNB). This block has been used in segmentation and provided high performance in terms of speed [14].

The innovative contributions of the proposed method can be summarized as follows:

- A hybrid tumor segmentation architecture can capture global context information at a high speed. The architecture combines two important modules and an effective approach: GCN, NLM, and multi-scale segmentation approach.
- Compared with state-of-the-art segmentation networks, the proposed network has reached a high mean intersection over union (IoU) value (68.6%) and has obtained a close pixel accuracy (PA) value. These results present the success of the proposed method for segmenting tumors in histopathological images of the lymph nodes of the breast.

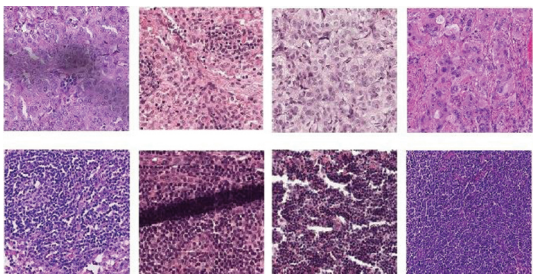

Figure 2 | Collected sample tissues patches. Top: tumor, bottom: normal.

The remainder of this paper is organized as follows: Section 2 provides a brief review of the related literature. We introduce the proposed hybrid architecture in Section 3. Section 4 lays out the experimental results and analysis. Discussion and future directions are given in Section 5. Finally, Section 6 presents a brief summary and our conclusions.

2. RELATED WORKS

2.1. Histopathological Image Segmentation

Fully convolution network (FCN) is the first image segmentation network using the convolutional layers. FCN has been developed by replacing the fully connected layers of standard CNNs used for classification with convolution layers [15]. This architecture has contributed to the development of many segmentation networks. Unet is one of the extended versions of FCN used in the medical images [16]. Layers close to the exit of the network makes the low-dimensional feature maps the same size as the input image using the deconvolution method. The segmentation accuracy is increased with skip connections. Skip connection is the addition of feature maps taken from different levels in the encoder to the decoder. Segmentation networks have been formed based on these networks [17–22].

There are some works segmenting tumors from histopathological medical images with interesting approaches. The work of Takahama *et al.* [23] is one of them. Patches obtained from WSI were given to the standard classification CNN to provide classification vector results. Then, these vectors were combined while taking into account the patch locations and given as an input to the segmentation network [9]. In another work is that the patches of different levels of WSIs were used for segmentation [10]. [24–26] are some of the current studies using CNN in the segmentation processes of histopathological image structures.

Hybrid network models in medical image segmentation have attracted a great deal of attention lately. For example, a segmentation architecture with fewer parameters has been developed by combining Unet, which is the classical and still the most popular architecture, with the residual learning framework for 3D MRI images segmentation [27]. In another study using a hybrid network, a stream that obtains shape information has been added to the classical Unet architecture [28].

2.2. Encoder–Decoder Architecture

The standard autoencoder architecture has an encoder and a decoder module. The encoder module is the part that gradually reduces feature maps size and captures advanced semantic

information. The decoder module, on the other hand, is the module that gradually converts discriminative features (low-level resolution) learned in the encoder module to high-resolution. This architecture is used in many different areas of image processing such as object detection [29], pedestrian detection [30], facial landmark detection [31], and semantic segmentation [16,32,33]. In image segmentation, newly developed modules are added to the features taken from different scales in the encoder [34–37].

2.3. Effective Convolution Modules

Filter masks with high kernel dimensions are used to detect the relationships between objects in high resolution images. However, increasing kernel size causes the expansion of the parameter space, thereby increasing the cost. GCN is a way to effectively use large kernels [12,36,38]. The internal structure of the GCN module is shown in Figure 3(a). It is based on the parallel use of different-sized convolution layers. Considering the k kernel size, the cost of convolution with the *kxk* mask is quite high. Instead, the combination of $1xk + kx1$ and $kx1 + 1xk$ convolution results is preferred. Thus, the cost of calculation decreases from $O\left(k^2\right)$ to $O\left(2/k\right)$ [12].

The convolution process causes shifts, scatter or thinning on the edge pixels (boundary regions) of the image objects. This situation causes problems in terms of perception. The boundary highlight module also called Boundary-aware Module (BM) is used to overcome such problems [12,36]. BM improves these disorders that occur in boundary information as a result of convolution. BM is integrated into the proposed network structure during the training. The internal structure of this module is given in Figure 3(b).

2.4. Nonlocal Module

NLM was first used in video classification. It was formulated with inspiration from nonlocal means used to reduce noise in the image. It concerns the self-attention (SA) module. SA mechanism allows inputs to interact with each other and find out which they should pay more attention to. SA was introduced as a module in language translation networks and later used in the field of images [39–42]. SA can be viewed as a form of the nonlocal mean [13].

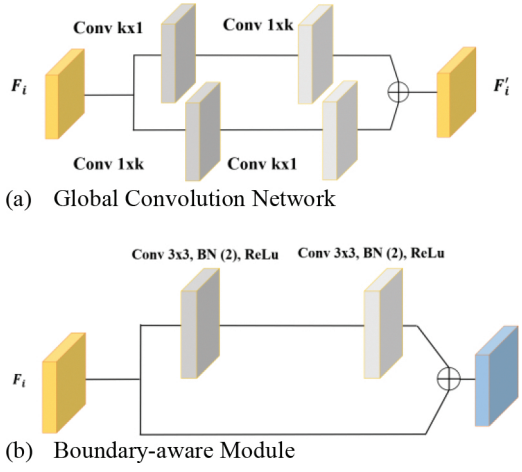

Figure 3 | Effective convolution modules: (a) Global Convolution Network (GCN) module and (b) Boundary-aware Module (BM).

3. METHOD

Potential differences can be found between the feature representations of the pixels of the same class. These differences cause inconsistency in the segmentation result when the contextual information of the image is not well coded [12]. NLM by analyzing the problem in this context is intended to create relationships between features.

The proposed network is based on fully convolutional network-based encoder–decoder architecture. The encoder achieves semantically effective features from the image and the decoder uses the achieved features to make semantic segmentation. The architecture of the network is given in Figure 4. The multi-stage segmentation strategy developed by Zhao *et al.* is used in the proposed network [36].

Between the stages of the decoder, 3 × 3 kernel convolutions, ReLu, and batch normalization, and finally 2 × 2 max pooling have been performed. Table 1 the dimensions of input and output features of the stages are given.

In each stage, the feature maps channels are reduced through the 1 × 1 convolution layer. We call the feature maps $\{F_i\}$ in the $i\,(i = 0, 1, 2, 3)$ stage. F_i first passes through the GCN module. The module output is called F_i'. Then, M_i is obtained by passing the module F_i'. through NLM. M_i combines with the next stage feature map F_{i+1}'. and passes though the NLM again. This process takes place for all stages. At the end, M_3 combines with F_3, F_2, F_1, F_0 respectively, and segmentation prediction is done in the last layer, as shown in Figure 4.

3.1. Large Kernel Convolution

In the proposed network, different-sized kernels are used for the modules (GCN), where feature maps pass through each stage. The dimensions of the feature maps passing through the stages decrease as they progress. General contexts are captured using size-appropriate kernels for feature maps that are shrinking in size. The kernel sizes are 15, 7, 5, and 3 respectively for F_0, F_1, F_2, and F_3.

3.2. Nonlocal Module

Standard NLM is given Figure 5. [13]. Let denote $F_i' \in \Re^{CxWxH}$ an input feature map to the NLM, where *i, C, W, H* represent the feature level, feature map's channel, width and height dimensions, respectively. Three transform functions $W\phi$, $W\theta$, and $W\gamma$ are used to flattened F_i' to size *CxN*, *N* represents the total number of the spatial locations. $N = HxW$. $\phi \in \Re^{C'xWxH}$, $\theta \in \Re^{C'xWxH}$ and $\gamma \in \Re^{C'xWxH}$ are outputs of transform functions. C' is the channel number of the new embeddings.

$$\phi = w_\phi\left(F_i'\right), \theta = w_\theta\left(F_i'\right), \gamma = w_\gamma\left(F_r^i\right) \quad (1)$$

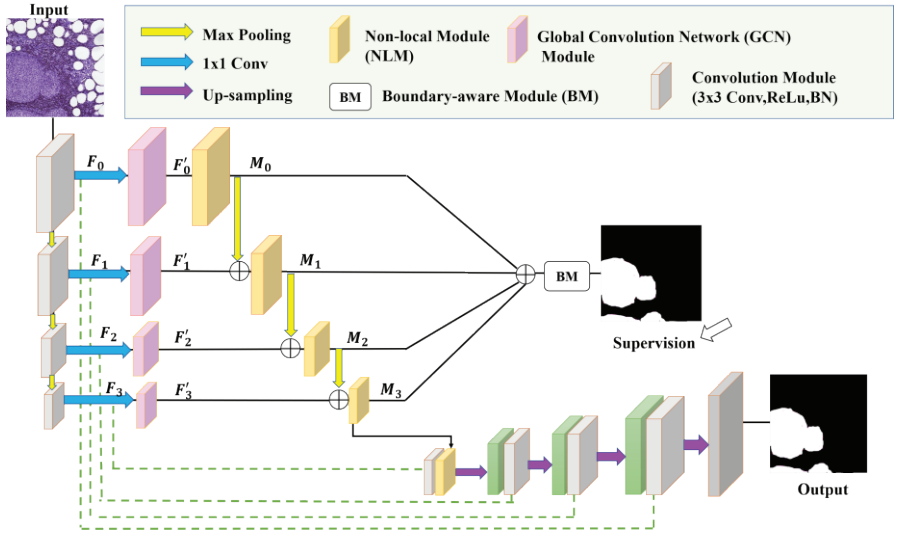

Figure 4 | Pyramid nonlocal network architecture (NonLocalSeg).

Table 1 | Input and output dimensions of the stages.

Stage	Input	Output
F_0	3×256^2	64×128^2
F_1	64×128^2	128×64^2
F_2	128×64^2	256×32^2
F_3	512×32^2	1024×16^2

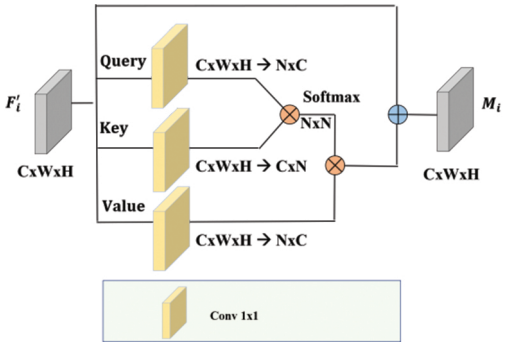

Figure 5 | Standard nonlocal module (NLM).

Then the similarity matrix $V \in \Re^{NxN}$ is calculated by a matrix multiplication as

$$V = \phi^T x \theta \tag{2}$$

Afterward, normalization is applied to V to get a unified similarity matrix as

$$\vec{V} = f(V) \tag{3}$$

As indicated by [13], the normalizing function f can take the structure SoftMax. SoftMax is comparable to the SA system and demonstrated to function admirably in numerous tasks, for example, machine translation and image generation.

For every location in γ, the output of the attention layer is

$$O = \vec{V} x \gamma^T \tag{4}$$

where $O \in R^{NxC}$ by referring to the design of the NLM, the final output is given by

$$M_i = O^T + F'_i \tag{5}$$

Feature maps $\left(F'_i\right)$ are passed through the NLM. Global context features M_i added with F'_{i+1} and pass through a block consisting of convolution, batch normalization and dropout.

3.3. Asymmetrical Pyramid Nonlocal Block

It is observed that Eqs. (2) and (4) calculation costs are high for standard NLM. The two matrix products have $O\left(CN^2\right) = O\left(CW^2H^2\right)$ time complexities. The large matrix multiplication is the leading cause of the inefficiency of the NLM. The method that Zhu *et al.* developed to reduce the cost is given Figure 6. [14]. The method is called APNB.

Returning to the design of the NLM, changing N to a small number S is equivalent to sampling several representative points from θ and γ instead of feeding all the spatial points, as illustrated in Figure 6. Consequently, computational complexity could be considerably decreased.

P_θ, P_γ are sampling modules applied to θ and γ respectively which are obtained in the standard NLM in Eq. (1). $\theta_P \in \Re^{\breve{C}xS}$, $\gamma_P \in \Re^{\breve{C}xS}$ are outputs of sampling modules. S is the number of sampling points. The remaining process after sampling is the same as the standard NLM process.

$$\theta_P = P_\theta(\theta), \gamma_P = P_\gamma(\gamma)$$

$$V_P = \phi^T x \theta_P \tag{6}$$

$$O_P = \vec{V}_P x \gamma_P^T \tag{7}$$

$$Y = cat\left(O_P^T, F\right) \tag{8}$$

APNB is faster than standard NLM. The number of feature maps channels was reduced using Spatial Pyramid Pooling (SPP) as shown in Figure 6. SPP is used as a sampling module. In this study, NLM is replaced with APNB to analyze its impact on segmentation accuracy.

3.4. Pyramidal Feature Gathering for Segmentation

Once the encoder has acquired features that contain contextual information, the decoder must effectively combine these features to complete the segmentation. The merging process starts with the addition of low-dimensional feature maps. As seen in Figure 4, the features containing contextual information combine with features containing spatial information from the encoder. After increasing the size of feature maps by up-sampling, it combines with feature maps from the previous stage.

3.5. Deep Supervision

Another important factor affecting success in segmentation is to use the segmentation results obtained from different stages of the network during training [43]. This process is called supervision and we have applied this extra control in the proposed network. Overall, the process is used as an additional cost in training by combining the prediction results obtained from the network stages. Thus, the middle layers of the network learn to distinguish semantic discriminatory features. There are similar studies in the literature [8,36,44]. Considering all the losses, the total objective function is

$$L_{SegTotal} = L_{final} + \lambda L_{supervision} \tag{9}$$

where the first term refers to the loss of the segmentation results at the end of the network. The second term refers to the supervision loss. Total cost ($L_{SegTotal}$) is the sum of two Binary Cross-Entropy (BCE) loss functions (L_{final},$L_{supervision}$). BCE is given in Eq. (10). Here, λ is a hyperparameter that controls the weighting of the loss and is determined as 0.4.

$$L = 1/N \sum_{n=1}^{N} \left(y_n \log y'_n + \left(1 - y_n\right) \log \left(1 - y'_n\right)\right) \tag{10}$$

y is the ground truth, y' is the prediction, and N is the number of pixels.

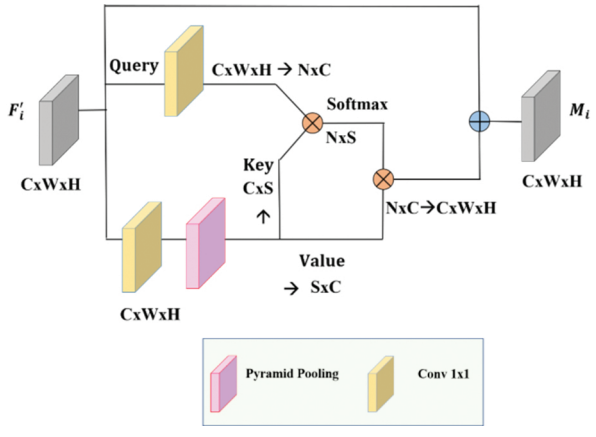

Figure 6 | Asymetrical pyramid nonlocal block (APNB).

3.6. Dataset

WSI used in this study was obtained from the contest named CAMELYON16 in 2016. WSI has 40× lens magnification. The average size of images is 1 GB and dimensions are 200,000 × 100,000. They are in the TIFF file format with different resolution levels.

The aim of the contest is the automatic tumor detection in the histopathological images of lymph nodes taken from the breast cancer patient. The areas with tumors in all WSIs were determined under the supervision of expert pathologists and recorded as two-level images (masks). The competition organization gave the position of tumor on slides to the researchers in XML file format. The slide images prepared in the pyramid structure have 8 levels between 0 and 7 in total. Mask images of WSIs can be obtained using the provided files. Table 2 provides detailed information about the dataset [45].

From the Camelyon16 training set, a total of 8000 patches with tumor and normal tissues were collected from 111 Tumor WSIs. These patches created were used in the training of networks. From the Camelyon16 test set, a total of 2000 patches with tumor and normal tissue were collected from 25 Tumor WSIs. These patches created were used to obtain the test results of the networks. The size of all patches in the data set is 512 × 512. The total patches size for the used dataset is 10.000.

Random image cropping, one of the data augmentation techniques, was used in the training and testing stages of networks. As an input to the networks, randomly cropped images with 256 × 256 size from 512 × 512 images are given. Data augmentation techniques are used in the training [46].

Figure 7(a) shows a WSI structure, it has many levels as shown. In Figure 7(b) shows the WSI 3rd level's patch containing tumor and normal tissue. Finally, Figure 7(c) shows the mask of the image in (b). The white part represents the tumor. The size of the images in the figure (b) and (c) is 512 × 512.

Table 2 | CAMELYON16 dataset.

	Normal	Tumor	Total
Train	160	111	271
Test	80	49	129
Total	240	160	400

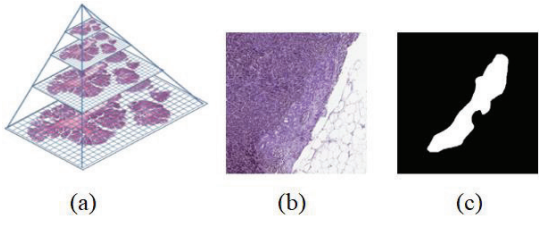

Figure 7 | (a) The structure of a whole slide image (WSI), (b) a patch of level 3 of WSI, and (c) a mask of the patch.

4. EXPERIMENTS

4.1. Experimental Setting

In this section, we have provided dataset, training details and evaluation metric used in networks.

4.1.1. Training and implementation details

The proposed network is trained with Adam optimizer, and weight-decay = 0.0005. In the experiments, the weight layers are started without any pretrained parameters and the initial bias is set to 0. At the same time, all the networks in Table 3 are run 10 epochs. The learning rate initial value is 0.0001, after the 5th epoch, the learning rate is divided by 10. We adopt dropout after each fusion of the NLM and GCN [47]. We fine-tune our models with Batch-Norm [48] enabled when it is applied (batch size is 2). Using the Pytorch library, we used a computer with a Nvidia Quadro M4000 (under CUDA 9.0 without other ongoing programs) graphics card with 8GB of GDDR5 GPU memory. Windows Server 2012 R operating system is installed on the computer.

4.1.2. Evaluation

IoU, also known as the Jaccard Index, is the standard measure of accuracy in semantic segmentation. *IoU* is measured as the ratio between the underlying truth and the predicted segmentation mask. The mean *IoU* is calculated by summing all the *IoU*s and dividing by the number of classes. In Eq. (11), let B_{ti} be the region area estimated for class i and B_{gi} real object region area. Eqs. (11) and (12) are given IoU_i and *mean IoU* formulate, respectively.

$$IoU_i = area\left(B_{ti} \cap B_{gi}\right) / area\left(B_{ti} \cup B_{gi}\right) \tag{11}$$

$$meanIoU = 1/N \sum_{i=1}^{N} IoU_i \tag{12}$$

Another measurement metric for segmentation is PA. The formula of the PA is given Eq. (13). PA is the percent of pixels in the image that are classified correctly.

$$PA = (TP + TN)/(TP + FP + TN + FN) \tag{13}$$

Abbreviations our case TP denotes correctly classified pixels, FN denotes pixels not detected, and FP denotes background pixels classified as parts of a class. And finally, TN denotes the background pixels which are classified correctly.

Table 3 Comparison with current segmentation networks.

Methods	Backbone	Pretrained	Mean IoU (%)	PA (%)
DeepLabV3+ [32]	Xception	Yes	61.9	**85.6**
FCN8s [15]	VGG16	Yes	15.1	30.3
UNet [16]		No	58.5	79.1
PSPNET [49]	ResNet101	Yes	53.1	78.0
ICNet [34][a]		No	49.2	66.4
SegNet [33][a]	VGG19	Yes	61.6	81.9
NonLocalSeg		No	**68.6**	**85.1**

(a) Real-time segmentation networks.
Bold indicates the success of the method used.

4.2. Result

4.2.1. Ablation study on the proposed model

We have made some ablation studies on the proposed network and compared the results in terms of accuracy (*mean IoU, PA*) and speed (*second*). Ablations results are given in Table 4. The efficiency of the modules (NLM, APNB) integrated into the GCN module base network has been investigated. The network with NLM increases the performance of the base network from 62.2% to 68.6% (*mean IoU*). By replacing NLM with APNB a minor performance decrease (68.6% → 67.7%). Table 4 also shows the PA values of the networks. In experimental studies, it has been observed that the network with NLM has the highest PA value.

In Table 4, the average execution time of the networks has been given in second for a 256 × 256 × 3 histopathological image. The network with NLM is the slowest compared to others, which is expected that because more products are made in NLM.

4.2.2. Comparison to state-of-the-art networks

A large number of epochs are needed to converge the training of networks with a high number of parameters. The proposed network has been converged faster than all the networks we have trained using the same epoch number. The success of the proposed network (68.6% *mean IoU*) has been shown in Table 3. Using low parameter space and the effective modules for a segmentation network have yielded better results. In Table 3, the 5th column gives the PA measurement values. The DeepLabV3+ network has the highest PA value. The proposed network has achieved the second-highest PA value, with a slight difference between the highest PA value.

Table 5 shows the number of parameters (million-M) of the successful networks that have passed the 60% mean *IoU* threshold and the average execution time spent to predict an image segmentation mask. The proposed network has the least number of parameters and is the fastest one. The average execution time of the proposed network is 0.040 seconds for a 256 × 256 × 3 histopathological image.

Table 4 Ablation study on the proposed network.

	With NLM	With APNB	Without NLM and APNB
Mean IoU (%)	68.6	67.7	62.2
PA (%)	85.1	84.1	82.5
Avg. Execution Time (second)	0.040	0.035	0.028

NLM, nonlocal module; APNB, asymmetric pyramid nonlocal block.

Table 5 Comparison of the execution time and the number of parameters.

Methods	Avg. Execution Time (Second)	# Param (M)
DeepLabV3+ [32]	0.148	54
SegNet [33]	0.072	9
NonLocalSeg	0.040	6

Figure 8 in Section 7 shows the predicted segmentation masks of networks which have more than 60% of *mean IoU*. The first and second columns of Figure 8 show histopathological images and their ground truth masks. In the remaining columns of Figure 8, the tumor probability masks produced by Deeplabv3+, SegNet, and the proposed network are given respectively. When the tumor probability masks produced by the networks are examined in Figure 8, it is seen that they are different from the ground truth masks. However, it is determined that the mask results of the proposed network closer to the ground truth masks. The edges of the SegNet results are not clear. In DeepLabv3+, it is seen that tumor and normal tissue borders are intertwined.

5. DISCUSSION AND FUTURE RESEARCH

The segmentation process is difficult in histopathological images due to heterogeneous structure and tissue diversity. We propose a segmentation network model has been proposed model to make this process faster and more successful than existing networks. By keeping the parameter space of the network lower, the network has been trained faster. A low epoch value is used to train all networks. The results are presented in tables and figures. The comparison networks and the proposed network's visual results are given in Figure 8. When the visual results of all networks are compared with the ground truth masks, the one-to-one similarity is not obtained. However, the visual results of the proposed network are the most similar to ground truth masks.

Additionally, the shortest average execution time belongs to the proposed network. Speed is crucial for high-resolution images (WSI) analysis so that fast segmentation results can be obtained.

For future work, some tactics are used during training to improve the performance of networks in segmentation. For example, [35,37] use higher batch size and epoch value during training. The impact on network performance will be investigated using these tactics.

6. CONCLUSION

In this study, we implement a novel architecture of focus for the segmentation of histopathological images. This architecture network integrates a multi-scale strategy to combine semantic information at various levels and NLM to gradually combine relevant contextual features. To approve our strategy, we conduct the test on WSIs from the CAMELYON16 Challenge. We give extensive tests to examine the impact of the individual components of the proposed architecture. In addition, we compare our model to existing methods that have been lately present for the natural scene and medical image segmentation. The achievement of our method may be explained by having an enhanced ability to model rich contextual dependencies over local features of images.

The proposed segmentation network integrates GCN [12], NLM [13], APNB [14] modules and uses pyramid feature fusion [36] and segmentation supervision [24,37] strategies. Finally, the proposed network model shows that it is a valid tool to assist in rapid tumor segmentation, with its low parameter size and performance of segmentation close to current segmentation networks.

7. QUALITATIVE RESULT

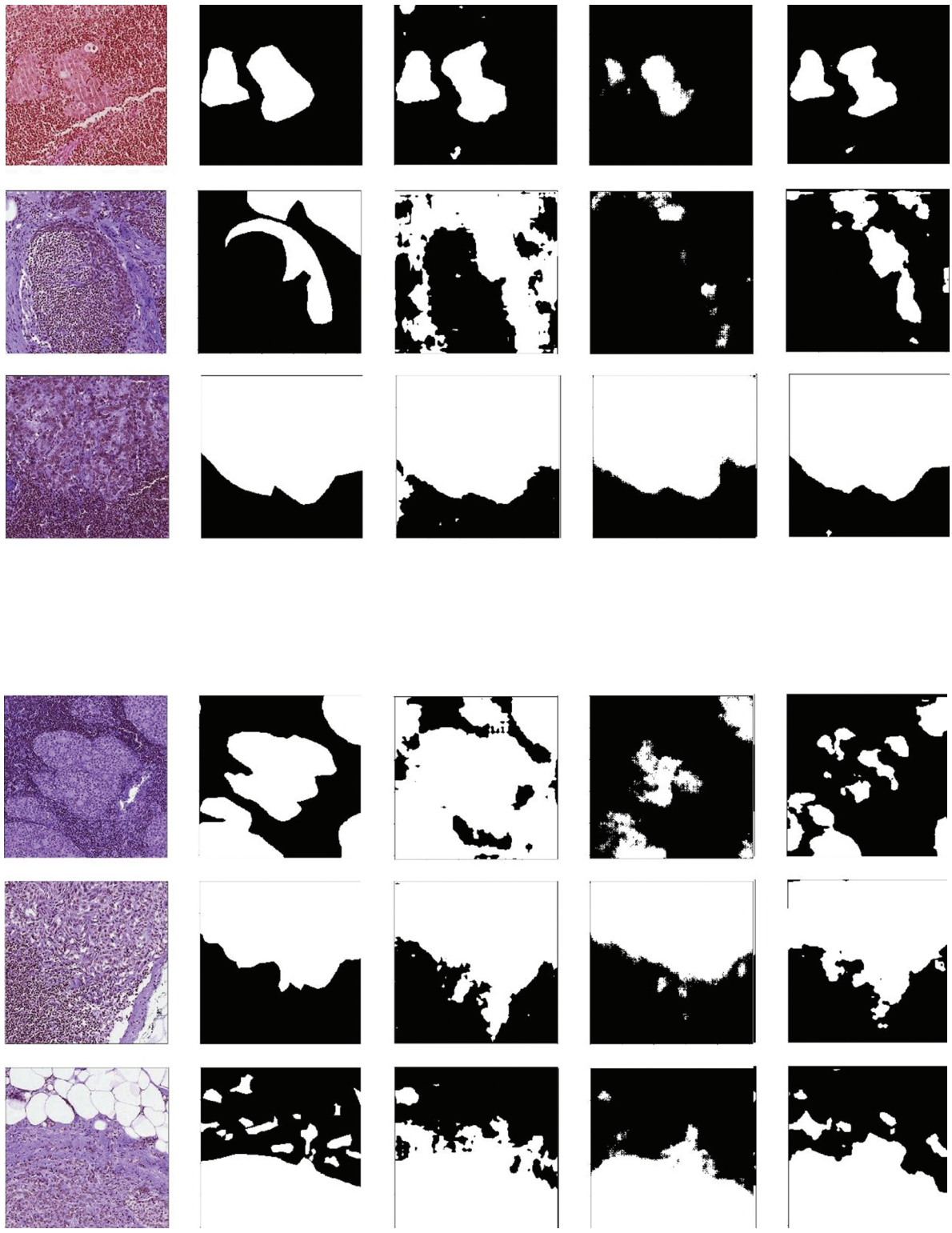

Figure 8 | Qualitative results of the networks are given. The first two columns show input images and ground truth masks. The remaining columns show probability masks of DeepLabv3+, SegNet, and NonLocalSeg networks, respectively.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

AUTHORS' CONTRIBUTIONS

Zehra Bozdag contributed to the design and implementation of the research, to the analysis of the results. Zehra Bozdag wrote the manuscript in consultation with Fatih Talu.

Funding Statement

The author(s) received no financial support for the research, authorship, and/or publication of this article.

REFERENCES

[1] Y. Song, J J. Zou, H. Chang, W. Cai, Adapting fisher vectors for histopathology image classification, in Proceeding of International Symposium Biomedical Imaging, Melbourne, Australia, 2017, pp. 600–603.

[2] B.E. Bejnordi, G. Litjens, M. Hermsen, N. Karssemeijer, J.A.W.M. van der Laak, A multi-scale superpixel classification approach to the detection of regions of interest in whole slide histopathology images, in Medical Imaging 2015: Digital Pathology, Orlando, Florida, USA, 2015, vol. 9420, p. 94200H.

[3] N. Riaz, S.L. Wolden, D.Y. Gelblum, J. Eric, Multi-instance multi-label learning for multi-class classification of whole slide breast histopathology images, IEEE Trans. Med. Imaging. 37 (2018), 316–325.

[4] S. Reis, *et al.*, Automated classification of breast cancer stroma maturity from histological images, IEEE Trans. Biomed. Eng. 64 (2017), 2344–2352.

[5] H. Ni, *et al.*, Multiple visual fields cascaded convolutional neural network for breast cancer detection, in: X. Geng, B.H. Kang (Eds.), PRICAI 2018: Trends in Artificial Intelligence, PRICAI 2018, Lecture Notes in Computer Science, vol. 11012, Springer, Cham, Switzerland, 2018.

[6] M. Peikari, M.J. Gangeh, J. Zubovits, G. Clarke, A.L. Martel, Triaging diagnostically relevant regions from pathology whole slides of breast cancer: a texture based approach, IEEE Trans. Med. Imaging. 35 (2016), 307–315.

[7] A.D. Belsare, M.M. Mushrif, M.A. Pangarkar, N. Meshram, Classification of breast cancer histopathology images using texture feature analysis, in IEEE Region 10th Annual International Conference on Proceedings/TENCON, Macao, China, 2016.

[8] Y. Wang, *et al.*, Deep attentive features for prostate segmentation in 3D transrectal ultrasound, IEEE Trans. Med. Imaging. 38 (2019), 2768–2778.

[9] S. Takahama, *et al.*, Multi-stage pathological image classification using semantic segmentation, in IEEE International Conference on Computer Vision, Seoul, South Korea, 2019, pp. 10701–10710.

[10] Z. Li, R. Tao, Q. Wu, B. Li, DA-RefineNet: a dual input whole slide image segmentation algorithm based on attention, 2019, pp. 1–11. http://arxiv.org/abs/1907.06358

[11] R. Krithiga, P. Geetha, Breast cancer detection, segmentation and classification on histopathology images analysis: a systematic review, Arch. Comput. Methods Eng. (2020).

[12] C. Peng, X. Zhang, G. Yu, G. Luo, J. Sun, Large kernel matters - improve semantic segmentation by global convolutional network, in Proceeding - 30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017), Honolulu, HI, USA, 2017, pp. 1743–1751.

[13] X. Wang, R. Girshick, A. Gupta, K. He, Non-local neural networks, in IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018, pp. 7794–7803.

[14] Z. Zhu, M. Xu, S. Bai, T. Huang, X. Bai, Asymmetric non-local neural networks for semantic segmentation, in 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, South Korea, 2019, pp. 593–602.

[15] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 2015.

[16] O. Ronneberger, P. Fischer, T. Brox, U-Net: convolutional networks for biomedical image segmentation, in: N. Navab, J. Hornegger, W. Wells, A. Frangi (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Lecture Notes in Computer Science, Springer, Cham, Switzerland, 2015.

[17] S. Lian, Z. Luo, Z. Zhong, X. Lin, S. Su, S. Li, Attention guided U-Net for accurate iris segmentation, J. Vis. Commun. Image Represent. 56 (2018), 296–304. http://www.sciencedirect.com/science/article/pii/S1047320318302372

[18] Z. Zhou, M.M.R. Siddiquee, N. Tajbakhsh, J. Liang, UNet++: redesigning skip connections to exploit multiscale features in image segmentation, IEEE Trans. Med. Imaging. 39 (2020), 1856–1867.

[19] Z. Zeng, W. Xie, Y. Zhang, Y. Lu, RIC-Unet: an improved neural network based on unet for nuclei segmentation in histology images, in IEEE Access. 7 (2019), pp. 21420–21428.

[20] J. Hu, *et al.*, S-UNet: a bridge-style U-Net framework with a saliency mechanism for retinal vessel segmentation, in IEEE Access. 7 (2019), pp. 174167–174177.

[21] H. Li, *et al.*, CR-Unet: a composite network for ovary and follicle segmentation in ultrasound images, in IEEE J. Biomed. Heal. Informat. 24 (2020), pp. 974–983.

[22] Y. Weng, T. Zhou, Y. Li, X. Qiu, NAS-Unet: neural architecture search for medical image segmentation, IEEE Access. 7 (2019), 44247–4457.

[23] T. Takikawa, D. Acuna, V. Jampani, S. Fidler, Gated-SCNN: gated shape CNNs for semantic segmentation, in 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, South Korea, 2019, pp. 5228–5237.

[24] T. Wan, L. Zhao, H. Feng, D. Li, C. Tong, Z. Qin, Robust nuclei segmentation in histopathology using ASPPU-Net and boundary refinement, Neurocomputing. 408 (2020), 144–156.

[25] Y. Kurmi, V. Chaurasia, N. Kapoor, Design of a histopathology image segmentation algorithm for CAD of cancer, Optik. 218 (2020), 164636.

[26] S. Rezaei, A. Emami, N. Karimi, S. Samavi, Gland segmentation in histopathological images by deep neural network, in 2020 25th International Computer Conference, Computer Society of Iran (CSICC), Tehran, Iran, 2020, pp. 1–5.

[27] M. Baldeon-Calisto, S.K. Lai-Yuen, AdaResU-Net: multiobjective adaptive convolutional neural network for medical image segmentation, Neurocomputing. 392 (2020), 325–340.

[28] J. Sun, F. Darbeha, M. Zaidi, B. Wang, SAUNet: shape attentive U-Net for interpretable medical image segmentation, 2020. http://arxiv.org/abs/2001.07645

[29] T.Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, S. Belongie, Feature pyramid networks for object detection, in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017.
[30] T. Liu, T. Stathaki, Faster R-CNn for robust pedestrian detection using semantic segmentation network, Front. Neurorobot. 12 (2018), 64. https://www.frontiersin.org/article/10.3389/fnbot.2018.00064
[31] H. Lai, *et al.*, Deep recurrent regression for facial landmark detection, IEEE Trans. Circuits Syst. Video Technol. 28 (2018), 1144–1157.
[32] L. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, Encoder-decoder with atrous separable convolution for semantic image segmentation, in: V. Ferrari, M. Hebert, C. Sminchisescu, Y. Weiss (Eds.), Computer Vision – ECCV 2018, Lecture Notes in Computer Science, Springer, Cham, Switzerland, 2018, pp. 833–851.
[33] V. Badrinarayanan, A. Kendall, R. Cipolla, SegNet: a deep convolutional encoder-decoder architecture for image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 39 (2017), 2481–2495.
[34] H. Zhao, X. Qi, X. Shen, J. Shi, J. Jia, ICNet for real-time semantic segmentation on high-resolution images, in: V. Ferrari, M. Hebert, C. Sminchisescu, Y. Weiss (Eds.), Computer Vision – ECCV 2018, Lecture Notes in Computer Science, Springer, Cham, Switzerland, 2018.
[35] Y. Yuan, X. Chen, J. Wang, Object-contextual representations for semantic segmentation, 2019. http://arxiv.org/abs/1909.11065
[36] Z. Zhao, H. Lin, H. Chen, P.A. Heng, PFA-ScanNet: pyramidal feature aggregation with synergistic learning for breast cancer metastasis analysis, in: D.Shen, *et al.* (Eds.), Medical Image Computing and Computer Assisted Intervention – MICCAI 2019, Lecture Notes in Computer Science, vol. 11764, Springer, Cham, Switzerland, 2019, pp. 586–594.
[37] K. Sun, *et al.*, High-resolution representations for labeling pixels and regions, 2019. http://arxiv.org/abs/1904.04514
[38] C. Peng, *et al.*, MegDet: a large mini-batch object detector, in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018.
[39] H. Zhang, I. Goodfellow, D. Metaxas, A. Odena, Self-attention generative adversarial networks, 2019. https://arxiv.org/pdf/1805.08318
[40] A. Brock, J. Donahue, K. Simonyan, Large scale GaN training for high fidelity natural image synthesis, 2019. https://arxiv.org/abs/1809.11096
[41] J. Fu, *et al.*, Dual attention network for scene segmentation, in 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 2018, pp. 3141–3149.
[42] X. Wang, Z. Cai, D. Gao, N. Vasconcelos, Towards universal object detection by domain attention, in 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 2019, pp. 7281–7290.
[43] C.Y. Lee, S. Xie, P.W. Gallagher, Z. Zhang, Z. Tu, Deeply-supervised nets, J. Mach. Learn. Res. 38 (2015), 562–570. https://arxiv.org/abs/1409.5185
[44] A. Sinha, J. Dolz, Multi-scale self-guided attention for medical image segmentation, IEEE J. Biomed. Heal. Informatics. (2020), 1.
[45] F.A. Spanhol, L.S. Oliveira, C. Petitjean, L. Heutte, A dataset for breast cancer histopathological image classification, IEEE Trans. Biomed. Eng. 63 (2016), 1455–1462.
[46] A. Mikołajczyk, M. Grochowski, Data augmentation for improving deep learning in image classification problem, in 2018 International Interdisciplinary PhD Workshop (IIPhDW), Swinoujście, Poland, 2018.
[47] G.E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, R.R. Salakhutdinov, Improving neural networks by preventing co-adaptation of feature detectors, arXiv: 1207. 0580v1 [cs. NE], 2012, pp. 1–18. https://arxiv.org/abs/1207.0580
[48] S. Ioffe, C. Szegedy, Batch normalization: accelerating deep network training by reducing internal covariate shift, 2015. https://arxiv.org/abs/1502.03167
[49] H. Zhao, J. Shi, X. Qi, X. Wang, J. Jia, Pyramid scene parsing network, in Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017), Honolulu, HI, USA, 2017, pp. 6230–6239.