Computers in Biology and Medicine 108 (2019) 223–233

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Binary tree-like network with two-path Fusion Attention Feature for cervical cell nucleus segmentation

Jianwei Zhang[a,*], Zhenmei Liu[a], Bohai Du[b], Junting He[a], Guanzhao Li[a], Danni Chen[a]

[a] School of Computer Science and Engineering, South China University of Technology, China
[b] Department of Pathology, The First Affiliated Hospital of Guangzhou University of Traditional Chinese Medicine, China

ARTICLE INFO

Keywords:
Cervical cell nucleus segmentation
Convolutional neural network
Binary tree structure
Dataset

ABSTRACT

Early detection of cervical lesion is of great significance in reducing mortality from cervical cancer, and segmentation of cervical cell nuclei plays an important role in screening for cervical lesion. Compared with traditional algorithms, several deep learning methods can improve the segmentation; however, due to blurred boundaries and complex gradients of medical images, the segmentation remains unsatisfactory. In addition, the existing datasets used by cervical cell nucleus segmentation research are lacking in terms of both quantity and diversity, so methods based on those datasets cannot effectively cope with challenging cases. This paper releases a new cervical cell dataset and proposes a network named Binary Tree-like Network with Two-path Fusion Attention Feature (BTTFA). The simplified diagram of BTTFA is similar to a binary tree structure and combines ResNeXt's last four layers of information by integrating adjacent pairs of layers each time; therefore, the information of multilayers can be fully integrated, and the information lost by the pooling layers can be recovered. BTTFA also applies a two-path fusion attention to selectively utilize information close to the root nodes, thereby taking full advantage of low-level detail and high-level semantic information and selectively emphasizing important features while suppressing less useful ones. Meanwhile, at some nodes of the binary tree-like network, focal loss is imposed to calculate the loss between the ground truth and the feature map during the training process. Experiments demonstrate that BTTFA performs better than the existing technology on our dataset and another public dataset.

1. Introduction

Cervical cancer is the fourth most common cancer in women [1], causing more than 25,000 deaths per year [2], so the early detection of cervical lesion is of great significance in reducing the mortality from cervical cancer [3]. For example, the morbidity and mortality of this in the United States have significantly declined by more than 70% since the 1950s. This decline has been mainly attributed to the introduction of cytology-based screening with the Pap smear test, originally developed by George N. Papanicolaou [57–60]. With the continuous development of smearing and staining technology, ThinPrep Cytology Test (TCT) published in 1996 [61] and Liquid-Based Cytology test (LBC) developed in mid-1990s [62] replaced some Pap smear techniques, and LBC has become the most widely used screening method in the field of cytology [4]. Presently, in most countries, the main smear screening technique available is manual reading, which is very cumbersome and subject to human error [5,6]. Several automated screening techniques have been developed to help pathologists analyze large amounts of data [7,8]; however, those techniques do not work well when processing images from various sources. Therefore, with the maturity of smearing and staining techniques, the automatic screening technology should also advance simultaneously, which is of great significance for screening of cervical cancer.

A key module in almost all of pathology pipelines is nucleus segmentation that extracts important features of every cell and is critical to detection of cervical cancer [9,10]. Studies such as [63] directly classify abnormal cells without nucleus segmentation; however, nucleus segmentation remains an important topic in this field because this module can provide critical features of cervical nuclei for subsequent classification and analysis. Several public datasets have been released for research of nucleus segmentation in cervical cytology images. These include the Herlev dataset and the LCH Pap smear dataset for single cell segmentation that were used by Refs. [11,12]. On the other hand, the ISBI challenge dataset and Shenzhen University (SZU) dataset for

* Corresponding author.
E-mail addresses: jwzhang@scut.edu.cn (J. Zhang), olinazhen@163.com (Z. Liu), 18922770796@189.cn (B. Du), jthescut@163.com (J. He), 2510213763@qq.com (G. Li), conniechen9469@gmail.com (D. Chen).

https://doi.org/10.1016/j.compbiomed.2019.03.011
Received 15 November 2018; Received in revised form 11 March 2019; Accepted 12 March 2019

multiple-cell segmentation were used by Refs. [13–16] to solve the problem of nucleus and overlapping cytoplasmic segmentation of cervical cells. It is relatively easy to segment the nuclei for the above existing datasets because they have the following characteristics: the nuclei of those datasets are almost separate, their shapes are mostly uniform with clear contours, and the color difference between the nucleus and the cytoplasm is clear. Based on those datasets, most of the papers [16–20] were devoted to the segmentation of overlapping cytoplasm, and a small part of the previous studies [21–23] also focused on segmentation of nuclei. However, in the real-world clinical data, overlapping or huddled nuclei often appear in the smears, and some nuclei have color similar to that of folded or overlapped cytoplasm. These features all lead to nucleus segmentation in cervical cell images becoming difficult.

Based on those existing datasets, many traditional approaches [14,19,20,24–26] to the cell nucleus segmentation have been proposed. For example, Yang et al. [14] proposed an algorithm based on the watershed, which grew the marker based on a flooding simulation process and would easily cause over- or under segmentation. When there were multiple cells in an image, it was difficult to find reliable and robust markers. Most of the nondeep learning methods do not work well in intractable cases, and much room for improvement remains.

In recent years, increasingly more networks such as U-Net [27–29] have been used for biological image segmentation. The difference between a deep neural network and nondeep learning algorithms is that if the dataset is large enough, the neural network can simultaneously handle various complicated nucleus segmentation cases. However, presently, on the one hand, the datasets in the literature have less diversity of nuclear appearance than that in a clinical environment; on the other hand, the strategy of fusing features in existing networks is not quite suitable for cervical cell nucleus segmentation.

For this paper, a group of cervical cell images based on Liquid-Based Cytology test was obtained from a hospital and a biomedical device company. This dataset contains 104 images of size 1024×768, and each image has a segmentation ground truth that has been labeled by professional pathologists. Additionally, a deep-learning framework named BTTFA is proposed that can process cervical cell images of challenging cases. BTTFA selectively integrates multilevel features to enhance their coordination. The Binary Tree-like Network proposed in BTTFA strategically incorporates features to compensate for the information loss caused by the pooling layers, and the features that are extracted by the basic network are integrated into *Final Feature map* to correctly classify each pixel into background and nucleus pixels. BTTFA also includes Two-path Fusion Attention, generating weights to selectively use low- and high-layer features to improve the role of important information and reduce the interference of unimportant information. Experiments on the released dataset demonstrate that the BTTFA proposed in this paper performs well. Furthermore, experiments on another existing public medical dataset are performed, and achieve the desired performance.

This paper proposes a new method for feature combination that flexibly integrates the multilevel features from the used basic network to cope effectively with challenging cases in nucleus segmentation. The difficult problems that we are trying to solve are clustered nuclei, granular cluster nuclei, similar colors of nucleus and cytoplasm, diverse shapes and sizes of nuclei, etc.

The manuscript is organized as follows. Section 2 provides a brief survey of the current segmentation and evaluation methods. Section 3 describes the used datasets. Section 4 describes the proposed method and the loss function. The experimental environments, parameter settings, experimental results and discussion are presented in Section 5. Finally, the conclusions are described in Section 6.

2. Related work

Nucleus segmentation is a problem that the academic community has been working on because the results of nucleus segmentation contribute to judging pathologies and medical diagnostics. In this section, we will review the nondeep and deep learning methods for nucleus segmentation.

2.1. Non-deep learning methods for cytological nucleus segmentation

Thus far, many cervical cytology nucleus segmentation methods have been proposed, most of which are based on traditional algorithms, such as edge enhancement [30], watershed [14], clustering [31,32], and thresholding [34,35]. For instance, Lin et al. [30] used a segmentation method based on a series of edge enhancement techniques, which did not perform well in the case of blurred contours of nuclei. The multi-pass fast watershed-based method (MPFW) [18] can deal with the task of segmenting separate cell nuclei, and was also used as an approach to segmentation of overlapping cervical cytoplasm. Bergmeir et al. [35] proposed an algorithm to localize the nuclei using a voting scheme and prior knowledge; the nuclear contours were extracted with a canny edge detection algorithm. Many cases, such as cervical cell images of nuclei with diverse shapes and sizes, were still not handled well by the abovementioned methods.

2.2. Deep learning methods for cytological nucleus segmentation

For the problem of challenging cytology nucleus segmentation, the performance of deep convolutional networks has improved with the prevalence of machine learning, and in most cases deep convolutional networks perform better than nondeep learning methods [18,33,36]. As a typical example, U-Net, a convolutional network proposed in 2015 for biomedical image segmentation, has an architecture that consists of a U-shaped network used to fuse the features extracted by previous convolutional layers, which is a symmetric expanding path that enables precise localization of nuclei and can also be applied to cervical cell nuclear segmentation [28]. Song [18] proposed a novel contour-seed pairs learning-based framework for robust and automated nucleus segmentation. First, a set of online convolutional features is learned in each layer; next, with the input from the extracted features, the method iteratively updates the locations and clustered object boundaries until convergence has been achieved. Although the method had a certain generalization ability, which could perform well in the case of touching nucleus segmentation, it was not an end-to-end program and was difficult to apply. Song [37] proposed a deep learning method that was used for detection of the region of interest for cervical cancer cell segmentation, and then a coarse-to-fine nucleus segmentation was also developed. The deep learning method of FCN is an important part of the method proposed by Zhang [38]. FCN is trained to learn the nucleus' high-level feature, and then generate the final nuclear map by combining it with the graph cost function. However, the method [38] has only been applied to cell nuclei from the Herlev Pap smear dataset. Most of the methods described above, such as [18,37,38], were not end-to-end methods and had high computational complexity. Although deep learning methods have helped make significant progress in the field of cytology nucleus segmentation, the need to develop more practical and effective methods remains.

2.3. Evaluation of segmentation methods

There are many widely accepted evaluation metrics for nucleus segmentation techniques. Such techniques are mainly evaluated according to two aspects, including object-level and pixel-level. The object-level method means that the nucleus is the basic unit used in calculating the recall rate, accuracy, etc. to represent the missed detection of ground truth objects and false detection of ghost objects. The pixel-level method calculates the overall accuracy and recall rate, etc. in pixels. Object-level evaluation and pixel-level evaluation are complementary. Six evaluation metrics are considered in this paper.

2.3.1. Object-level evaluation

F1-score is a commonly used object detection metric. For every ground truth object G_n and segmented object S_m, True-Positive (*TP*) is the intersection of the ground truth object set $\{G_n\}$ and the segmented object set $\{S_m\}$. The segmented objects are counted in *TP* only if the Dice coefficients of them and their corresponding ground truth objects are greater than a certain threshold T. False-Positive (*FP*) refers to all segmented objects without corresponding ground truth objects. False Negative (*FN*) contains all ground truth objects without corresponding detected objects. F1-score is a measure that comprehensively considers the precision and the recall rate, as defined by Formula (1).

$$\mathrm{F1-score} = \frac{2TP}{2TP + FP + FN} \tag{1}$$

Object-level detection accuracy called $prec_O$ is the ratio of the number of true-positive nuclei to the total number of all segmented nuclei in the test set, as defined by Formula (2).

$$prec_o = \frac{TP}{TP + FP} \tag{2}$$

Object-level detection recall, named, rec_O is the ratio of the number of true-positive nuclei to the total number of all nuclei in the ground truth set, as defined by Formula (3).

$$rec_o = \frac{TP}{TP + FN} \tag{3}$$

2.3.2. Pixel-level evaluation

Object-level evaluation is insufficient for an accurate assessment of the segmentation results, so three pixelwise metrics are used in this paper. Pixelwise Dice coefficient [10], a similarity measure function that is widely used to calculate the similarity of two sets, is the most common pixel-level evaluation. In this paper, the mean Dice coefficient is used, defined as the average value of Dice coefficients of all nuclei. First, the Dice coefficients of all pairs of nuclei G_n and S_m .are calculated, in which G_n is a ground truth nucleus, and S_m is the corresponding nucleus. The two sets are all pixels in G_n and all pixels in S_m , respectively. Only if the Dice coefficient (DC) of S_m and G_n. exceeds a certain threshold *T*, S_m. can be counted as a true positive that affects the Dice coefficient [14]. Then, the mean Dice coefficient is calculated. The Dice coefficient is defined by Formula (4) (where |. | denotes the number of pixels in the object):

$$\mathrm{DC} = \frac{2|G_n \cap S_m|}{|G_n \cap S_m| + |G_n \cup S_m|} \tag{4}$$

Pixel-level accuracy called $prec_p$ is ratio of the number of intersecting pixels in all ground truth nuclei G_n. and all their corresponding segmented nuclei S_m. to the number of all pixels in all segmented nuclei S_m. Metric $prec_p$ is defined by Formula (5).

$$prec_p = \frac{|U_n G_n \cap U_m S_m|}{|U_m S_m|} \tag{5}$$

Pixel-level recall, called rec_p, refers to the ratio of the number of intersecting pixels of all ground truth nuclei and their corresponding segmented nuclei to the number of pixels in all the ground truth nuclei and is defined by Formula (6).

$$rec_p = \frac{|U_n G_n \cap U_m S_m|}{|U_n G_n|} \tag{6}$$

The mentioned pixel-level and object-level evaluation metrics adopted in this paper can comprehensively assess the performance of the proposed nucleus segmentation method.

3. Labeled datasets

Several public cervical cytology datasets exist, such as the Herlev dataset, the LCH Pap smear dataset, the ISBI challenge datasets, etc. In most of existing public datasets, there is only a single cervical cell in each image, which makes it unsuitable for implementing the entire process flow for the real-world data in a clinic. This paper releases a dataset for cervical nucleus segmentation, containing complicated cases from clinical practice. To prove the efficacy of the proposed algorithm, a common cervical cell dataset is chosen for comparison. The two datasets used in this paper are described below.

3.1. Released cervical cell dataset based on Liquid-Based Cytology test

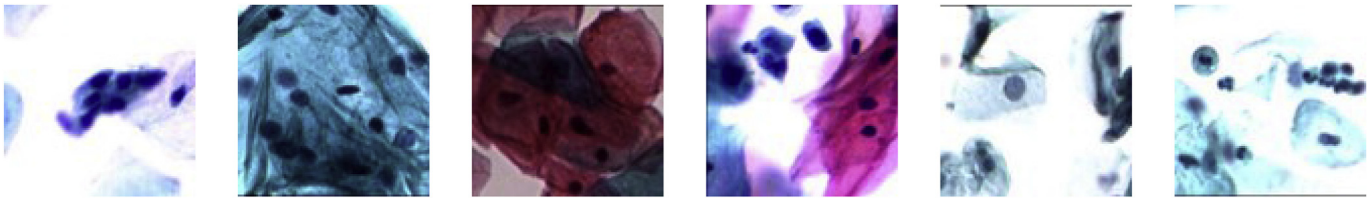
In deep learning algorithms, the size of dataset is critical to experimental results, and labeled datasets are necessary for supervised learning methods. It is time-consuming to build a tagged dataset of cervical cell images; additionally, this task requires specialized medical knowledge. This leads to the lack of datasets. This paper has released a real-world clinical dataset with well-annotated nuclei and diverse cervical cell images with the primary aim of promoting the development of cervical cytology imagery.

The images are based on the Liquid-Based Cytology test from the pathology departments of a Chinese hospital and a biomedical device company. Under the guidance of a professional pathologist, we chose qualified images among numerous LCT cervical images. The dataset contains 104 images of size 1024 × 768., and each image has a ground truth that has been manually segmented by a professional pathologist. All original images were created using the Olympus microscope B x 51 with 200x magnification, and their effective pixel size is 0.32 μm × 0.32 μm. The used PlanN 20x/0.4 objective produces a 20x magnification with a numerical aperture of 0.4, and the illumination is white.

In clinical data, several cases often appear that are difficult to handle, including intercellular overlapping and self-folding, clustered nuclei, diverse shapes and sizes of nuclei, nuclei with blurred contours, similar colors of nucleus and cytoplasm, and granular clustered nuclei, as shown in Fig. 1. At the same time, some images contain a few impurities. To meet the practical need, the normal-looking images and the intractable images shown above must be all accurately segmented.

3.2. ISBI challenge dataset

The dataset provided by the ISBI 2014 challenge organizers in their baseline method [40] includes 45 training images and two sets of test



(a) (b) (c) (d) (e) (f)

Fig. 1. Several cases that are difficult to handle in released dataset. (a) clustered nuclei; (b) nuclei with blurred contours; (c) intercellular overlap and self-folding; (d) diverse shapes and sizes of nuclei; (e) similar color of nucleus and cytoplasm; (f) granular cluster nuclei.

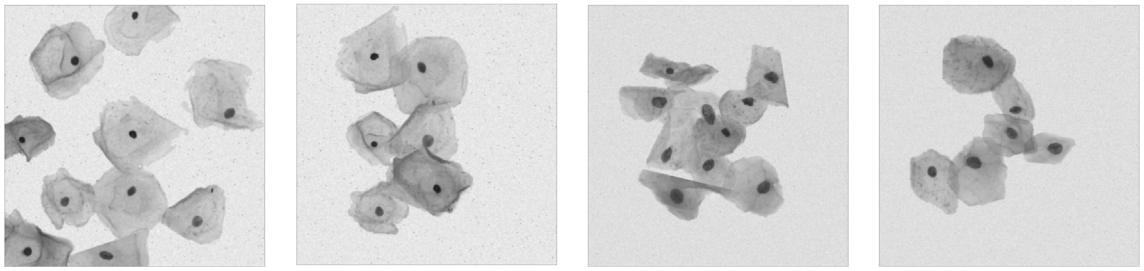


Fig. 2. Samples of ISBI challenge dataset.

images: the ISBI-14 test-90 and test-810 synthetic datasets. All those images were 512 × 512 grayscale images with annotated nuclei. In this paper, this dataset consisting of 45 training images and 900 testing images is adopted. All those synthetic images were generated by overlapping a collection of isolated cells extracted from real Pap smear slides, reconstructed by applying a random linear brightness transformation and a random rigid transformation, and finally located on the synthetic image [41]. Several samples of ISBI challenge dataset are shown in Fig. 2.

4. Method

The proposed network applies ResNeXt [42] as the basis network for the multilayer feature extraction to produce a set of feature maps of the same scale. Additionally, Binary Tree-like Network Topology and Two-path Fusion Attention are presented. The overall structure of the network is shown in Fig. 3. For the deep neural network, the deeper layers are more semantic but spatially coarser [43]. The shallower layers express more detail, but the convolution kernel has a small receptive field, so the semantic information is not described well. To better perform the roles of both high-level semantic features and low-level detail features, features in different layers are integrated into nodes by the method of Binary Tree-like Network proposed in this paper. The basic network generates five layers of features, called *layer0*, *layer1*, *layer2*, *layer3*, and *layer4* from shallow to deep, respectively, shown in Fig. 3. *Layer3* and *layer4* can be considered high-level semantic features compared to other layers, while *layer2* and *layer1* can be regarded as low-level detail features. *Layer4* is convolved to generate a feature map *F1* that then is concatenated with the upsampled *layer3*; finally, convolution processing is used to generate the feature map named *Feature-High* that can be regarded as the integration of high-level semantic information. The integration operations of *layer2* and *layer1* are same as those of *layer4* and *layer3*, and result in *Feature-Low* being obtained. To better use high-level semantic information and detail features, Two-path Fusion Attention is proposed to fuse *Feature-High* and *Feature-Low* and generate the *Fusion Attention feature*. To obtain a sufficiently low-level feature before generating the *Final Feature map*, the *Fusion attention feature* is concatenated with the *layer0* feature by the method proposed by this paper, called the root refinement. Additionally, for feature maps *F1* and *F2*, *Feature-Low*, *Feature-High*, *Fusion Attention Feature* and the *Final Feature map*, Focal Loss [44] is imposed to calculate the loss during the training process. Finally, the *Final Feature map* in Fig. 3 at the root node of the Binary Tree-like Network is used as the final output of the proposed network. The following subsections will elaborate on building the Binary Tree-like Network and the Two-path Fusion Attention.

4.1. Binary tree-like network topology

Since the features extracted by the basic network are sufficient, including semantic and detailed information, it is common to explore the fusion methods for multilayer features to improve the segmentation performance, such as DSS [45] and R3 [46]. Based on the fact that the segmented nuclei in the cell image are small, the nucleus segmentation methods have more relevance to low-level detail features. To adapt to the computer storage capacity, most of the current networks often have pooling layers, such as VGG [52], but when the network comes to the deeper layers, many details are lost due to the pooling operation. To fully integrate the low-level and high-level information of the neural network, the U-Net [28] combines the information of layers in symmetric positions. In contrast, FCN-like structures [27] gradually fuse the features from upper to lower ones. In general, to fully use semantic information and low-level detail features, the general idea involves skip connections from shallower to deeper stages for merging the feature at different scales and resolutions. However, the skips are linear in existing approaches, e.g., FCN, U-Net, and FPN [39]. Furthermore, when the network becomes very deep, it is not enough to use the feature of the last layer of the network alone. The middle level of the network has a significant amount of detail information that is very helpful. In 2018, Yu [43] introduced a deep layer aggregation structure to extract the full spectrum of semantic and spatial information from a network. The goal of deep layer aggregation is to aggregate blocks of the basic network (Resnet, VGG) and add more flexibility to the network. The method works well in the visual recognition task if the dataset is sufficiently large, but involves a large computational cost and is unsuitable for tasks

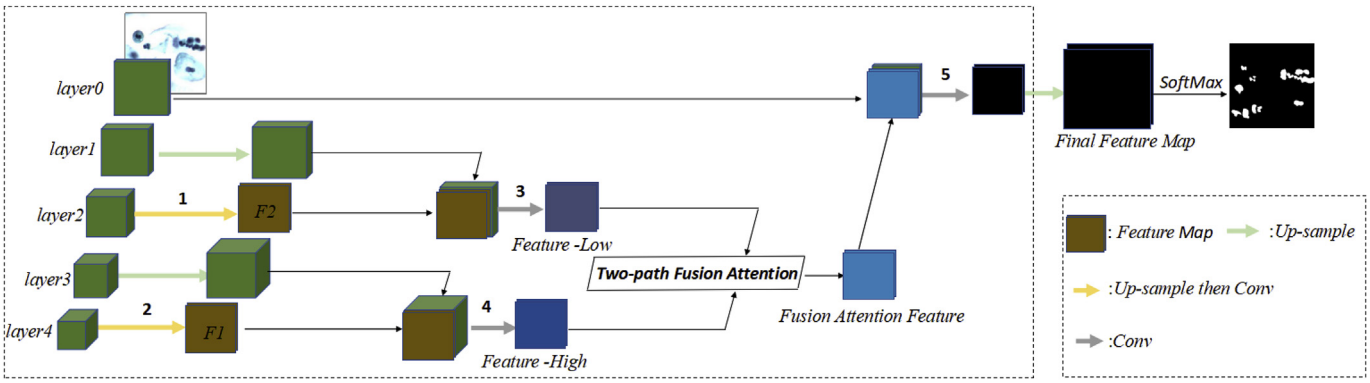


Fig. 3. Binary-tree-like network.

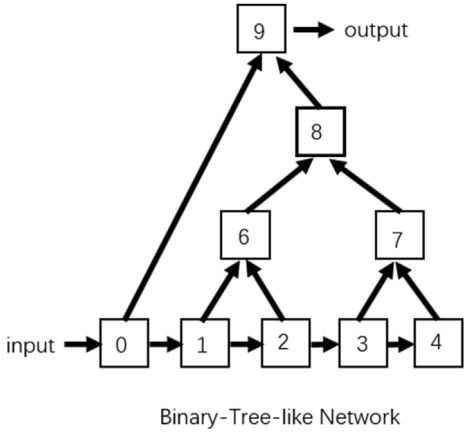


Fig. 4. Binary-tree-like network topology (BTN).

with small datasets, such as nucleus segmentation.

To accommodate small datasets and obtain more discriminating features, the Binary Tree-like Network is proposed, which aggregates the features captured by the basic network by means of a hierarchical structure, and has a lower computational cost. The Binary Tree-like Network compensates for the information loss caused by the shallow pooling layers by integrating the information obtained from the various pooling layers. Additionally, at several nodes, a series of losses are calculated between feature maps and the ground truth to preserve more details of cervical nuclei. The last four layers of the basic network are divided into two sets of high-level semantic information and low-level detailed features, and the two features in each set are combined; next, nodes 6 and 7 are obtained, which are *Feature-Low* and *Feature-High*, respectively, as shown as Fig. 4. Afterwards, nodes 6 and 7 are re-combined into the *Fusion Attention Feature* by the method of Two-path Fusion Attention. Finally, before the *Final Feature map* is generated, BTN establishes a skip connection to *layer0* with the *Fusion Attention Feature*, named Node 9 in Fig. 4, to emphasize the role of the low-level features in nucleus segmentation.

For the segmentation task, especially segmentation of small objects such as nuclei, the low-level details are crucial. If the extracted lowest feature layer0 is ignored, the segmentation result will be affected. This effect is shown in the experiment of Component and Structure Analysis in this paper. The experiment proves that layer0 has helped retain more details and improve the segmentation result. In addition to the network topology shown in Fig. 4, the other three structures in Fig. 5 are designed for a comparison study. In the network structure in Fig. 5(1), layer0 is combined with node 1, while in Fig. 5(2) it is combined with node 6, and in Fig. 5(3), the convolution is applied to layer0, and then the result is combined with node 8. A comparison experiment is performed, and the result in Table 4 shows that the network topology in Fig. 4 is better than the three other structures shown in Fig. 5. Therefore, layer0 is combined as described in Fig. 4 in the proposed BTTFA method.

The cervical cell image datasets in this paper are small, so the features captured by ResNeXT [42] are divided into only five levels. If the dataset is large, the Binary Tree-like Network can become bigger, and the basic network features can be divided into more levels; then, the features captured by the basic network can be combined in the same ways as in the presented aggregation method; finally, the high-level features can be combined by the method of Two-path Fusion Attention.

4.2. Two-path fusion attention

Hu [48] has proposed an attention mechanism to assign a higher weight to important information, that can flexibly capture global and local connections. Attention can be viewed as a tool to bias the allocation of available processing resources towards the most informative components of an input signal; such a mechanism has been described across a range of tasks, including those in the image domain [49,50]. It is often used on top of one or more layers representing higher-level semantic information for adaptation between modalities [48]. Inspired by Hu, we propose Two-path Fusion Attention (TFA) in this paper.

TFA has two inputs (*Feature-Low* and *Feature-High*) and one output (*Fusion Attention Feature*). First, the two inputs are concatenated and subsequently convolved to generate common fusion features. Next, the fusion features are concatenated with the two respective inputs and convolved to generate two weight maps. Each generated weight map is related to the fused feature of Feature-High and Feature-Low; hence, this method is called Two-path Fusion Attention. Finally, the two inputs are multiplied by the two weights and fused to form the Fusion Attention Feature. Compared to the proposed TFA, the two more common integration patterns are shown in Fig. 6(a) and (b). In Fig. 6(a), the Feature-Low and Feature-High are concatenated directly, while Fig. 6(b) shows the standard Two-path Attention without fusion between Feature-Low and Feature-High. In TFA proposed in this paper, Feature-Low and Feature-High are interwoven and adequately fused. During network training, TFA can integrate *Feature-High* and *Feature-Low* better, as will be illustrated in the experiments in this paper.

4.3. Loss function

As shown in Fig. 3, the six feature maps of the proposed network obtained in all steps are *F1*, *F2*, *Feature-High*, *Feature-Low*, *Fusion Attention Feature* and the *Final Feature Map*. During the training process, Focal Loss [44] is applied to calculate loss for these feature maps as defined by Formula (7).

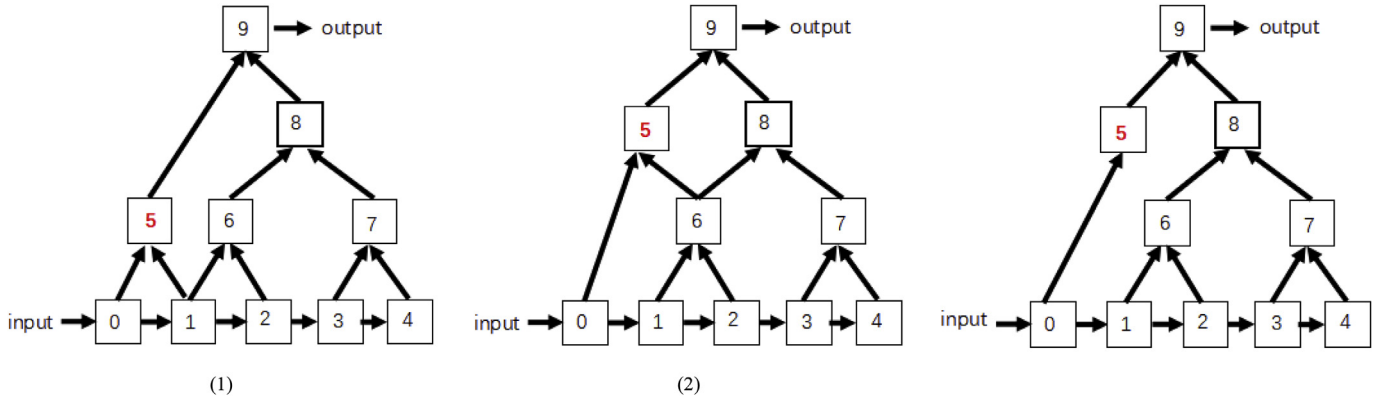


Fig. 5. Other three ways of combining with layer0. (1) layer0 combines with node 1; (2) layer0 combines with node 6; (3) layer0 is processed by convolution, then combined with node 8.

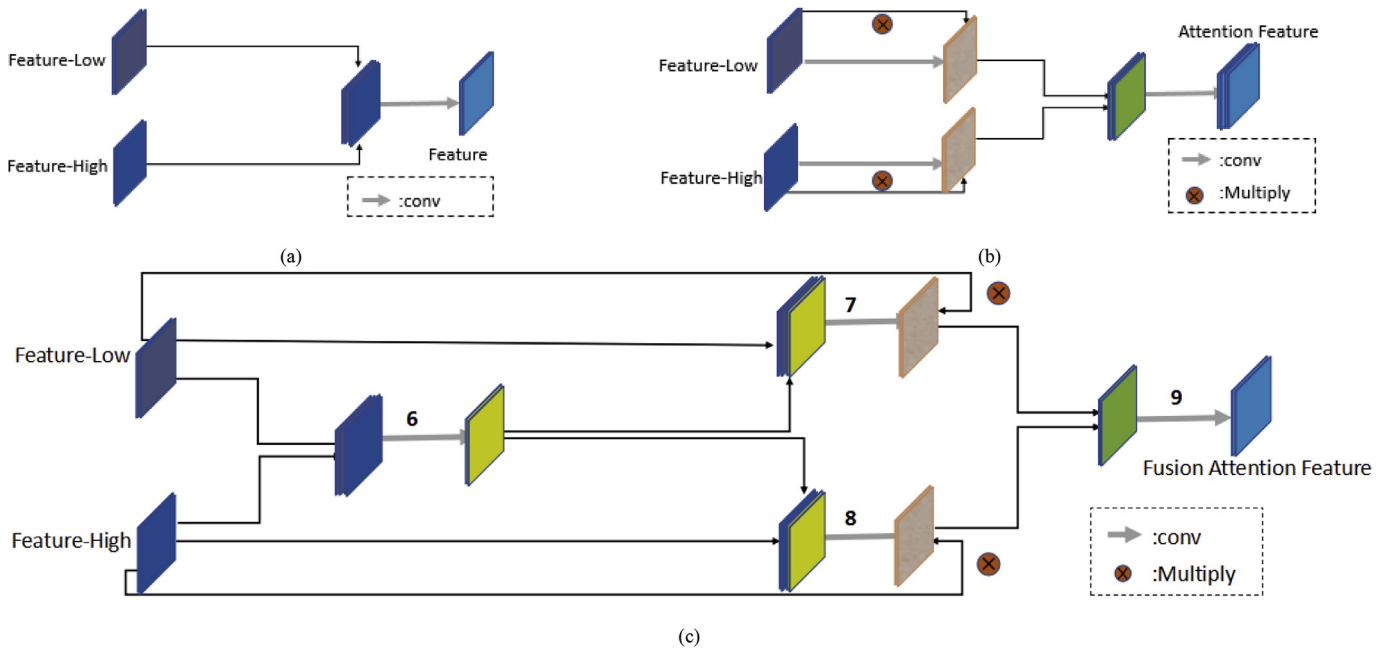


Fig. 6. Three kinds of integration methods: (a) merging features with a normal connection; (b) Two-path Attention without fusion; (c) Two-path Fusion Attention (TFA).

$$FL(p_t) = -(1 - p_t)^y \log(p_t) \tag{7}$$

The tunable focusing parameter is $y \geq 0$; and in our experiment, y is set to 1, and p_t is defined by Formula (8), where p ∈[0,1]:

$$p = \begin{cases} p & if \quad y = 1 \\ 1 - p & otherwise \end{cases} \tag{8}$$

The total loss of the proposed network is defined as the sum of the loss on all the predicted feature maps, as shown in Formula (9). Although the loss is related to those 6 feature maps during the training, the loss of the test process is only related to the last loss of the *Final Feature Map*.

$$Loss_{all} = \sum_{i=1}^{6} w_i y_i \tag{9}$$

where w_i .and y_i .are the weight and loss in the six feature predictions. In our experiment, we empirically set all the weights w_i. to 1, and use the total loss in Formula (9) to train the entire network simultaneously. The Focal Loss has been proposed by Kaiming He [54], and is modified based on the cross-entropy loss. It can make the presented model to more carefully learn the difficult-to-classify samples by reducing the loss weight of the easily categorized samples. This paper applies the Focal Loss due to its somewhat better performance than that of cross-entropy loss. Experimental results are shown in Table 5.

5. Experimental results

In this section, we mainly consider three aspects to evaluate the performance of the presented BTTFA network. The metrics used to quantitatively evaluate segmentation results include *F1-score*, object detection accuracy ($prec_O$.), object detection recall rate (rec_O.), Dice coefficient (Dice),.pixel-level accuracy ($prec_p$), and pixel-level recall rate (rec_p). For all of these metrics, higher values indicate better results. More detailed information about the evaluation metrics is provided in Section 2.3.

5.1. Experimental environments and parameter settings

To speed up the training process, the parameters of the basic network are initialized using the well-trained ResNeXt network on ImageNet [42] (see Fig. 3), and the parameters of the other layers use the default initialization method of PyTorch. Adaptive moment estimation (ADAM) is used to train the network with the learning rate of 0.001, and the batch size of 10; the training step stopped after 10,000 iterations. The images of the size 1024 × 768 in the released dataset are randomly rotated, horizontally flipped, and repeatedly cropped to size 128 × 128. for data augmentation during the training process. Once the training step has completed, cervical cell images of various size can be segmented by the network, and in this paper, each test image is averagely divided into 12 images of size 256 × 256 pixels considering the experiment efficiency. The input to the network is three channels of color images without preprocessing. In the training process, the size of the images in the public dataset is 512 × 512 pixels, and the process in the public dataset is almost the same as that of the images in the released dataset. The only difference is that the grayscale images in the public dataset have only one channels. In this paper, when training on the public dataset, one channel of the grayscale image is copied to generate the second and third channels. The proposed framework is trained on a Nvidia GeForce GTX 1080 GPU with 12 GB of RAM. The training of BTTFA takes approximately 1 min and 40 s for each epoch, and its test speed is 7 images of size 256 × 256 per second.

In Figs. 3 and 6, layer0 to layer4 are directly obtained from the first four layers of ResNeXt [42], and the operation of up-sample is bilinear interpolation without parameters. Table 1 shows the details of the convolution operation in BTTFA. Numbers 1, 2, 3, etc. in the first column in Table 1 correspond to the convolution numbers in Fig. 3.

5.2. Experimental results on the dataset released in this paper

Three comparisons are performed between the proposed method and the segmentation results of the fully convolutional network (FCN) [47], U-Net [28] and DeepLabv3+ [51] using the released dataset. The best segmentation results of other three methods are obtained by retraining their models using the public implementations. The best segmentation results of other three methods are obtained by retraining their models using the public implementations and adjusting training parameters. To test the accuracy of the algorithm, the fivefold cross-validation method is utilized to obtain the experimental results, as

Table 1
The convolution operation details of BTTFA and the first column correspond to the convolution numbers in Fig. 3.

Convolution numbers	Parameter	# Parameters	Input dimensionality	Output dimensionality
1	3 × 3 × 512 × 128 3 × 3 × 128 × 128 1 × 1 × 128 × 2	737536	64 × 64 × 512	64 × 64 × 2
2	3 × 3 × 2048 × 128 3 × 3 × 128 × 128 1 × 1 × 128 × 2	2507008	16 × 16 × 2048	16 × 16 × 2
3	3 × 3 × 258 × 128 3 × 3 × 128 × 128 1 × 1 × 128 × 2	444928	64 × 64 × 258	64 × 64 × 2
4	3 × 3 × 2026 × 128 3 × 3 × 128 × 128 1 × 1 × 128 × 2	2493184	16 × 16 × 2026	16 × 16 × 2
5	3 × 3 × 66 × 128 3 × 3 × 128 × 128 1 × 1 × 128 × 2	223744	256 × 256 × 66	256 × 256 × 2
6	3 × 3 × 4 × 128 3 × 3 × 128 × 128 1 × 1 × 128 × 2	152320	256 × 256 × 4	256 × 256 × 2
7	3 × 3 × 4 × 64 3 × 3 × 64 × 64 1 × 1 × 64 × 2	39296	256 × 256 × 4	256 × 256 × 2
8	3 × 3 × 4 × 64 3 × 3 × 64 × 64 1 × 1 × 64 × 2	39296	256 × 256 × 4	256 × 256 × 2
9	3 × 3 × 4 × 128 3 × 3 × 128 × 128 1 × 1 × 128 × 2	152320	256 × 256 × 4	256 × 256 × 2

Table 2
Comparison with other segmentation methods on the dataset released in this paper.

Method	$prec_o$ (MEAN)	rec_o (MEAN)	$prec_p$ (MEAN)	rec_p (MEAN)	Mean DC (MEAN)	F1-Score (MEAN)
[28] U-Net	0.73	0.84	0.79	0.75	0.84	0.78
[47] FCN	0.87	0.86	0.77	0.83	0.82	0.87
[51] DeepLabv3+	0.77	0.90	0.87	0.82	0.89	0.83
Proposed	**0.91**	**0.94**	**0.90**	**0.89**	**0.91**	**0.93**

shown in Table 2. In the experiment, the segmented objects are counted as True-Positive only if the Dice coefficients of them and their corresponding ground truth objects are greater than 0.5. The experiments have been performed with other thresholds such as 0.7 and 0.6, and the results demonstrate that the proposed method is always the best on all evaluation metrics. Table 2 shows the results for the Dice threshold of 0.5.

As Table 2 shows, although U-net [28] with a U-shaped structure that can effectively combine shallow and deep features is a classic network for biological image segmentation, its performance is slightly worse than that of other methods among the four. FCN [47] is a fully convolutional network for semantic segmentation and is adapted from a contemporary classification network (namely, VGG [52]). Compared with U-Net, the basic framework of FCN is VGG, which can extract the detailed features well. DeepLabv3+ performs slightly better than FCN overall, and the former also performs better than the latter on the PASCAL VOC 2012 semantic segmentation benchmark [53]. DeepLabv3+ is by far the most advanced semantic segmentation network that applied an optimized version of Resnet [54] to extract features. Compared with BTTFA, DeepLabv3+ is more suitable for segmenting larger objects; as the nucleus is too small, this method does not obtain the best performance. It can be observed that our method consistently outperforms others on all the metrics, indicating the improvement of BTTFA.

Fig. 7 visualizes the segmentation results using three colors. White indicates the intersection regions of a ground truth nucleus and its corresponding segmented nucleus (TP); red refers to the intersection regions of a ground truth nucleus without the corresponding segmented nucleus (FN), and blue corresponds to the intersection regions of a segmented nucleus without the corresponding ground truth nucleus (FP). Several cases include clustered nuclei, intercellular overlapping and self-folding, diverse shapes and sizes of nuclei, and blurred contours of nuclei. At the same time, some images contain a few impurities.

According to the first, fourth, fifth, and seventh columns, in the case of cervical cell images with blurred contours of nuclei, DeepLabv3+ is observed to be superior to all the other compared methods; it is the most advanced method of semantic segmentation, and has been improved based on the previous paper in the series of [13,55,56], which proposed ASPP to further process high-level semantic information. However, compared with the proposed method, DeepLabv3+ is prone to misclassifying impurities nuclei, yet slightly outperforms FCN overall. At the same time, according to the second and third columns, although U-Net has identified most of the nuclei, it could not produce their intact segmentation. Overall, BTTFA outperforms the other methods in distinguishing among the nucleus, impurities and the background; however, some segmentation results are not ideal. For example, in the fourth column's legend, if the cytoplasm and nucleus are close in color, BTTFA also cannot segment the nucleus well. Although the method proposed in this paper may occasionally segment a few impurities as nuclei, the impurities might be distinguished by future work, classification. However, in most cases BTTFA's segmentation is closest to the ground truth.

5.3. Experimental results on the ISBI challenge dataset

This paper performs comparisons of the segmentation results of the

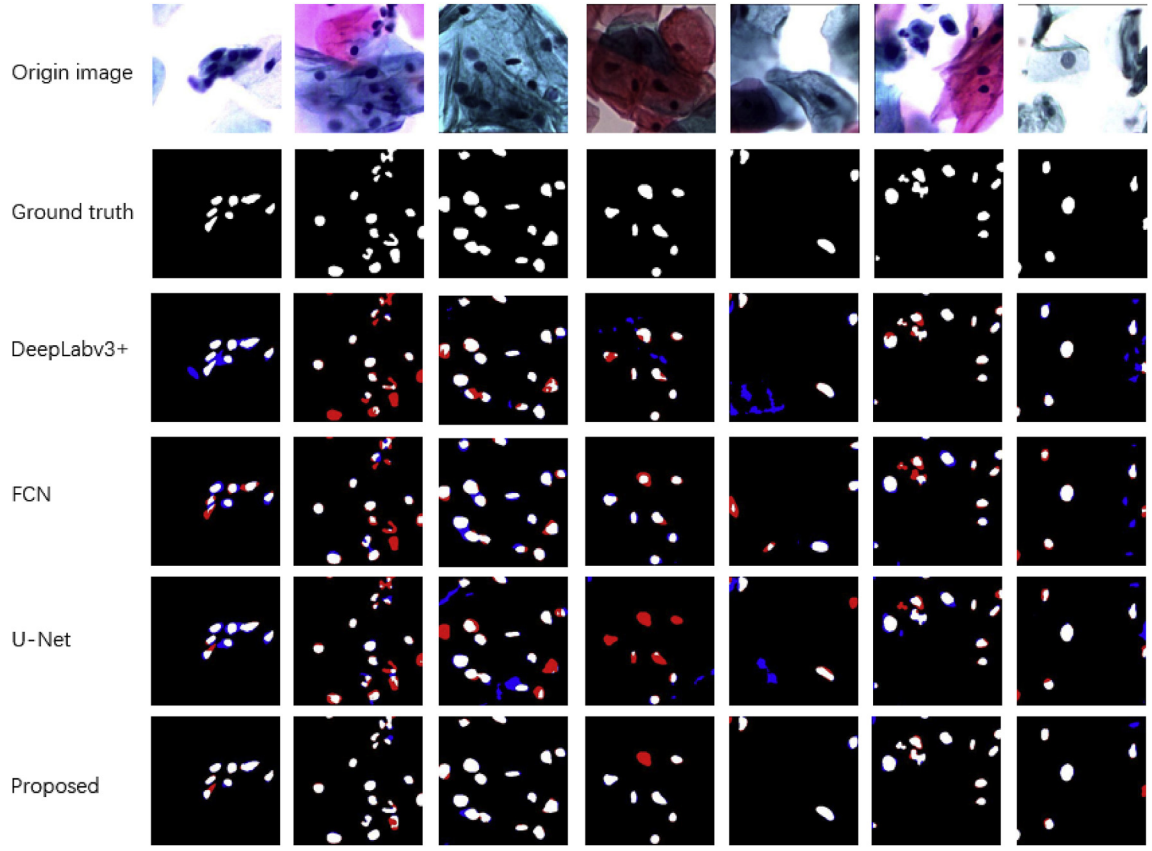


Fig. 7. Visual comparison of nucleus segmentation results. White color is the intersection regions of ground truth nucleus and its corresponding segmented (TP); red color refers to the intersection regions of ground truth nucleus without corresponding segmented nucleus (FN); blue color is the intersection regions of segmented nucleus without corresponding ground truth nucleus (FP).

Table 3
Comparison of segmentation results on the dataset released by ISBI-2014.

Method	$prec_o$	rec_o	$prec_p$	rec_p	Mean DC
2014 [25]	0.903	0.893	0.901	0.916	0.900
2015 [24]	0.958	0.895	**0.968**	0.871	0.914
2015 [26]	0.977	0.883	0.942	0.912	0.921
2018 [18]	0.983	0.959	0.906	0.950	0.925
Proposed	**0.990**	**0.971**	0.902	**0.954**	**0.931**

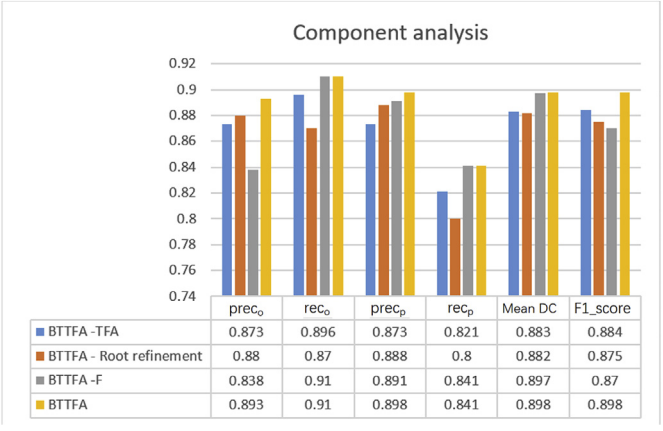


Fig. 8. Component analysis.

Table 4
Comparison segmentation results with different structures.

Method	$prec_o$ (MEAN)	rec_o (MEAN)	. $prec_p$ (MEAN)	rec_p (MEAN)	Mean DC (MEAN)	F1-Score (MEAN)
Fig. 5(1)	0.86	0.901	0.88	0.861	0.9	0.904
Fig. 5 (2)	0.888	0.903	0.889	0.851	0.894	0.895
Fig. 5 (3)	0.891	0.904	**0.89**	0.842	0.895	0.902
BTTFA	**0.901**	**0.908**	0.88	**0.861**	**0.9**	**0.904**

proposed method and the results of [18,24–26]. The results of methods being compared to are obtained directly from the published studies. The experimental results for the cervical cell dataset released by the ISBI-2014 challenge are shown in Table 3. The segmented objects are counted as True-Positive (*TP*) only if the Dice coefficients of them and their corresponding ground truth objects are greater than a certain threshold. The threshold in this experiment was set to 0.7 and was consistent with the threshold set by the methods to be compared.

The nucleus segmentation results of the proposed method and the ISBI-14 segmentation methods of [18,24–26], using the ISBI-14 synthetic test dataset, are shown in Table 3. The MPFW [18] method published in March 2018 that is a traditional method based on a watershed is the best performer on the ISBI-2014 dataset. Compared with the earlier methods [24–26] and the state-of-art method [18], the method proposed in this paper produces the best segmentation according to almost all metrics except $prec_p$. The method proposed by Daniela [24] resulted in the highest value of 0.968; however, this was at the expense of reducing the pixel-level recall rate (rec_p). These results indicate that BTTFA also performs well on the ISBI-2014 dataset.

Table 5
Comparison segmentation results between Focal loss and cross entropy loss.

Method	$prec_o$ (MEAN)	rec_o (MEAN)	. $prec_p$ (MEAN)	rec_p (MEAN)	Mean DC (MEAN)	F1-Score (MEAN)
CEloss	0.901	0.903	**0.89**	0.842	0.898	0.902
BTTFA	**0.901**	**0.908**	0.88	**0.861**	**0.9**	**0.904**

5.4. Component and structure analysis

Several experiments are designed to evaluate the effectiveness of the components of proposed method, as shown in Fig. 8. The orange color indicates the results of the proposed network without the root refinement of combining l*ayer0*, i.e., *layer0* is not connected to Node 9 in Fig. 4. The blue columns show the results of the proposed network without Two-path Fusion Attention and merge node6 and node7 with a normal connection shown in Fig. 6(a) to verify its effectiveness. The gray columns (BTTFA-F) show the effect of replacing TFA by Two-path Attention (TA) without fusion, as shown in Fig. 6(b), which removes the intermediate fusion step from TFA. Yellow columns indicate the best performance of BTTFA.

Fig. 8 shows that each component leads to a significant improvement in performance. The blue and the yellow columns alone illustrate that the straightforward segmentation result of merging features with a normal connection (BTTFA-TFA) is inferior to BTTFA. The orange columns demonstrate that in the absence of the root refinement (BTTFA-Root refinement), the effectiveness of segmentation is also reduced, especially according to metrics rec_o .and rec_p. There are three indicators of BTTFA-F that are very close to those of TFFTA, but they are obtained at the expense of reducing the metrics $prec_o$ and *F1-score*. The generated weight maps of BTTFA-F are only related to a single input feature, as proposed by Hu [48], and the effect is not as satisfactory as that of BTTFA. Since the weight map generated by BTTFA is related to the fusion feature, considering the normal fusion feature of *Feature-High* and *Feature-Low* results in a better performance.

This paper also compares several tree-like structures and loss functions. The experimental results are shown in Table 4 and Table 5, respectively.

The first three rows of Table 4 correspond to other three structures as shown in Fig. 5. Compared with BTTFA, these other three methods have more parameters than BTTFA does, and in general underperform BTTFA.

Focal Loss is modified based on the cross-entropy loss. If we replace the Focal loss by cross-entropy loss, Focal loss produces a slightly better result, shown in Table 5.

The necessity of each part of the BTTFA method can be proven by the overall trend of component and structure analysis experiments. In Fig. 8, Tables 4 and 5, the segmented objects are counted as True-Positive (*TP*) only if the Dice coefficients of them and their corresponding ground truth objects are greater than 0.5. In this paper, the experimental results with the Dice thresholds of 0.6 and 0.7 have also been obtained, and BTTFA always performs the best. Due to space limitations, experimental results for other thresholds are not shown here.

6. Conclusions

Accurate nucleus segmentation is critical for detection of cervical cancer, for it helps distinguish between abnormal and malignant cells. This paper designs a novel deep neural network, named Binary Tree-like Network, for cervical cell nucleus segmentation by effectively combining the characteristics of all layers. Furthermore, the Two-path Fusion Attention is applied to automatically learn a set of weights to indicate the importance of the features that are close to the output of the Final Feature map. A labeled dataset from clinical cervical cytology test is released in this paper. Experiments on two datasets demonstrate that the segmentation results of the proposed method are superior to those of the state-of-the-art methods. In addition, the proposed network has the potential to be used for other nucleus segmentation tasks.

Conflict of interest

None Declared.

Acknowledgments

We are grateful for the help of Dr. Bohai Du, The First Affiliated Hospital, Guangzhou University of Chinese Medicine, who provided and labeled the dataset.

References

[1] International Agency for Research on Cancer, Latest World Cancer Statistics Global Cancer Burden Rises to 14.1 Million New Cases in 2012: Marked Increase in Breast Cancers Must Be Addressed vol. 12, World Health Organization, 2013.
[2] World Health Organization, World Health Organization Reproductive Health, World Health Organization Chronic Diseases, & Health Promotion, Comprehensive Cervical Cancer Control: a Guide to Essential Practice, World Health Organization, 2006.
[3] D. Saslow, D. Solomon, H.W. Lawson, M. Killackey, S.L. Kulasingam, J. Cain, ... N. Wentzensen, American cancer society, American society for colposcopy and cervical pathology, and American society for clinical pathology screening guidelines for the prevention and early detection of cervical cancer, Ca - Cancer J. Clin. 62 (3) (2012) 147–172 https://doi.org/10.1016/j.ypat.2012.11.013.
[4] R.K. Gibb, M.G. Martens, The impact of liquid-based cytology in decreasing the incidence of cervical cancer, Reviews in Obstetrics and Gynecology 4 (Suppl 1) (2011) S2.
[5] E. Bengtsson, Computerized cell image analysis: past, present, and future, Scandinavian Conference on Image Analysis, Springer, Berlin, Heidelberg, 2003, June, pp. 395–407 https://doi.org/10.1007/3-540-45103-x_54.
[6] E. Bengttson, Recognizing signs of malignancy—the quest for computer assisted cancer screening and diagnosis systems, Proc. IEEE Conf. Comput. Intell. Comput. (ICCIC), 2010, December, pp. 1–6 https://doi.org/10.1109/iccic.2010.5705885.
[7] R. Sparks, A. Madabhushi, Explicit shape descriptors: novel morphologic features for histopathology classification, Med. Image Anal. 17 (8) (2013) 997–1009 https://doi.org/10.1016/j.media.2013.06.002.
[8] L.F. Handfield, B. Strome, Y.T. Chong, A.M. Moses, Local statistics allow quantification of cell-to-cell variability from high-throughput microscope images, Bioinformatics 31 (6) (2014) 940–947 https://doi.org/10.1093/bioinformatics/btu759.
[9] M.E. Plissiti, C. Nikou, A review of automated techniques for cervical cell image analysis and classification, Biomedical Imaging and Computational Modeling in Biomechanics, Springer, Dordrecht, 2013, pp. 1–18 https://doi.org/10.1007/978-94-007-4270-3_1.
[10] M.E. Plissiti, C. Nikou, Overlapping cell nuclei segmentation using a spatially adaptive active physical model, IEEE Trans. Image Process. 21 (11) (2012) 4568–4580 https://doi.org/10.1109/tip.2012.2206041.
[11] K. Li, Z. Lu, W. Liu, J. Yin, Cytoplasm and nucleus segmentation in cervical smear images using Radiating GVF Snake, Pattern Recogn. 45 (4) (2012) 1255–1264 https://doi.org/10.1016/j.patcog.2011.09.018.
[12] T. Chankong, N. Theera-Umpon, S. Auephanwiriyakul, Automatic cervical cell segmentation and classification in Pap smears, Comput. Methods Progr. Biomed. 113 (2) (2014) 539–556 https://doi.org/10.1016/j.cmpb.2013.12.012.
[13] A. Kale, S. Aksoy, Segmentation of cervical cell images, Proceedings of the 2010 20th International Conference on Pattern Recognition, IEEE Computer Society, 2010, August, pp. 2399–2402.
[14] X. Yang, H. Li, X. Zhou, Nuclei segmentation using marker-controlled watershed, tracking using mean-shift, and Kalman filter in time-lapse microscopy, IEEE Transactions on Circuits and Systems I: Regular Papers 53 (11) (2006) 2405–2414 https://doi.org/10.1109/tcsi.2006.884469.
[15] H.X. Fei, Y. Qing, A new threshold segmentation method based on the genetic algorithm for enhancing the images, Intelligent System Design and Engineering Application (ISDEA), 2010 International Conference on, vol. 1, IEEE, 2010, October, pp. 103–106.
[16] H.A. Phoulady, D.B. Goldgof, L.O. Hall, P.R. Mouton, April). A new approach to detect and segment overlapping cells in multi-layer cervical cell volume images, Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on, IEEE, 2016, pp. 201–204 https://doi.org/10.1109/ISBI.2016.7493244.

[17] Y. Song, E.L. Tan, X. Jiang, J.Z. Cheng, D. Ni, S. Chen, ... T. Wang, Accurate cervical cell segmentation from overlapping clumps in pap smear images, IEEE Trans. Med. Imaging 36 (1) (2017) 288–300 https://doi.org/10.1109/TMI.2016.2606380.
[18] J. Song, L. Xiao, Z. Lian, Contour-seed pairs learning-based framework for simultaneously detecting and segmenting various overlapping cells/nuclei in microscopy images, IEEE Trans. Image Process. 27 (12) (2018) 5759–5774 https://doi.org/10.1109/tip.2018.2857001.
[19] A. Tareef, Y. Song, H. Huang, D. Feng, M. Chen, Y. Wang, W. Cai, Multi-pass fast watershed for accurate segmentation of overlapping cervical cells, IEEE Transactions on Medical Imaging, 2018 https://doi.org/10.1109/tmi.2018.2815013.
[20] Y. Song, J.Z. Cheng, D. Ni, S. Chen, B. Lei, T. Wang, April). Segmenting overlapping cervical cell in Pap smear images, Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on, IEEE, 2016, pp. 1159–1162 https://doi.org/10.1109/isbi.2016.7493472.
[21] C. Bergmeir, M.G. Silvente, J.M. Benítez, Segmentation of cervical cell nuclei in high-resolution microscopic images: a new algorithm and a web-based software framework, Comput. Methods Progr. Biomed. 107 (3) (2012) 497–512 https://doi.org/10.1016/j.cmpb.2011.09.017.
[22] Y. Al-Kofahi, W. Lassoued, W. Lee, B. Roysam, Improved automatic detection and segmentation of cell nuclei in histopathology images, IEEE (Inst. Electr. Electron. Eng.) Trans. Biomed. Eng. 57 (4) (2010) 841–852 https://doi.org/10.1109/tbme.2009.2035102.
[23] S. Ali, A. Madabhushi, An integrated region-, boundary-, shape-based active contour for multiple object overlap resolution in histological imagery, IEEE Trans. Med. Imaging 31 (7) (2012) 1448–1460 https://doi.org/10.1109/tmi.2012.2190089.
[24] D.M. Ushizima, A.G. Bianchi, C.M. Carneiro, Segmentation of Subcellular Compartments Combining Superpixel Representation with Voronoi Diagrams (No. LBNL-6892E), Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States), 2015.
[25] M. Nosrati, Ghassan Hamarneh, A variational approach for overlapping cell segmentation, ISBI Overlapping Cervical Cytology Image Segmentation Challenge, 2014, pp. 1–2.
[26] Z. Lu, G. Carneiro, A.P. Bradley, An improved joint optimization of multiple level set functions for the segmentation of overlapping cervical cells, IEEE Trans. Image Process. 24 (4) (2015) 1261–1272 https://doi.org/10.1109/tip.2015.2389619.
[27] J. Minnema, M. van Eijnatten, W. Kouw, F. Diblen, A. Mendrik, J. Wolff, CT image segmentation of bone for medical additive manufacturing using a convolutional neural network, Comput. Biol. Med. 103 (2018) 130–139 https://doi.org/10.1016/j.compbiomed.2018.10.012.
[28] O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation, International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, Cham, 2015, October, pp. 234–241.
[29] A. Sanchez-Gonzalez, B. García-Zapirain, D. Sierra-Sosa, A. Elmaghraby, Automatized colon polyp segmentation via contour region analysis, Comput. Biol. Med. 100 (2018) 152–164 https://doi.org/10.1016/j.compbiomed.2018.07.002.
[30] C.H. Lin, Y.K. Chan, C.C. Chen, Detection and segmentation of cervical cell cytoplast and nucleus, Int. J. Imaging Syst. Technol. 19 (3) (2009) 260–270 https://doi.org/10.1002/ima.20198.
[31] M.H. Tsai, Y.K. Chan, Z.Z. Lin, S.F. Yang-Mao, P.C. Huang, Nucleus and cytoplast contour detector of cervical smear image, Pattern Recogn. Lett. 29 (9) (2008) 1441–1453 https://doi.org/10.1016/j.patrec.2008.02.024.
[32] K. Li, Z. Lu, W. Liu, J. Yin, Cytoplasm and nucleus segmentation in cervical smear images using Radiating GVF Snake, Pattern Recogn. 45 (4) (2012) 1255–1264 https://doi.org/10.1016/j.patcog.2011.09.018.
[33] N. Kumar, R. Verma, S. Sharma, S. Bhargava, A. Vahadane, A. Sethi, A dataset and a technique for generalized nuclear segmentation for computational pathology, IEEE Trans. Med. Imaging 36 (7) (2017) 1550–1560 https://doi.org/10.1109/tmi.2017.2677499.
[34] A. GençTav, S. Aksoy, S. ÖNder, Unsupervised segmentation and classification of cervical cell images, Pattern Recogn. 45 (12) (2012) 4151–4168 https://doi.org/10.1016/j.patcog.2012.05.006.
[35] C. Bergmeir, M.G. Silvente, J.M. Benítez, Segmentation of cervical cell nuclei in high-resolution microscopic images: a new algorithm and a web-based software framework, Comput. Methods Progr. Biomed. 107 (3) (2012) 497–512 https://doi.org/10.1016/j.cmpb.2011.09.017.
[36] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, J. Garcia-Rodriguez, A Review on Deep Learning Techniques Applied to Semantic Segmentation, (2017) 1704.06857.
[37] Y. Song, L. Zhang, S. Chen, D. Ni, B. Li, Y. Zhou, ... T. Wang, A deep learning based framework for accurate segmentation of cervical cytoplasm and nuclei, Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE, IEEE, 2014, August, pp. 2903–2906 https://doi.org/10.1109/EMBC.2014.6944230.
[38] L. Zhang, M. Sonka, L. Lu, R.M. Summers, J. Yao, Combining fully convolutional networks and graph-based approach for automated segmentation of cervical cell nuclei, Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on, IEEE, 2017, April, pp. 406–409 https://doi.org/10.1109/ISBI.2017.7950548.
[39] T.Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, S. Belongie, Feature pyramid networks for object detection, CVPR 1 (2) (2017, July) 4.
[40] Z. Lu, G. Carneiro, A.P. Bradley, Automated nucleus and cytoplasm segmentation of overlapping cervical cells, International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, Berlin, Heidelberg, 2013, September, pp. 452–460.
[41] Overlapping Cervical Cytology Image Segmentation Challenge-ISBI, V1, (2014) http://cs.adelaide.edu.au/carneiro/isbi14 challenge/.
[42] S. Xie, R. Girshick, P. Dollár, Z. Tu, K. He, Aggregated residual transformations for deep neural networks, Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on, IEEE, 2017, July, pp. 5987–5995 https://doi.org/10.1109/cvpr.2017.634.
[43] F. Yu, D. Wang, E. Shelhamer, T. Darrell, Deep layer aggregation, 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, IEEE, 2018, June, pp. 2403–2412.
[44] T.Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss for dense object detection, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018 https://doi.org/10.1109/iccv.2017.324.
[45] Q. Hou, M.M. Cheng, X. Hu, A. Borji, Z. Tu, P.H. Torr, Deeply supervised salient object detection with short connections, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 3203–3212 https://doi.org/10.1109/cvpr.2017.563.
[46] Z. Deng, X. Hu, L. Zhu, X. Xu, J. Qin, G. Han, P.A. Heng, R3Net: recurrent residual refinement network for saliency detection, Proceedings of the 27th International Joint Conference on Artificial Intelligence, AAAI Press, 2018, July, pp. 684–690 https://doi.org/10.24963/ijcai.2018/95.
[47] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3431–3440 https://doi.org/10.1063/1.5033713.
[48] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7132–7141.
[49] C. Cao, X. Liu, Y. Yang, Y. Yu, J. Wang, Z. Wang, ... D. Ramanan, Look and think twice: capturing top-down visual attention with feedback convolutional neural networks, Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 2956–2964 https://doi.org/10.1109/iccv.2015.338.
[50] M. Jaderberg, K. Simonyan, A. Zisserman, Spatial transformer networks, Advances in Neural Information Processing Systems, 2015, pp. 2017–2025.
[51] L.C. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, Encoder-decoder with Atrous Separable Convolution for Semantic Image Segmentation, (2018) 1802.02611.
[52] K. Simonyan, A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition, (2014) 1409.1556 https://doi.org/10.1109/acpr.2015.7486599.
[53] M. Everingham, S.A. Eslami, L. Van Gool, C.K. Williams, J. Winn, A. Zisserman, The pascal visual object classes challenge: a retrospective, Int. J. Comput. Vis. 111 (1) (2015) 98–136.
[54] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
[55] L.C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A.L. Yuille, Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs, IEEE Trans. Pattern Anal. Mach. Intell. 40 (4) (2018) 834–848 https://doi.org/10.1109/tpami.2017.2699184.
[56] L.C. Chen, G. Papandreou, F. Schroff, H. Adam, Rethinking Atrous Convolution for Semantic Image Segmentation, (2017) 1706.05587.
[57] G.A. Mishra, S.A. Pimple, S.S. Shastri, An overview of prevention and early detection of cervical cancers, Indian J. Med. Paediatr. Oncol.: official journal of Indian Society of Medical & Paediatric Oncology 32 (3) (2011) 125.
[58] M. Safaeian, D. Solomon, P.E. Castle, Cervical cancer prevention—cervical screening: science in evolution, Obstet. Gynecol. Clin. N. Am. 34 (4) (2007) 739–760 https://doi.org/10.1016/j.ogc.2007.09.004.
[59] S.H. Landis, T. Murray, S. Bolden, P.A. Wingo, Cancer statistics, 1999, Ca - Cancer J. Clin. 49 (1) (1999) 8–31.
[60] P.A. Wingo, C.J. Cardinez, S.H. Landis, R.T. Greenlee, L.A. Ries, R.N. Anderson, M.J. Thun, Long-term trends in cancer mortality in the United States, 1930–1998, Cancer: Interdiscip. Int. J. Am. Cancer Soc. 97 (S12) (2003) 3133–3275 https://doi.org/10.1002/cncr.11380.
[61] H.X. Zhang, Y.M. Song, S.H. Li, Y.H. Yin, D.L. Gao, K.S. Chen, Quantitative detection of screening for cervical lesions with ThinPrep Cytology Test, Clinical Oncology and Cancer Research 7 (5) (2010) 299–302 https://doi.org/10.1007/s11805-010-0535-7.
[62] V.B. Singh, N. Gupta, R. Nijhawan, R. Srinivasan, V. Suri, A. Rajwanshi, Liquid-based cytology versus conventional cytology for evaluation of cervical Pap smears: experience from the first 1000 split samples, Indian J. Pathol. Microbiol. 58 (1) (2015) 17.
[63] J. He, T. Chen, Z. Liu, D. Chen, Abnormal region detection in cervical smear images based on fully convolutional network, IET Image Processing, 2018.

Jianwei zhang is an associate professor at School of Computer Science and Engineering, South China University of Technology. In 1991, she graduated from Zhejiang University with a bachelor's degree in applied mathematics. In 1997, she obtained a master's degree in applied mathematics from Zhejiang University and has worked in the School of Computer Science of South China University of Technology. In June 2006, she received a Ph.D. in School of Computer and Science Engineering from South China University of Technology. From October 2009 to October 2010, she conducted a visiting study at the School of Computer Science at Carnegie Mellon University. Her research interests include computer vision and machine learning.

Her published papers are as follows:

Journal Papers

(1)Jianwei Zhang, Zhenpeng Hu, Guoqiang Han, Xiaozhen He, Segmentation of Overlapping Cells in Cervical Smears Based on Spatial Relationship and Overlapping Translucency Light Transmission Model, Pattern Recognition,2016,60:286–295

(2)Kole T. Roybal Taráz E. Buck, Xiongtao Ruan, Baek Hwan Cho, Danielle J. Clark, Rachel Ambler, Helen M. Tunbridge, Jianwei Zhang, Paul Verkade, Christoph Wülfing, Robert F. Murphy, Computational spatiotemporal analysis identifies WAVE2 and cofilin as joint regulators of costimulation-mediated T cell actin dynamics, Science Signaling. 2016 Apr 19; 9(424):rs3.

(3)ZHANG Jian-wei, HUANG Da-cheng, GUI Jiang-qin, YEWen-zhong, 2D registration based on contour matching for partial matching images, J. Cent. South Univ. (2014) 21: 4553–4562.

(4)Zhang Jian-wei, Huang Da-cheng, Hu Zhen-peng, Zhang Jing. Multi-Scale Structure Tensor Image Retrieval Algorithm Based on Edge Tangent Flow Field. Journal of South China University of Technology(Natural Science), 2015,43(5):107–113

(5)Zhang Jian-wei, Han Guo-qiang, Wo Yan, Image registration method based on mutual information about distance field of image edges. Journal on Communications,2006,27(7): 87–93.

(6)Zhang Jian-wei, Han Guo-qiang, Image Registration Approach Based on Nearest-Point Pseudo Gravitation Field, Journal of South China University of Technology(Natural Science),2006,34(6):6–11.

Conference Papers

(1)Jian-wei Zhang, Shan-shan Zhang, Guo-hong Yang, Da-cheng Huang, Lin Zhu,Dongfa Gao. Adaptive Segmentation of Cervical Smear Image Based on GVF Snake Model. Proceedings of the 2013 International Conference on Machine Learning and Cybernetics, 2013:890–895.

(2)Jian-wei Zhang,Min-chao Lian,Wan-peng Wang, Lin Zhu. Detection of Abnormal Nuclei in Cervical Smear Images Based on Visual Attention Model.Proceedings of the 2013 International Conference on Machine Learning and Cybernetics, 2013:920–924.

Zhenmei Liu received the B.E. degree from the School of Computer Science and Technology, Jiangxi Normal University, China, in 2017. And she is currently a graduate student at the School of Computer Science and Engineering, South China University of Technology. Her research interests include image segmentation, object detection and machine learning algorithms.

Bohai Du is a doctor in the Department of Pathology, the First Affiliated Hospital of Guangzhou University of Traditional Chinese Medicine. His research interests is the diagnosis of cervical liquid-based cells.

Junting He received the B.S. degree from the School of Computer Science and Engineering, South China University of Technology, China, in 2016. And from 2016, she is a graduate student and her research interests include computer vision and machine learning.

Guanzhao Li received the B.E. degree from the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China, in 2017. And he is currently a graduate student at the School of Computer Science and Engineering, South China University of Technology. He focuses on the area of computer vision.

Danni Chen received the B.S. degree from the School of Computer Science and Engineering, South China University of Technology, China, in 2016, where she is currently pursuing the M.S. degree. Her current research interests include image generation, natural language generation and machine learning algorithms.