Dicle University
Journal of Engineering
https://dergipark.org.tr/tr/pub/dumf
duje.dicle.edu.tr

Research Article

Multi-scale Residual Segmentation Network for Histopathological Image

Zehra BOZDAĞ KARAKEÇİ[1*], Muhammed Fatih TALU[2]

[1] Harran University, Software Engineering Department, zbozdag@harran.edu.tr, Orcid No: 0000-0002-1119-5275

[2] Inonu University, Computer Engineering Department, fatihtalu@gmail.com, Orcid No: 0000-0003-1166-8404

ARTICLE INFO

Article history:

Received 13 July 2024
Received in revised form 5 September 2024
Accepted 24 September 2024
Available online 30 September 2024

Keywords:

deep learning, histopathological image segmentation, convolutional neural network

Doi: 10.24012/dumf.1500666

* Corresponding author

ABSTRACT

Deep learning is used in all areas of the image processing like object detection/localization, synthetic image generation, segmentation, tracking, and others. It is frequently used especially in medical image segmentation field since it provides rapid response during the treatment process. The fact that natural images contain different types of noise, patterns, and structures and the lack of distinctive quantitative information still makes the segmentation problem very challenging. The classical networks having high parameters have a long training time. The need of less training time for high parameter networks and high segmentation accuracy has led us to develop a new network. In this study, a state-of-the-art autoencoder network (MSRSegNet) is proposed to perform segmentation. Unlike conventional autoencoder approaches, it consists of encoder, fusion and decoder blocks. In encoder and decoder blocks, Multi-scale Residual Blocks are used to share information between blocks and to detect features on different scales. In fusion block, Atrous Spatial Pyramid Pooling (ASPP) module is used to preserve multi-scale contextual information. Segmentation architectures, such as Deeplabv3+, SegNet, and U-Net, with well-established backbones from the literature, are used for performance comparison on histopathological images extracted from the Camelyon16 dataset. As a result, it was observed that the proposed segmentation network has high accuracy (69% mean intersection over unit (mIOU)) and fast segmentation performance (0.061sec. for an image with 256x256)

Introduction

According to the data of the World Health Organization, breast cancer ranks first among the causes of women's death [1]. In order to diagnose this disease and to determine the treatment, tissues are taken from the body by biopsy method. Tissues are examined under a light microscope for the manifestation of the disease. This procedure is called histopathological examination in medical language. Generally, the shape, size, color, and distribution of nucleoids are taken into consideration when the diagnosis of breast cancer is considered in histopathological examination [2]. Traditional histopathological examination is performed manually. The pathologist examines the tissues under a light microscope and diagnoses them accordingly. Sometimes he consults with other pathologists to make the diagnosis because the variety of tissues is very high. This approach has at least two major drawbacks: the first is a laborious process, the tissues need to be carefully examined. Second is that the decision of a pathologist is not objective [3].

In digital pathology, there are special scanners that are used for scanning pathologic slides. The images produced by these scanners are called Whole slide Images (WSI). WSIs are glass slides of high magnification digitized glass samples. In the early days, this WSI was used for telepathology, clinical training to assist pathologists. With the development of computer-based medical image analysis in recent years, these images have been used intensively [3]. For example, many studies such as tissue classification, detection, registration and segmentation in WSI of different tissue samples have been done [4]–[9]. The most important benefit of using WSI in computer-based analysis is that it helps pathologists make quick decisions. It reveals overlooked findings when analyzing WSI by the pathologist. Therefore, it reduces workload and eliminates objectivity. In the first studies using computer-based methods, after obtaining the determining features such as shape, texture, and color of the image then machine learning techniques were applied [10]–[15].

Although these studies are successful up to a certain level, it can be time-consuming and inefficient considering the feature selection process. There is a huge increase in the amount of medical data especially visual data with the technological developments in hardware and software. The use of a convolutional neural network (CNN) is inevitable in this abundance of data because the success of CNN depends on seeing a lot of data. The most important benefit

of using CNN is one doesn't need to choose features for the medical image tasks. Also, it gives results in a short time with high performance [16]. There are several studies that have been using CNN for histopathological image, for example WSI classification, cancer metastasis detection, mitotic cells detection, tumor proliferation score prediction, and so on [8], [17]–[19]. A systematic review of the use of deep learning histopathological images can be found in [20], [21].

Segmentation of histopathological images has attracted considerable attention, with deep learning architectures originally developed for natural image segmentation being adapted for these applications [22], [23]. However, training these complex and deep networks is often time-consuming, highlighting the need for a network specifically adapted to biomedical images that can be trained quickly while effectively capturing the contextual information of the images.

Recent advances in biomedical image segmentation have focused on improving computational efficiency while maintaining accuracy. For instance, FU-Net employs bottleneck convolutional layers, which significantly reduce the number of parameters and accelerate the training process without compromising segmentation performance [24]. Similarly, Sharp U-Net enhances feature fusion through deep convolution, minimizing artifacts and improving segmentation quality without increasing model complexity [25]. These studies, among others, underscore the critical importance of developing lightweight and efficient architectures to address the computational challenges inherent in deep learning models for biomedical image segmentation [26]–[29].

In this study, we present a new approach that is one of the first approaches to use the convolution block used for increasing image resolution to task of histopathological image segmentation. To the best of our knowledge, it is the first segmentation method to segment tumors in histopathological images of H&E-stained breast lymph nodes using convolution blocks from increasing image resolution. The novelty of this work can be summarized as:

- A unique convolutional neural network (CNN) architecture with a multi-scale feature fusion strategy is introduced for tumor segmentation in histopathological images.
- The proposed model achieves excellent results with fewer parameters, making it computationally efficient and accurate.
- Comprehensive ablation studies provide detailed insights into the network's efficiency, distinguishing it from existing segmentation networks.

Related works

Recently, deep learning has proven its success in many areas, especially in computer vision. Networks used for vision tasks such as ResNet, DenseNet, Inception are quite deep [30]–[32]. These networks are also used as a decoder in the image segmentation process to reveal the characteristics of the image [33]–[35].

However, they take too much time when they are trained because many parameters need time to learn. We developed an architecture for segmentation that has a small number of parameters to capture and emit image content.

Medical image segmentation

With the development of a fully convolutional network (FCN), it opened a new period in semantic segmentation. The difference between CNN which has made a single prediction output for images and FCN is giving semantic segmented outputs in the desired size. This success is followed by Unet, which performs natural and biomedical image segmentation. Unet also has applied encoder-decoder architecture in the field. The encoder part obtains the features of an input image passing it to low-resolution. The decoder part uses these low-resolution features to get segmented output with high-resolution. Unet has presented the concept of skip connections. Skip connection is the improvement strategy technique applied in the decoder part due to the loss of spatial features of the in the encoder part. The features coming from the encoder section are concentrated with the features of the decoder section and the performance increases with this technique [36][37].

Recently, new segmentation architectures have been developed for biomedical images using different modules such as gated attention mechanism and long-short term memory network (LSTM) [22], [33], [38]. In our study, we analyzed its performance and obtained high performance by using a feature extraction block that was not used in segmentation architectures.

Feature extraction block

With the development of deep learning architectures, many feature extractions blocks have been developed such as the inception module, dense module, and residual module. These blocks used different kernel sizes and different convolution combinations. In all areas related to the image, such as classification, generation, detection, and semantic segmentation tasks have been used these blocks and gotten shown high performance [30]–[32].

Li et al. are developed a multi-scale residual block to obtain a high-level image from a low-level image. We have been generated a segmentation network using this block. Our segmentation network, which we will be explained in detail later, has achieved good performance [39]. A considerable body of research has leveraged the multi-scale block technique in various applications [40]–[42].

Proposed method

In this section, we are presented the segmentation network we developed. It is illustrated in Fig 1. Our network has an

encoder-decoder architecture. The encoder stream reveals contextual contexts in an image. The decoder stream connects the low-level features coming from the encoder stream with skip connections.

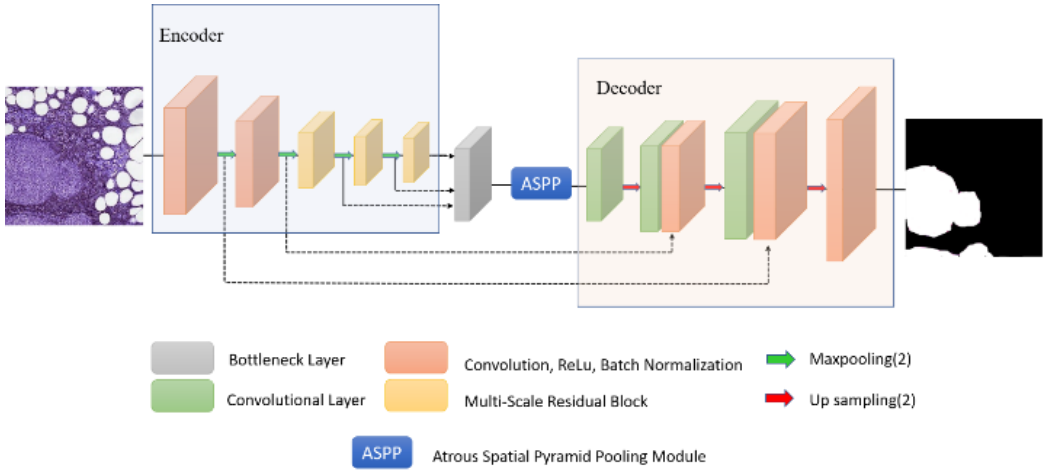

Figure 1: An illustration of the proposed segmentation network for histopathological images.

Encoder stream. Denoted as $\mathcal{E}_\theta(I)$ with parameters θ, takes input image $I \in \mathbb{R}^{CxHxW}$ with channel C (3), with height H, and weight W. This stream produces dense pixel features. The first two blogs of the encoder consist of convolution, activation function (ReLu) and batch normalization. Subsequently, the features in the image are extracted using three Multi-scale Residual Blocks (MSRB). Each scale features from MSRBs are collected in a bottleneck layer. We denote the output features from each MSRB as $e_i \in \mathbb{R}^{Cx\frac{H}{m}x\frac{W}{m}}$, where m is the stride of the encoder stream and i represents MSRB number.

Fusion stream. Each MSRB's output has specific features. The bottleneck layer is used to prevent these features from disappearing between convolutional layers. The bottleneck layer also performs dimension reduction to avoid computational cost with convolution layer (1x1 kernel). In Equation (1) is given output of the bottleneck layer, where $e_0, e_1, and\ e_2$ are the output features from MSRBs. BL represents the bottleneck layer, which concatenates and processes the input features to produce the combined feature e. Combined features pass through the Atrous Spatial Pyramid Pooling (ASPP) module to preserve multi-scale contextual information. Atrous convolution, also known as dilated convolution, differs from regular convolution by inserting zeros (or gaps) between kernel elements based on a defined dilation rate. This expands the receptive field exponentially, allowing the network to capture more extensive spatial context without increasing the kernel size or the number of learnable parameters. As a result, the ASPP module effectively extracts features at multiple scales, making it a valuable component in contemporary segmentation [43]. We denoted as ASPP(e), takes the bottleneck layer's output as an input, as an output features representation F formulate in Equation (2). ASPP(e) applies multiple dilated convolutions with different dilation rates to the input features e, expanding the receptive field without increasing the number of parameters.

$$e = BL([e_0, e_1, e_2]) \quad (1)$$

$$F = ASPP(e) \quad (2)$$

Decoder stream. Finally, low-resolution features merge with skip connections from the initial blocks of the network. Skip connections are essential for segmentation networks because they keep spatial properties that may be lost passing through many convolution layers. This stream, denoted as $\mathcal{D}_\gamma$ with parameters γ, takes input the dense feature representation F coming from the ASPP module and produce a refined segmentation output

Multi-scale residual block

Different convolution blocks have been developed to capture wide and rich contextual representations. Li et al. used a Multi-scale Residual Block (MSRB) to improve image resolution [39]. Until today, this blog has not been used for semantic segmentation in our knowledge. MSRB blog is given in Fig 2. As seen, the block consists of two parts. The first part is multi-scale features merge, the second part is the residual part.

Multi-scale features fusion: There are two bypass networks. Each network has different convolution kernels. The purpose of the bypass is to share information between networks and to detect features on different scales. The process is defined by the Equation (3), (4), (5), (6), (7) are given below,

$$S_1 = \sigma(\omega^1_{3\times3} * F_{n-1} + b^1) \quad (3)$$

$$P_1 = \sigma(\omega^1_{5\times5} * F_{n-1} + b^1) \quad (4)$$

$$S_2 = \sigma(\omega^2_{3\times3} * [S_1, P_1] + b^2) \quad (5)$$

$$P_2 = \sigma(\omega^2_{5\times5} * [P_1, S_1] + b^2) \quad (6)$$

$$S' = \omega^3_{1\times1} * [S_2, P_2] + b^3 \quad (7)$$

where b^1 , b^2 and b^3 represent bias of first, second and third layers of the block respectively. Same as bias $\omega^1, \omega^2, and\ \omega^3$represent weights of first, second and third layer of the block respectively. F denote the number of feature maps pass to the MSRB. σ represent an activation function (ReLu). The number of input features of the first layer and output layer are the same, but the number of middle layer features is twice of them. Equation (8) gives the description of MSRB.

$$F_n = S' + F_{n-1} \tag{8}$$

where F_n and F_{n-1}represent the input and output features of the MSRB. F_{n-1}helps to preserve information to pass the next block. The operation between $F_{n-1}$and S' are element-wise addition. Adding a shortcut connection to the output of block increases the performance of the network. At the end of each block, we put the pooling layer to reduce the computational cost.

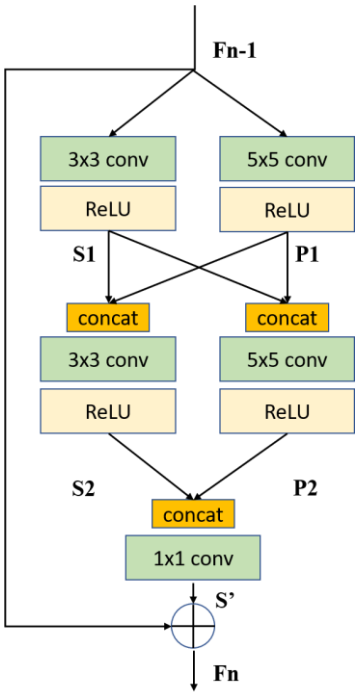

Figure 2: Multi-scale residual block (MSRB) structure.

Pyramid feature fusion module

We are used many blocks consecutively to get contextual features of image, during this transmission important features may be gradually disappearing. To prevent this loss, many methods have been developed in the literature. The effortless and useful of these methods is using skip connections.

To get outperform segmentation accuracy, we have been collected all the output features of multi-scale residual blocks. While doing this, we are brought the properties of different sizes to the first block output size by up-sampling. All these features pass through a convolutional layer with kernel size 1x1 to reduce channel size otherwise computational complexity will be increased.

The structure of the fusion module is described as follows:

$$F = w * [F_1, F_2, F_3] + b \tag{9}$$

where F_1, F_2, and F_3represents the concatenation of features from different MSRB blocks. b is the bias term. w is the weight of the final convolution layer.

Image segmentation

To obtain the output in the input size, the features maps must pass through the decoder section. The task of the decoder is to complete segmentation bypassing low-resolution feature maps into a set of convolution layers. During this task, bilinear interpolation is applied to increment the feature maps size. Bilinear interpolation helps to find out unknown pixel value by taking the weighted average of the known pixels surrounding. We use a fixed upscaling factor (x2) for the whole network.

The spatial information of contextual feature maps obtained by decreasing to low resolution is lost. As a solution, we have been applied skip connection, a technique used by many segmentation networks. In this technique, the feature maps of the encoder section received in certain scales are fused with certain scale feature maps as seen in Fig 1.

Results and discussion

To evaluate the success of the proposed segmentation network we have been used histopathological images that have tumors. First, the dataset and evaluation metrics used have been introduced. Then, the implementation of network and training details is given. Eventually, we have obtained results of the ablation studies and compared the network result with state-of-the-art segmentation networks. In addition, the quantitative results of the proposed network and segmentation networks are presented.

Dataset and evaluation metrics

In our experiments, we used Whole Slide Images (WSI) that we have obtained from the competition called Camelyon16 [44]. Camellyon16 is organized for automatic tumor detection in breast lymph nodes. The competition gives 400 WSIs. Table 2 provides detailed information about the dataset. The WSI size ranges from 1 to 4 gigabytes. They have a pyramid structure with multiple levels which are between 0 to 7. The competition has provided mask information for WSI which contains tumors.

Table 2-CAMELYON16 dataset

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

For training and testing the network, patches with 512x512 size have been prepared from level 3 of WSIs which have the tumor and normal tissue. The dataset contains 10,000

samples, with 9,000 used for training and 1,000 for testing purposes. During the training, 256x256 size patches are cropped randomly from images. Also, data augmentation techniques have been applied to the patches. In Fig 3, the images are cropped from a 512x512 image is given.

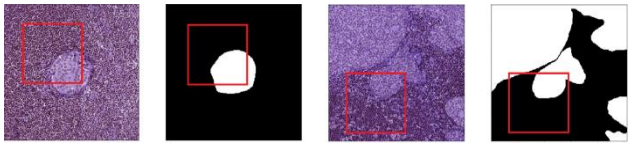

Figure 3: Histopathological images and their masks. The sizes of images are 512x512. The red square patches are randomly cropped with 256x256 size.

Evaluation metrics. In semantic segmentation, intersection over union (IoU) is the standard unit of measure. The Equation (10) is given IoU's formula for class i. Abbreviations our case tp denotes correctly classified pixels, fn denotes pixels not detected, and fp denotes background pixels classified as parts of the class. mIoU is obtained by dividing the sum of the IoU of each class by the number of classes. And Equation (11) is shown the formula of mIoU. N is the number of the classes.

$$\mathrm{IoU_i} = \frac{\mathrm{tp}}{\mathrm{tp+fp+fn}} \tag{10}$$

$$\mathrm{mIoU} = \frac{1}{\mathrm{N}}\sum \mathrm{IoU_i} \tag{11}$$

Pixel Accuracy (PA) also is used in the evaluation of the segmentation network performance. In Equation (12) is given the formula of PA;

$$PA = \frac{\mathrm{tp+tn}}{\mathrm{tp+fp+fn+tn}} \tag{12}$$

where tn denotes the background pixels which are classified correctly.

Training and Implementation Details

We applied Stochastic Gradient Descent (SGD) optimizer with the momentum of 0.9 and the weight decay of 0.0001. We have been used the learning rate of 0.0001 and batch size of 2. The total epochs number for all network training is 10. All our networks are implemented in Pytorch. Training is performed on a server computer with NVIDIA QUARD 4000 GPU under CUDA 9.0 without other ongoing programs.

We use the cross-entropy as the loss function of our network. The cross-entropy function;

$$\mathcal{L} = -\frac{1}{N}\sum_{n=0}^{N}\left(y_n \log y'_n + (1-y_n)\log\left(1-y'_n\right)\right) \tag{13}$$

where y is the ground truth, y' denotes the counterpart prediction segmentation maps, N is the total number of the pixels.

Ablation Study on the Proposed Network

Applied MSRB block and ASPP to catch content features for better histopathological image understanding. To confirm the performance of the network, we conducted experiments with different settings in Table 2. The first column of the table gives names of varieties of the network, the second, third, fourth and fifth columns give the number of kernels of their first, second, third and fourth blogs respectively. The last column shows if ASPP is used.

Table 3 gives the pixel accuracy (PA) and mIOU results of networks given in Table 2. As shown in Table 3, we are used 4 MSRBs for the first developed network, it yields a result of 62.69% mIOU and 78.00% PA. In the second experiment, we achieved a performance improvement of about 5% mIOU with 3 blocks. It is observed that adding more blocks to the network doesn't improve the performance. After the second experiment, kept the number of blocks fixed to 3 for other experiments. In the third experiment, to measure the effect of ASPP is been removed from the network which brings around %5 mIOU decline. Finally, the success of the network is tested by keeping the kernel size different for each block. This tactic has not improved the performance.

Table 2: Ablation study on the proposed network.

Network Name	1.MSRB Kernel Size	2. MSRB Kernel Size	3. MSRB Kernel Size	4. MSRB Kernel Size	ASPP
MSRSegNet_V1	5, 3	5, 3	5, 3	5, 3	Yes
MSRSegNet_V2	5, 3	5, 3	5, 3		Yes
MSRSegNet_V3	5, 3	5, 3	5, 3		No
MSRSegNet_V4	15, 7	7, 5	5, 3		Yes

Table 3: Ablation study on different proposed network on the histopathological image dataset.

Network	mIoU (%)	PA (%)
MSRSegNet_V1	62.69	78.00
MSRSegNet_V2	67.29	81.85
MSRSegNet_V3	63.14	78.47
MSRSegNet_V4	65.46	80.37

Table 4: Performance comparison between using and not using dropout.

Network	Dropout (0.05)	mIoU (%)	PA (%)
MSRSegNet_V1	Yes	**69.21**	**83.66**
MSRSegNet_V2	No	67.29	81.85

We adopted a dropout strategy to improve performance further. Dropout usually has been used by deep networks. We are gotten around 2.0% mIOU and pixel accuracy improvement with a very small probability value for dropping out nodes. The results show in Table 4.

Comparisons with State-of-the-art Methods

Table 5 shows the results from state-of-art segmentation networks. Compared to the mIOU criterion, the segmentation network we developed has great success among others. There is a difference of approximately 5.0% between the second successful network and ours. These segmentation networks have a strong backbone. Compared to the PA criteria, the Deeplabv3+ [45] segmentation network is the most successful. There is a 2.0% difference with our network.

Table 6 shows the average execution time (seconds) and parameter numbers of the networks that achieved more than 60% mIOU success. The developed network is acceptable according to the number of parameters and time criteria.

Table 5: Performance comparison between the state-ot-the-art segmentation networks. * Real-time segmentation networks

Networks	Backbone	Pretrained	mIoU(%)	PA(%)
FCN8s[37]	VGG16	Yes	15.16	30.33
UNet[36]			58.54	79.15
PSPNET [46]	ResNet101	Yes	53.11	78.05
ICNet[47]*			49.20	66.42
SegNet[48]*	VGG19	Yes	61.67	81.92
Deeplabv3+ [45]	Xception	No	61.90	**85.65**
MSRSegNet_V2			**69.21**	83.66

Table 6: Performance comparison between the state-ot-the-art segmentation networks.

Networks	Average Execution Time (second)	# Parameters (Million)
SegNet[48]	0.072	9.789
Deeplabv3+[45]	0.148	54.612
MSRSegNet_V2	0.061	2.873

In experiments, all networks are trained with a fixed number of epochs. Since DeepLabv3 + has a lot of parameters, it may need more epochs to show better performance. The small epoch number may have been affected its success. Our network, which has a small number of parameters, has been successful compared to DeepLabv3+ as it is trained faster with the same epoch number

Qualitative analysis

A sample of histopathological images, the ground truth mask of it, the mask generated by DeeplLabv3+, and the mask generated by the developed network are given in Figure 4. As can be seen, the tumors are caught in both network masks. It can be observed that there is no network result exactly like the ground truth mask.

We think this is due to tumor border lines. Once the ground truth is examined, it is observed that the edges of the border regions were rounded. While the pathologist draws the boundary of the tumor area, he can draw a little rough. Small normal tissues may remain within the tumor tissue border. Likewise, small tumor tissues may remain within the normal tissue border. Watch out that the boundary lines of the masks created by the networks are indented and protruding. We had been expecting for this situation because the real tumor boundaries are not clear like the ground truth mask. The aim of the network is to reveal the contextual dependencies of the image. Finally, the network places textures with similar content in the same classes.

Qualitative results are given in Figure 5. Histopathological image samples are given in the first column and ground truth masks are given in the second column of Figure 5. In the third and last column give predicted segmentation masks by Deeplabv3+ and our network respectively. The

masks of our network are more like the ground truth mask compared to the other network masks.

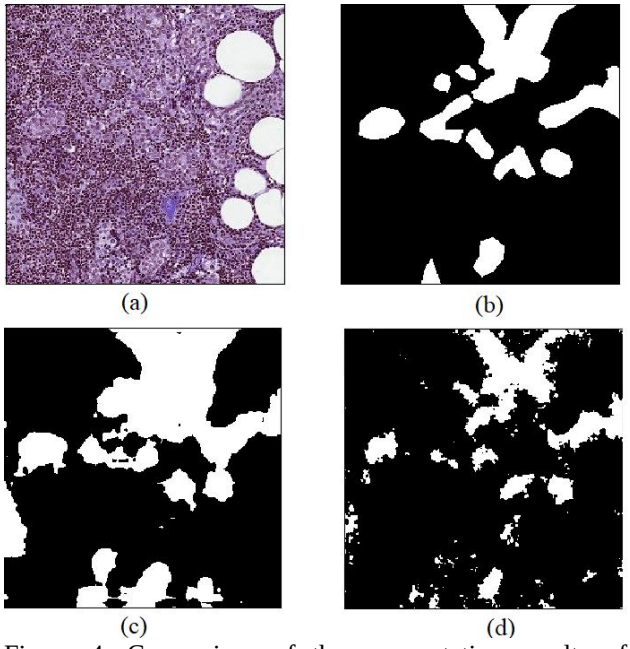

Figure 4: Comparison of the segmentation results of Deeplabv3+ and our method. (a) a sample of histopathological image, (b) the ground truth mask, (c) the prediction mask of Deeplabv3+, (d) the prediction mask of our segmentation network.

Conclusions

Early diagnosis saves lives in cancer patients. Therefore, automatic and rapid tumor segmentation is very important in histopathological images. In this paper, we proposed a new encoder-decoder CNN architecture (named as MSRSegNet) for tumor segmentation in histopathological images. Unlike the classic autoencoder approaches, the proposed architecture consists of two different properties: 1) Information sharing between encoder and decoder blocks; 2) Atrous Spatial Pyramid Pooling (ASPP) module. Information sharing (collecting contextual features) allows the coded content to be taken into consideration during the expansion phase and ensures that the local/global features in the image can be captured. ASPP module preserve multi-scale contextual information. A series of ablation experiments have been carried out in order to evaluate the performances of segmentation methods in different characteristic image regions in a healthy way. The ability to achieve higher segmentation accuracy with fewer parameters than the Deeplabv3 method, which is frequently used in medical segmentation, clearly demonstrates the superiority of the proposed method.

In future research, more diverse and comprehensive datasets, including images from a broader range of cancer types, may need to be incorporated to rigorously evaluate and improve the generalizability of MSRSegNet. Furthermore, integrating advanced techniques, such as attention mechanisms or exploring transformer-based architectures, could potentially lead to further improvements in the model's performance. Additionally, developing a lighter version of the proposed architecture, optimized for real-time segmentation, could significantly increase its applicability in clinical settings.

Ethics committee approval and conflict of interest statement

There is no need to obtain permission from the ethics committee for the article prepared.

There is no conflict of interest with any person / institution in the article prepared.

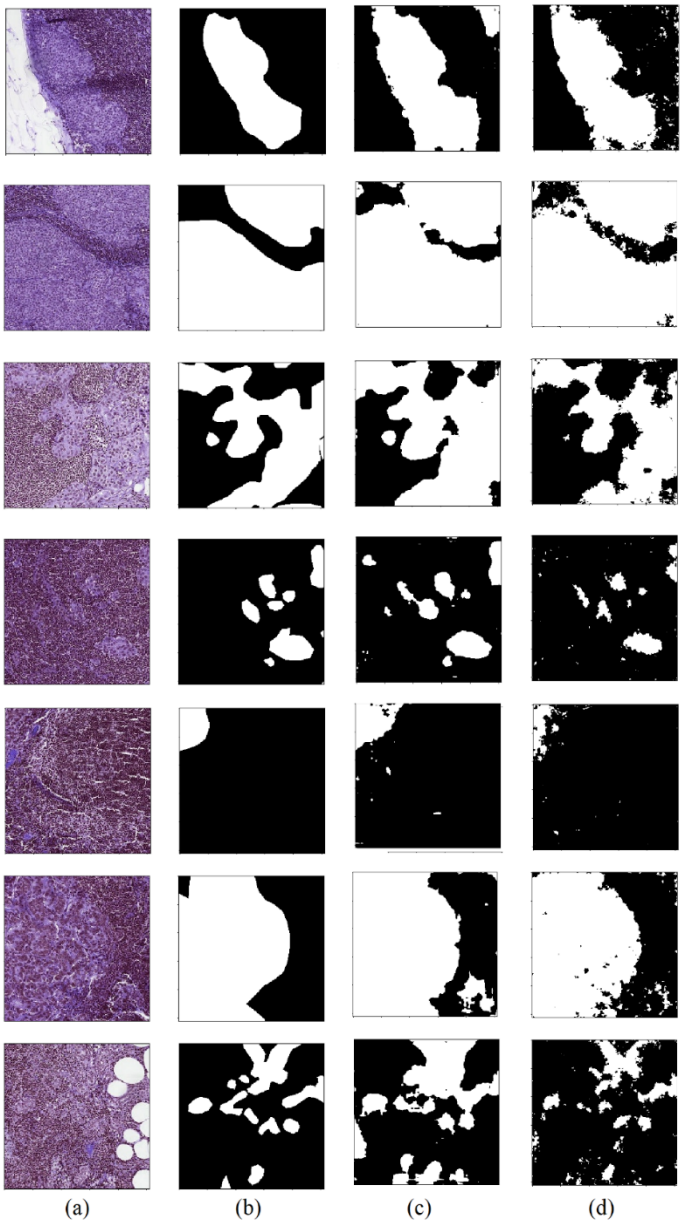

Figure 5: Comparison of the segmentation results of Deeplabv3+ and our method. (a) samples of histopathological image, (b) the ground truth masks, (c) the prediction masks of Deeplabv3+, (d) the prediction masks of our segmentation network.

References

[1] World Health Organization, "WHO | Breast cancer," *Who*, 2018. https://www.who.int/cancer/prevention/diagnosis-screening/breast-cancer/en/.

[2] M. N. Gurcan, L. E. Boucheron, A. Can, A. Madabhushi, N. M. Rajpoot, and B. Yener, "Histopathological Image Analysis: A Review," *IEEE Rev. Biomed. Eng.*, 2009, doi: 10.1109/RBME.2009.2034865.

[3] Z. Gandomkar, P. Brennan, and C. Mello-Thoms, "Computer-based image analysis in breast pathology," *J. Pathol. Inform.*, vol. 7, no. 1, p. 43, 2016, doi: 10.4103/2153-3539.192814.

[4] K. Das, S. Conjeti, A. G. Roy, J. Chatterjee, and D. Sheet, "Multıple Instance Learnıng Of Deep Convolutıonal Neural Networks For Breast Hıstopathology Whole Slıde Classıfıcatıon Kausik Das , Sailesh Conjeti , Abhijit Guha Roy Department of Electrical Engineering , IIT Kharagpur , India School of Medical Science an," no. Isbi, pp. 578–581, 2018.

[5] F. Gu, N. Burlutskiy, M. Andersson, and L. K. Wilén, "Multi-resolution Networks for Semantic Segmentation in Whole Slide Images," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11039 LNCS, pp. 11–18, 2018, doi: 10.1007/978-3-030-00949-6_2.

[6] J. W. Wei, L. J. Tafe, Y. A. Linnik, L. J. Vaickus, N. Tomita, and S. Hassanpour, "Pathologist-level classification of histologic patterns on resected lung adenocarcinoma slides with deep neural networks," *Sci. Rep.*, vol. 9, no. 1, p. 3358, 2019, doi: 10.1038/s41598-019-40041-7.

[7] Y. Celik, M. Talo, O. Yildirim, M. Karabatak, and U. R. Acharya, "Automated Invasive Ductal Carcinoma Detection Based Using Deep Transfer Learning with Whole-Slide Images," *Pattern Recognit. Lett.*, 2020, doi: 10.1016/j.patrec.2020.03.011.

[8] K. Paeng, S. Hwang, S. Park, M. Kim, and S. Kim, "A Unified Framework for Tumor Proliferation Score Prediction in Breast Histopathology," Dec. 2016.

[9] J. Wei, J. Wei, C. Jackson, B. Ren, A. Suriawinata, and S. Hassanpour, "Automated detection of celiac disease on duodenal biopsy slides: A deep learning approach," *J. Pathol. Inform.*, vol. 10, no. 1, p. 7, 2019, doi: 10.4103/jpi.jpi_87_18.

[10] B. E. Bejnordi *et al.*, "Deep Learnıng-Based Assessment Of Tumor-Associated Stroma For Dıagnosıng Breast Cancer In Hıstopathology Images Diagnostic Image Analysis Group , Radboud University Medical Center , Nijmegen , Netherlands , Beth Israel Deaconess Medical Center , Harvard M," *2017 IEEE 14th Int. Symp. Biomed. Imaging (ISBI 2017)*, pp. 929–932, 2017, doi: 10.1109/ISBI.2017.7950668.

[11] Y. Song, Q. Li, H. Huang, D. Feng, M. Chen, and W. Cai, "Low Dimensional Representation of Fisher Vectors for Microscopy Image Classification," *IEEE Trans. Med. Imaging*, vol. 36, no. 8, pp. 1636–1649, 2017, doi: 10.1109/TMI.2017.2687466.

[12] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local Neural Networks," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 7794–7803, 2018, doi: 10.1109/CVPR.2018.00813.

[13] S. Wan *et al.*, "Integrated local binary pattern texture features for classification of breast tissue imaged by optical coherence microscopy," *Med. Image Anal.*, vol. 38, pp. 104–116, 2017, doi: 10.1016/j.media.2017.03.002.

[14] P. J. Sudharshan, C. Petitjean, F. Spanhol, L. E. Oliveira, L. Heutte, and P. Honeine, "Multiple instance learning for histopathological breast cancer image classification," *Expert Syst. Appl.*, vol. 117, pp. 103–111, 2019, doi: 10.1016/j.eswa.2018.09.049.

[15] B. E. Bejnordi, G. Litjens, M. Hermsen, N. Karssemeijer, and 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," *Med. Imaging 2015 Digit. Pathol.*, vol. 9420, p. 94200H, 2015, doi: 10.1117/12.2081768.

[16] C. Higgins and C. Higgins, "Applications and challenges of digital pathology and whole slide imaging Applications and challenges of digital pathology," vol. 0295, 2015, doi: 10.3109/10520295.2015.1044566.

[17] H. Lin, S. Member, H. Chen, S. Graham, and S. Member, "Fast ScanNet : Fast and Dense Analysis of Multi-Gigapixel Whole-Slide Images for Cancer Metastasis Detection," *IEEE Trans. Med. Imaging*, vol. PP, no. c, p. 1, 2018, doi: 10.1109/TMI.2019.2891305.

[18] D. K. Das and P. K. Dutta, "Efficient automated detection of mitotic cells from breast histological images using deep convolution neutral network with wavelet decomposed patches," *Comput. Biol. Med.*, vol. 104, pp. 29–42, Jan. 2019, doi: 10.1016/j.compbiomed.2018.11.001.

[19] G. ÇELİK, "Histopatolojik Görüntülerden Kolon Kanseri Tespiti için EfficientNetB0 ve DVM Tabanlı Yaklaşım," *Fırat Üniversitesi Mühendislik Bilim. Derg.*, 2023, doi: 10.35234/fumbd.1323422.

[20] O. Jimenez-del-Toro *et al.*, "Analysis of Histopathology Images," in *Biomedical Texture Analysis*, Elsevier, 2017, pp. 281–314.

[21] J.-M. Chen *et al.*, "Computer-aided prognosis on breast cancer with hematoxylin and eosin histopathology images: A review," *Tumor Biol.*, vol. 39, no. 3, p. 101042831769455, Mar. 2017, doi: 10.1177/1010428317694550.

[22] W. T. Xiao, L. J. Chang, and W. M. Liu, "Semantic Segmentation of Colorectal Polyps with DeepLab and LSTM Networks," 2018, doi: 10.1109/ICCE-China.2018.8448568.

[23] M. Sebai, T. Wang, and S. A. Al-Fadhli, "PartMitosis: A Partially Supervised Deep Learning Framework for Mitosis Detection in Breast Cancer Histopathology Images," *IEEE Access*, vol. 8, pp. 45133–45147, 2020, doi: 10.1109/ACCESS.2020.2978754.

[24] B. Olimov, K. Sanjar, S. Din, A. Ahmad, A. Paul, and J. Kim, "FU-Net: fast biomedical image segmentation model based on bottleneck convolution layers," 2021, doi: 10.1007/s00530-020-00726-w.

[25] H. Zunair and A. Ben Hamza, "Sharp U-Net: Depthwise convolutional network for biomedical image segmentation," *Comput. Biol. Med.*, 2021, doi: 10.1016/j.compbiomed.2021.104699.

[26] Z. Bozdag and M. F. Talu, "Pyramidal position attention model for histopathological image segmentation," *Biomed. Signal Process. Control*, 2023, doi: 10.1016/j.bspc.2022.104374.

[27] Z. Bozdağ and F. M. Talu, "Pyramidal Nonlocal Network for Histopathological Image of Breast Lymph Node Segmentation," *Int. J. Comput. Intell. Syst.*, 2020, doi: 10.2991/ijcis.d.201030.001.

[28] J. M. J. Valanarasu, V. A. Sindagi, I. Hacihaliloglu, and V. M. Patel, "KiU-Net: Overcomplete Convolutional Architectures for Biomedical Image and Volumetric Segmentation," *IEEE Trans. Med. Imaging*, 2022, doi:

10.1109/TMI.2021.3130469.

[29] A. Srivastava *et al.*, "MSRF-Net: A Multi-Scale Residual Fusion Network for Biomedical Image Segmentation," *IEEE J. Biomed. Heal. Informatics*, 2022, doi: 10.1109/JBHI.2021.3138024.

[30] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," 2017, doi: 10.1109/CVPR.2017.243.

[31] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *2016 IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 770–778, Dec. 2015, doi: 10.1109/CVPR.2016.90.

[32] C. Szegedy *et al.*, "Going deeper with convolutions," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2015, pp. 1–9, doi: 10.1109/CVPR.2015.7298594.

[33] T. Takikawa, D. Acuna, V. Jampani, and S. Fidler, "Gated-SCNN: Gated Shape CNNs for Semantic Segmentation," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Oct. 2019, pp. 5228–5237, doi: 10.1109/ICCV.2019.00533.

[34] J. Fu *et al.*, "Dual Attention Network for Scene Segmentation," *2019 IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, pp. 3141–3149, Sep. 2018, doi: 10.1109/CVPR.2019.00326.

[35] Z. Zhu, M. Xu, S. Bai, T. Huang, and X. Bai, "Asymmetric Non-Local Neural Networks for Semantic Segmentation," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Oct. 2019, pp. 593–602, doi: 10.1109/ICCV.2019.00068.

[36] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *MICCAI2015*, 2015, pp. 234–241.

[37] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2015, pp. 3431–3440, doi: 10.1109/CVPR.2015.7298965.

[38] A. Pfeuffer, K. Schulz, and K. Dietmayer, "Semantic Segmentation of Video Sequences with Convolutional LSTMs," *IEEE Intell. Veh. Symp. Proc.*, May 2019, doi: 10.1109/IVS.2019.8813852.

[39] J. Li, F. Fang, K. Mei, and G. Zhang, "Multi-scale Residual Network for Image Super-Resolution," *Eur. Conf. Comput. Vis.*, pp. 527–542, 2018, doi: 10.1007/978-3-030-01237-3_32.

[40] R. Lan, L. Sun, Z. Liu, H. Lu, C. Pang, and X. Luo, "MADNet: A Fast and Lightweight Network for Single-Image Super Resolution," *IEEE Trans. Cybern.*, 2021, doi: 10.1109/TCYB.2020.2970104.

[41] Y. Chen, R. Xia, K. Yang, and K. Zou, "MFFN: image super-resolution via multi-level features fusion network," *Vis. Comput.*, 2024, doi: 10.1007/s00371-023-02795-0.

[42] C. He *et al.*, "Camouflaged Object Detection with Feature Decomposition and Edge Reconstruction," 2023, doi: 10.1109/CVPR52729.2023.02111.

[43] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation," 2018, doi: 10.1007/978-3-030-01234-2_49.

[44] B. Ehteshami Bejnordi *et al.*, "Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer," *JAMA*, vol. 318, no. 22, p. 2199, Dec. 2017, doi: 10.1001/jama.2017.14585.

[45] L. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation," pp. 833–851, Feb. 2018, doi: 10.1007/978-3-030-01234-2_49.

[46] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017, vol. 2017-Janua, pp. 6230–6239, doi: 10.1109/CVPR.2017.660.

[47] H. Zhao, X. Qi, X. Shen, J. Shi, and J. Jia, "ICNet for Real-Time Semantic Segmentation on High-Resolution Images," 2018, doi: 10.1007/978-3-030-01219-9_25.

[48] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2017, doi: 10.1109/TPAMI.2016.2644615.

.