GCSBA-Net: Gabor-Based and Cascade Squeeze Bi-Attention Network for Gland Segmentation

Zhijie Wen, Ru Feng, Jingxin Liu, Ying Li, and Shihui Ying

***Abstract*—Colorectal cancer is the second and the third most common cancer in women and men, respectively. Pathological diagnosis is the "gold standard" for tumor diagnosis. Accurate segmentation of glands from tissue images is a crucial step in assisting pathologists in their diagnosis. The typical methods for gland segmentation form a dense image representation, ignoring its texture and multi-scale attention information. Therefore, we utilize a Gabor-based module to extract texture information at different scales and directions in histopathology images. This paper also designs a Cascade Squeeze Bi-Attention (CSBA) module. Specifically, we add Atrous Cascade Spatial Pyramid (ACSP), Squeeze Position Attention (SPA) module and Squeeze Channel Attention module (SCA) to model semantic correlation and maintain the multi-level aggregation on the spatial pyramid with different dilations. Besides, to solve the imbalance of data distribution and boundary blur, we propose a hybrid loss function to response the object boudary better. The experimental results show that the proposed method achieves state-of-the-art performance on the GlaS challenge dataset and CRAG colorectal adenocarcinoma dataset, respectively.**

***Index Terms*—Gabor-based encoder module, cascade squeeze bi-attention, gland segmentation, hybrid loss function.**

I. Introduction

COLORECTAL cancer [1] (carcinoma of colon and rectum) is a common malignant tumor in the gastrointestinal tract. Its incidence and mortality are second only to gastric cancer, esophageal cancer and primary liver cancer in the digestive system malignant tumors [2]. It can be characterized by a glandular structure. Pathologists use gland morphology to assess the degree of grading or differentiation of various adenomas such as breast, prostate, and colon. In clinical practice, pathologists visually analyze the biopsy tissue slides under the microscope. Current manual assessment of Hematoxylin and Eosin (H&E) stained histology slides is time-consuming and subjective and it can often be challenging to make a diagnosis in ambiguous areas. Therefore, pathologists urgently need to use the computer-aided technology for histopathology image analysis. Nevertheless, this task remains challenging due to several reasons: First, accurate gland segmentation is an important prerequisite for obtaining reliable morphological data and the glandular morphology significantly varies between grades. Second, due to the interaction between cells, the glands always adhere to each other. The adherent glands must be separated and treated as independent glands. Third, the glands may have inconsistent staining and blurred contours, which are difficult to distinguish. The above mentioned difficulties are highlighted in Fig. 1.

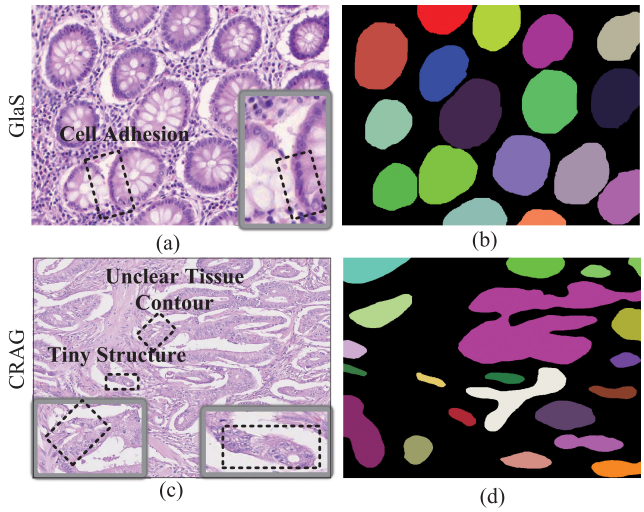


Fig. 1. (a) and (c) are the pathological images with different shapes and blurred contours. (b) and (d) are the ground truth. The mentioned difficulties are zoomed in the grey boxes.

Pathologists are prone to be fatigue and misdiagnosis after working for a long time. With the development of Whole Slide Images (WSIs) technology, the number and types of histological images are increasing rapidly. More and more histopathological data are needed to be analyzed. However, it is impractical for

Manuscript received December 4, 2019; revised April 26, 2020 and July 15, 2020; accepted August 6, 2020. Date of publication August 11, 2020; date of current version April 5, 2021. This work was supported in part by the National Natural Science Foundation of China under Grants 11701357, 11971296, 81830058, and 11601315, and in part by The Capacity Construction Project of Local Universities in Shanghai under Grant 18010500600. *(Corresponding author: Shihui Ying.)*

Zhijie Wen, Ru Feng, and Shihui Ying are with the Department of Mathematics, College of Sciences, Shanghai University, Shanghai 200444, China (e-mail: wenzhijie@shu.edu.cn; smileruru@shu.edu.cn; shying@shu.edu.cn).

Jingxin Liu is with the College of Information Engineering, Shenzhen University, Shenzhen 518060, China, and also with the Histo Pathology Diagnostic Center, Shanghai 200444, China (e-mail: jingxin.liu@outlook.com).

Ying Li is with the School of Computer Engineering and Science, Shanghai University, Shanghai 200444, China (e-mail: yinglotus@shu.edu.cn).

Digital Object Identifier 10.1109/JBHI.2020.3015844

pathologists to extract important morphological features from large-scale histology images. The development of digital pathology has led to new methods of the computational pathology, mainly for computer aided diagnosis systems. In this field, tumor type prediction and grading, tissue segmentation, and cancer cells identification form several hot topics.

In this paper, we present an efficient Gabor-based and cascade squeeze bi-attention network (GCSBA-Net) for helping pathologists to make diagnostic decisions. Pathological images can be regarded as typical texture images. Due to the fixed encoder-decoder structure, U-Net [3] is not suitable for processing texture images. It cannot extract the features of different scales and directions and Gabor wavelets just make up for this shortcoming. Therefore, in the feature extraction process, we use the Gabor-based encoder module to obtain rich texture information of the objects in the histological image. Gabor wavelets also modulate the CNN kernel in the neural network and increase its interpretability. Moreover, to fuse multiple location information, the original attention mechanism [4], [5] needs to be generated a huge attention map, which leads to the massive parameters. To overcome the above shortage, we also design a Cascade Squeeze Bi-Attention (CSBA) module to capture global dependencies, which is more lightweight and efficient. It aggregates multi-level attention to obtain the relationship of multi-channel and multi-position through a local and global weight adjustment mechanism. It is critical for segmentation, especially the glandular boundary segmentation. After feature extraction, we use the idea of DenseNet [6] to restore the image information in a densely connected way. The improved decoder module restores more pixel information, thereby improving the segmentation performance. Besides, a new hybrid loss function has been developed to help our network achieve a better diagnosis given an imbalanced data and capture the structural information in an image.

The contributions of this work are summarized as follows:

- We propose a novel Gabor-based encoder module to learn more texture features of the histology images.
- The Bi-Attention mechanism is introduced in the proposed Cascade Squeeze Bi-Attention module, which captures the spatial and channel information at different scales for gland segmentation.
- A novel hybrid loss function is proposed to balance the imbalanced data distribution.

This paper is organized as follows: Section II describes the related work. In Section III, we show the proposed approach. The detailed experiments and analysis are described in Section IV and Section V. The conclusion is given in Section VI.

II. Related Work

In this section, we first summarize the current gland segmentation methods, then review the Gabor wavelet transform, and finally introduce the attention model.

Semantic Segmentation: The typical histopathological segmentation methods were based on morphology [7], [8], texture [9], simple threshold method [10], watershed algorithm and the glandular structure [11], [12]. Moreover, Convolutional neural networks (CNNs), especially Fully Convolutional Networks (FCNs) [13] have achieved the best performance in medical image cell segmentation [14]–[17]. Ronneberger *et al.* [3] proposed the U-Net, which is based on the FCN network architecture and added an asymmetric extension path to enable accurate segmentation results with less training images. Chen *et al.* [18], [19] proposed a deep contour-aware network based on the FCN learning framework to segment contours and gland objects simultaneously, and won the 2015 MICCAI gland segmentation challenge [20]. Raza *et al.* [21] proposed the Micro-Net, which used a multi-resolution of the input image and multi-resolution deconvolution filters for the segmentation of various objects in microscopy images. Graham *et al.* [22] used the original image to compensate for the information loss caused by max pooling. Also, it adopted atrous spatial pyramid pools (ASPP) with different expansion rates to maintain the resolution of the features. It achieved state-of-the-art performance on the dataset of the MICCAI Gland Segmentation challenge. Compared with the traditional segmentation methods, the CNN-based segmentation methods exhibit excellent performance.

Segmentation using Gabor wavelets: Gabor wavelet simulates human vision system, which can detect multi-directional and multi-scale features. It is suitable for texture representation and recognition [23]. It is a special convolution with invariance to rotation, scale and translation, which is therefore widely used in image processing. Kinnikar *et al.* [24] used Gabor as a preprocessing tool to generate Gabor features and then used them as the input to CNNs. To reduce the complexity of CNN training and improve the robustness of feature representation, Sarwar *et al.* [25] used a Gabor filter in the first or second convolutional layer. Luan *et al.* [26] proposed Gabor convolutional networks (GCNs or Gabor CNNs) such that the robustness of learned features against the orientation and scale changes can be reinforced. A recent related work, Yoo *et al.* [27] proposed a wavelet as an alternative to traditional pooling, which accurately reconstructed the local information of the image.

Attention mechanism: CNNs based on attention module architecture have recently been shown to be helpful for models focusing on the useful information in a wide variety of tasks while suppressing irrelevant information [28]–[30]. Squeeze-and-Excitation Networks (SE-Net) [31] introduced the idea of attention. For each channel, it meant that a weight was used to indicate the importance of this channel in the next stage. Recently, some researchers have explored the self-attention module [5]. Dual attention network (DANet) [4] used two attention modules in the network to capture pixel relationships and channel dependencies, respectively. Point-wise spatial attention network (PSANet) [32] learned an attention map to connect each location in the feature map with all other locations to achieve dynamic aggregation of the context information.

III. Methodology

In this section, we describe the proposed method for gland segmentation. As shown in Fig. 2, our network consists of three parts: Gabor-based auto-encoder, CSBA module and Dense auto-decoder. The Gabor-based auto-encoder module is a densely supervised structure similar to the U-Net [3]. It includes two 3×3 convolution layers, which increase the

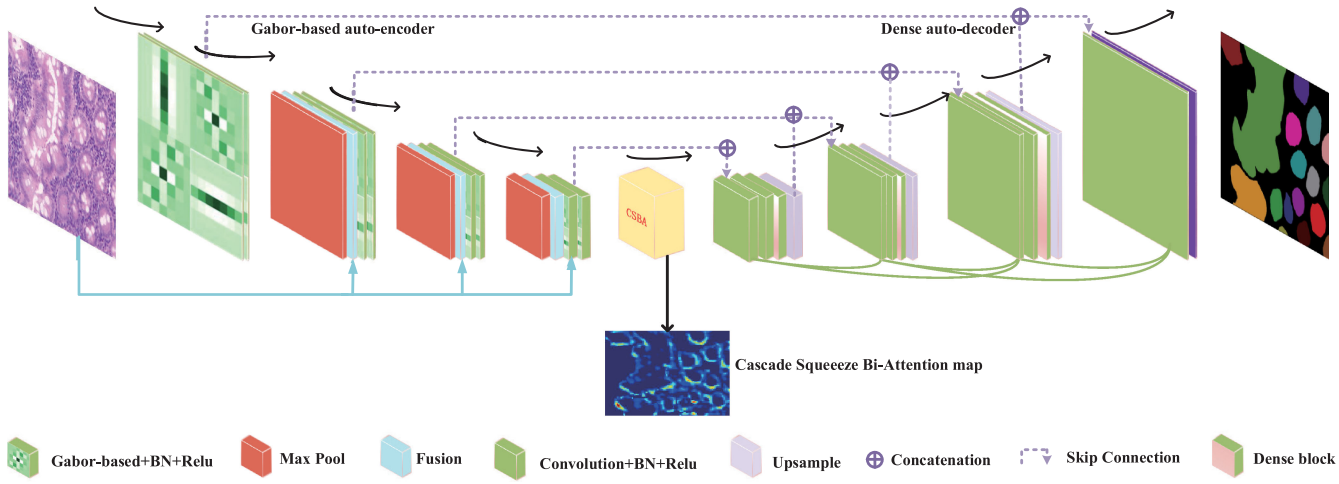


Fig. 2. An overview of the proposed GCSBA-Net for gland segmentation.

TABLE I
THE PARAMETER SETTINGS OF THE PROPOSED NETWORK

Layers	Filter Shape	Output Size
GaborBlock_1	$\begin{Bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{Bmatrix}$	$512\times512\times64$
GaborBlock_2	$\begin{Bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{Bmatrix}$	$256\times256\times128$
GaborBlock_3	$\begin{Bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{Bmatrix}$	$128\times128\times256$
GaborBlock_4	$\begin{Bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{Bmatrix}$	$64\times64\times512$
CSBA	refer to Figs. 4 and 5.	$64\times64\times512$
DenseConvBlock_4	$\begin{Bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{Bmatrix}$	$64\times64\times512$
DenseConvBlock_3	$\begin{Bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{Bmatrix}$	$128\times128\times256$
DenseConvBlock_2	$\begin{Bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{Bmatrix}$	$256\times256\times128$
DenseConvBlock_1	$\begin{Bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{Bmatrix}$	$512\times512\times64$

non-linear expression ability of the network and reduce the number of parameters. Each convolution is followed by a nonlinearity activation function (ReLU) and a 2×2 max pooling layer with a step size of 2 for downsampling. In each downsampling step, in order to map features to higher dimensions, we double the number of feature channels. The most critical part is to use the Gabor-based kernel to extract features. It makes the network more concerned with shape and texture information. The CSBA module captures spatial and channel information at different scales in the network through a multi-scale strategy and the weighted idea. In order to recover the image information, we have added a dense block and dense connection to the decoder. The parameter settings of the convolutional layers are listed in Table I. Besides, the general training process of the proposed GCSBA-Net can be briefly summarized as Algorithm 1.

In the following section, we introduce the Gabor-based encoder module and then describe our newly designed Cascade Squeeze Bi-Attention module in detail. Finally, we introduce the novel loss function.

A. Gabor-Based Auto-Encoder Module

Most of the histopathology images have complex texture features and different styles, and the Gabor wavelet is a more advantageous mathematical tool for processing texture images. [33]. It has been shown that Gabor filters are quite useful to extract highly informative features. Therefore, we use Gabor kernel to extract multi-directional and multi-scale features from the histopathology image. The two-dimensional Gabor function is expressed as Equation (1).

$$g_{\theta,v}(x,y)=\frac{1}{2\pi\sigma^2}\exp\left(-\frac{x'^2+y'^2}{2\sigma^2}\right)\exp\left(i\left(vx'+\psi\right)\right), \quad (1)$$

where $x'=x\cos(\theta)+y\sin(\theta)$, $y'=-x\sin(\theta)+y\cos(\theta)$. (x, y) represents the pixel coordinate position. θ represents the orientations. v is the scales. ψ is the phase offset. σ is the standard deviation of the Gaussian envelope.

In the process of image segmentation, the existing methods generally use deep feature maps with high-level semantics, while ignoring shallow feature maps with rich low-level object shapes and texture information. To overcome the above issue, we propose an encoder module based on the Gabor wavelets.

Fig. 3 shows the process of the typical convolutional kernels modulated by the Gabor wavelets. The Gabor-based feature maps can be obtained after training. Compared to the typical CNN, Gabor-based convolutions learn more rubust spatial features. In addition, compared with Gabor, Gabor-based convolutions enhanced the nonlinear fitting ability. The Gabor-based convolution is denoted as Equation (2).

$$C^{v}_{i,\theta}=C_i\circ g_{\theta,v}(x,y), \quad (2)$$

where $\circ$ is an element-by-element product operation. C_i is a 3×3 convolution filter. $g_{\theta,v}(x,y)$ represents Gabor filters. v and θ are scales and orientations, respectively. $C^{v}_{i,\theta}$ is the modulated filter. There are two parameters of the Gabor-based filters, scales and directions. For each convolutional layer, the number of scales is initially set as 4, i.e., 1, 2, 3, 4. For each

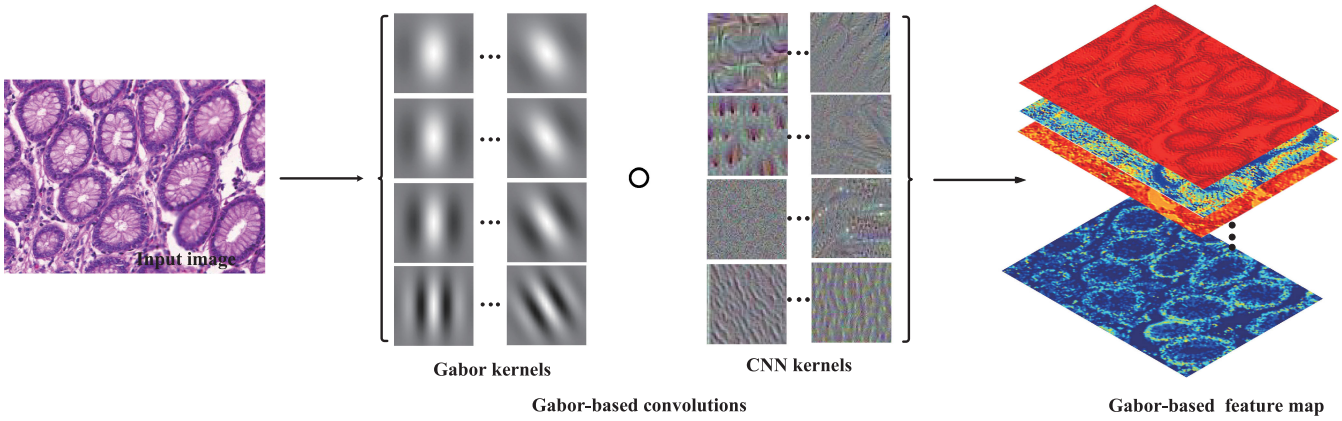


Fig. 3. Gabor-based convolutions and their feature maps.

scale, the number of directions equals to a quarter of the number of the output channels.

The backpropagation (BP) process is shown as Equation (3).

$$C_i = C_i - \eta . \frac{\partial L}{\partial C_i} \tag{3}$$

L is the loss function. In back propagation, the learned CNN kernel C_i need to be updated, but Gabor function does not change. Therefore, we only need to calculate the gradient of the convolution kernel. From the above analysis, we can find that this module captures multi-scale and multi-directional features on each layer.

B. Cascade Squeeze Bi-Attention Module

Considering the local and global weight adjustment mechanisms, we design the Cascade Squeeze Bi-attention (CSBA) module to capture channel and spatial information at different scales. It consists of three modules: the Atrous Cascade Spatial Pyramid (ACSP) module, the Squeeze Position Attention (SPA) module and the Squeeze Channel Attention (SCA) module.

Inspired by PSPNet [34], DeepLabv3+ [35] and the Inception module [36]–[38], an ACSP module is designed, which is shown in Fig. 4. ACSP module uses different receptive fields in each branch to learn features. With the gradual increase of atrous convolution rates, the receptive fields encode the global and local environments at four scales of 3, 7, 9, and 17, respectively. In each branch, the features after Relu activation are fed into the SPA module and SCA module, respectively. Finally, we sum the features extracted by the bi-attention modules of different scales. Small reception fields capture the small objects. On the other hand, large reception fields usually get the large objects. In general, they all obtain the high-level semantic information.

As shown in Fig. 5(a), the SPA module extends the weight adjustment channel of the SE-Net [31]. We compress the spatial information to obtain the pixel-level feature. This method extracts the feature $F_k(i,j)$ of the kth channel from local feature $F(i,j) \in R^{C\times H\times W}$, which is obtained by each branch of the ACSP module. The position attention module first applies two 1×1 convolution layers to this feature, which generates two different feature maps Q_k and K_k, $k \in \{1, 2, \ldots, C\}$. Then the

Algorithm 1: Gabor-Based and Cascade Squeeze Bi-Attention Network.

Input: Training dataset $X = \{x_1, x_2, \ldots, x_n\}$ annotations $\mathrm{Y} = \{\mathrm{y}_1, y_2, \ldots, y_n\}$.
Output: pixel-level segmentation results.

Def Gabor-based:
    # Set v and θ (v is the scales. u represents the orientations)
    $g_{\theta,v}(x,y) = \frac{1}{2\pi\sigma^2} \exp(-\frac{x'^2+y'^2}{2\sigma^2}) \exp(i(vx' + \psi))$
    Return $g_{\theta,v}(x,y)$
Start training
Train patch-based network.
Modulate the typical convolutions using Equation (2).
Cascade squeeze bi-attention module based on Equations (8) and (10).
Calculate the novel hybrid loss function using Equation (11) and then
update parameters.

feature P_k is obtained by merging an average pooling layer and max pooling layer on Q_k. The position squeeze weights S_k are obtained after two 1×1 convolution layers and a sigmoid activation layer. Then S_k are multiplied with $K_k(i,j) \in R^{C\times W\times H}$ to get the attention map. Finally, we multiply it by a scale parameter λ_p, which is then summed with the features $F_k(i,j)$ to get the weighted attention map-$F_k^{SPA}(i,j)$.

$$K_k(i,j) = k_{1\times1} * F_k(i,j), \tag{4}$$

$$Q_k(i,j) = k_{1\times1} * F_k(i,j), \tag{5}$$

$$P_k = \frac{1}{H \times W} \sum_{i=1}^{W} \sum_{j=1}^{H} Q_k(i,j) + \max\{Q_k(i,j)\}, \tag{6}$$

$$S_k = \sigma(k_{1\times1} * (k_{1\times1} * P_k)), \tag{7}$$

$$F_k^{SPA}(i,j) = \lambda_p(S_k \circ K_k(i,j)) + F_k(i,j), \tag{8}$$

where λ_p is a weight and it is initialized as 0. $*$ indicates the convolution operation. As can be seen from Equation (8), the feature $F_k^{SPA}(i,j)$ is the original feature of kth channel added

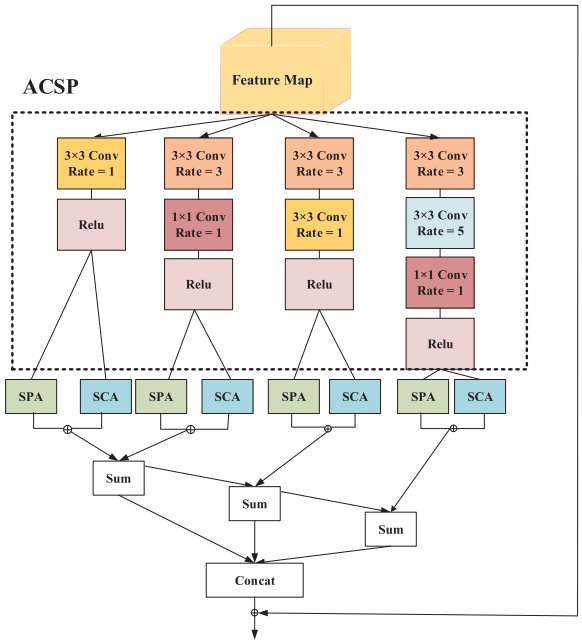


Fig. 4. The Cascade Squeeze Bi-Attention module (CSBA).

with the weighted feature of the spatial location. Finally, we concatenate the features of all channels to form the F_{SPA} feature map. Such spatial information retained by the SPA module enhances the accuracy of pixel-level dense prediction.

The SCA module is illustrated in Fig. 5(b). We first obtain the squeeze channel weight $S(i,j)$ by the original feature $F(i,j) \in R^{C\times H\times W}$ through a 1×1 convolution layer and a sigmoid function. Then perform an element-wise sum operation between $F(i,j)$ and $S(i,j)$ to get the channel attention map. Finally, we multiply the result by a scale parameter λ_c and perform an element-wise sum operation with the feature $F(i,j)$ to obtain the final output.

$$S(i,j) = \sigma(k_{1\times1} * F(i,j)), \tag{9}$$

$$F^{SCA}(i,j) = \lambda_c(S(i,j) \circ F(i,j)) + F(i,j). \tag{10}$$

The formula integrates the weighted features of all squeeze channels into the original features. It emphasizes the class-related feature mappings and improves the feature discriminating ability between different classes.

Finally, the SPA features and the SCA features are summed in an element-wise way to generate the bi-attention feature effectively.

C. Loss Function

In the area of medical image segmentation, the overlap, border blurring and imbalanced data distribution are still challenging. To solve these problems, we propose a novel hybrid loss function, which is divided into three parts: $F_\beta{-}focal$ loss function, binary cross entropy (BCE) loss function and structural similarity ($SSIM$) loss function. In order to solve the problem of imbalanced data distribution and overlap adaptively, this paper designs a $F_\beta{-}focal$ loss function. BCE loss function [39] is widely used in segmentation tasks. $SSIM$ loss function [40] is introduced to solve the problem of the unclear boundaries in the segmentation task. The proposed total loss function can be formulated as Equation (11).

$$L = \lambda_1 L_{F_\beta - focal} + \lambda_2 L_{bce} + \lambda_3 L_{ssim}, \tag{11}$$

where λ_1, λ_2 and λ_3 are hyperparameters.

$L_{F_\beta - focal}$ is a loss function to improve the overlap and data distribution imbalance. The F_β is shown as Equation (12).

$$F_\beta(c) = \frac{TP_c}{TP_c + \frac{\beta^2}{1+\beta^2}FN_c + \frac{1}{1+\beta^2}FP_c}, \tag{12}$$

where $TP_c = \sum_{i=1}^{N} p_{ic}g_{ic}$, $FN_c = \sum_{i=1}^{N}(1-p_{ic})g_{ic}$ and $FP_c = \sum_{i=1}^{N} p_{ic}(1-g_{ic})$ are the true positives, false negatives and false positives being class c, respectively. $p_{ic} \in [0,1]$ is the predicted probability for pixel i being class c. $g_{ic} \in \{0,1\}$ is the ground truth for pixel i being class c. N represents the total number of pixels in the image. β is a hyperparameter that controls the balance between the precision and the recall (FPs and FNs). The F_β score is adopted to the loss function by minimizing $\sum_c 1 - F_\beta(c)$.

According to the Focal loss function [41], by reducing the weight of easy classification examples, the training focus is on the difficult classification examples. This idea has been extended in recent work, for example, the Dice loss function [42], [43], a combination of the Dice loss function and Focal loss function [44], and a combination of the BCE and IOU loss function [45]. Similarly, the proposed loss function $L_{F_\beta - focal}$ is represented as:

$$L_{F_\beta - focal} = \sum_c (1 - F_\beta(c))^{\frac{1}{\gamma}}, \tag{13}$$

where γ is a hyperparameter. When a pixel is misclassified and $F_\beta(c)$ is large, the loss function is unaffected. However, when $F_\beta(c)$ is small and the pixels are misclassified, the value of $L_{F_\beta - focal}$ decreases significantly.

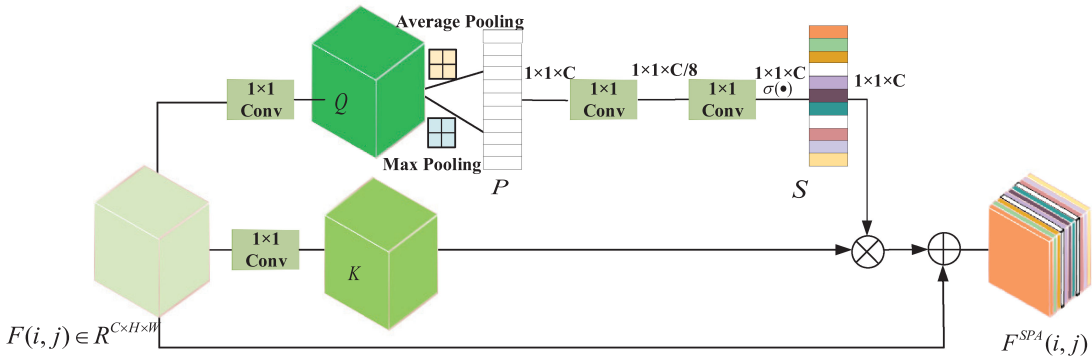
We assume that using a higher β helps us shift the focus to improve the recall, thus achieve a better balance between the precision and the recall. $L_{F_\beta - focal}$ with hyper-parameter β, γ can be generalized to Dice similarity coefficient, the Jaccard index, and the Tversky Loss function [46]. When $\beta{=}\ 1$ and $\gamma{=}\ 1$, $L_{F_\beta - focal}$ is equal to Dice loss function. When $\beta{=}\ 2$ and $\gamma{=}\ 1$, $L_{F_\beta - focal}$ simplifies to the F_2 score.

BCE loss function [39] is expressed by

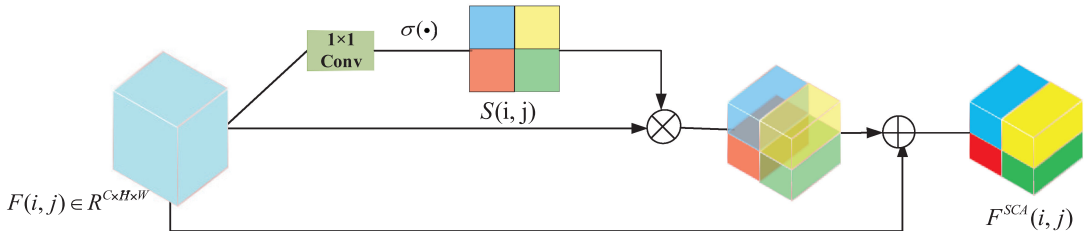
$$L_{bce} = -\frac{1}{N}\sum_{i=1}^{N} g_i \log(p_i) + (1-g_i)\log(1-p_i), \tag{14}$$

where $p \in [0,1]$ and $g \in \{0,1\}$ are the prediction probability map and the ground-truth label, respectively. N is the total number of pixels in the segmentation image.

SSIM loss function [40] is a well-known quality metric that measures the similarity between two images. It is based on a perceptual model designed to improve the effectiveness of image quality assessment in the human visual system. SSIM models image distortion as a combination of correlation loss, brightness



(a) Squeeze Position Attention Module



(b) Squeeze Channel Attention Module

Fig. 5. The details of Squeeze Position Attention Module (SPA) and Squeeze Channel Attention Module (SCA) are illustrated in (a) and (b).

distortion, and contrast distortion. SSIM gets a form as follows.

$$L_{ssim(x,y)} = \left[\frac{2u_x u_y + \varepsilon_1}{u_x^2 + u_y^2 + \varepsilon_1}\right]\left[\frac{2\sigma_x\sigma_y + \varepsilon_2}{\sigma_x^2 + \sigma_y^2 + \varepsilon_2}\right], \quad (15)$$

where $u_x = \bar{x}$, $u_y = \bar{y}$, $\sigma_x = \mathrm{var}(x)$, $\sigma_y = \mathrm{var}(y)$. σ_{xy} is their covariance and $\varepsilon_1, \varepsilon_2 \ll 1$ are two small positive constants. At this case, we set $\varepsilon_1 = 0.01^2, \varepsilon_2 = 0.02^2$.

IV. Experimental Results

We carry out a comprehensive experiment on the CRAG and GlaS datasets. We first introduce two datasets and implementation details. Then, the results on CRAG and Glas datasets are exhibited.

A. Datasets

The Gland Segmentation (GlaS) challenge dataset [19] The GlaS dataset mainly consists of 16 WSIs, which are obtained by the Zeiss MIRAX MIDI Slide Scanner. We split them into two parts: 85 (benign 37, malignant 48) images for training and 80 images for testing (Part A: 60, Part B: 20). All training images have associated ground truth for instance-level segmentation.

The colorectal adenocarcinoma gland (CRAG) [22] We have 213 $H\&E$ stained CRAG images from 38 WSIs scanned by an Omnyx VL120 scanner with a pixel resolution of 0.55 um/pixel. The CRAG datasets are divided into 173 training images and 40 testing images. Each image has approximately 1512×1516 pixels with a corresponding instance-level ground truth.

For both datasets, we set 25% of the training set to evaluate the performance of our model during training. We extract patches with size 512×512. To reduce overfitting, we use a data augmentation strategy to expand the training dataset. Data augmentation includes the following operations, such as random flipping, rotation, elastic deformation, cropping and brightness distortion, which makes the model more robust to the size and shape of various input cells.

B. Evaluation Metrics

In this paper, we use three types of metrics, including the pixel-level metric F_1 score, the object-level Dice index and the measure of shape similarity Hausdorff distance.

F_1 score: the metric F_1 score is adopted to evaluate performance, which is the harmonic mean of the precision and the recall. It is defined as:

$$F_1 = 2 \times \frac{\mathrm{pre} \times rec}{pre + rec}. \quad (16)$$

The higher the F_1 score is, the better the intersection between the ground truth and the predicted segmentation mask is.

Dice index: The Dice index is the measure of set similarity. We assume that $G = \cup_{i=1,2,\ldots K} G_i$ is the ground truth object and $\mathrm{P} = \cup_{i=1,2,\ldots K} P_i$ is the predicted result. The Dice index is defined by

$$Dice = \frac{2\,|G \cap P|}{|G| + |P|}. \quad (17)$$

A higher Dice index indicates a better result.

Hausdorff distance: The Hausdorff distance is a measure of the similarity between the shape of the segmented object and the

TABLE II
THE COMPARATIVE ANALYSIS ON THE CRAG DATASET

Methods	F1 Score	Obj. Dice	Obj. Hausdorff	Rank
FCN-8 [13]	0.558	0.64	435.43	7
U-Net [3]	0.6	0.654	354.09	6
DCAN [19]	0.736	0.794	218.76	3
SegResNet [47]	0.638	0.742	238.43	5
DeepLabv3+ [35]	0.653	0.762	288.31	4
MILD-Net [22]	0.825	0.875	160.14	2
GCSBA-Net	**0.836**	**0.894**	**146.77**	1

shape of the ground truth object.

$$\text{Hausdorff} = \max\left\{\sup_{x\in G}\inf_{y\in P}\|x-y\|, \sup_{y\in P}\inf_{x\in G}\|x-y\|\right\}. \tag{18}$$

The smaller the Hausdorff distance is, the higher the similarity between the two sets is.

C. Implementation Details

We use the Pytorch [49] to implement GCSBA-Net and ablation studies. Adam algorithm [50] is adopted to train the network, with default values of $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The initial learning rate is set to 10^{-4} and the batch size is set to 2. We use a poly learning rate schedule where the initial learning rate is multiplied by $1 - (\frac{iter}{\max\_iter})^{power}$ with power $= 0.9$. All of our models are trained on NVIDIA GTX 1080 Ti GPU.

D. Comparison With Other Methods

Quantitative evaluation: As can be seen from Table II, GCSBA-Net gets 83.6% in F_1 score which is 1.1% higher than MILD-Net [22], and gets 89.4% in Dice index which is 1.9% higher than MILD-Net [22]. Our method is 13.37 higher than MILD-Net in Hausdorff distance. Fig. 7 shows the ROC curves and AUC values of different algorithms. It shows that the AUC value of the proposed method (0.91) is superior to the AUC values of other methods. In order to emphasize the good generalization ability of our method on a different dataset, we also implement the experiment on GlaS datasets. The top ten methods in Table III are the rank listing methods on the 2015 MICCAI GlaS Challenge Dataset. The proposed method achieves the state-of-the-art result in test A.

Qualitative evaluation: In Fig. 6 and Fig. 11, we present qualitative results for both the CRAG and GlaS datasets, further illustrating the superior performance of GCSBA-Net. The proposed method pays more attention to the boundary features of cells in histology images. It can accurately segment the challenging segmentation scenes. For example, the second row in Fig. 6, tissue cells are crowded. Deeplabv3+, FCN-8, SegResNet, and U-Net all recognize different cells to be the same instance. Instead, the GCSBA-Net method divides cells into different instances. The cell boundaries become clearer. U-Net generates label mappings for holes in cell regions, but the proposed GCSBA-Net avoids this error with the atrous cascade spatial pyramid module. To further demonstrate the effectiveness of the approach, we also perform a visual analysis on the Glas dataset, as shown in Fig. 11. Compared with other methods, our GCSBA-Net accurately detects objects and differentiates them into different instances.

V. DISCUSSION

To further verify the efficiency of GCSBA-Net, the ablation study is designed for GCSBA-Net on the validation dataset.

A. Visualization of Gabor Kernel

The detailed differences between the features extracted by the typical CNN and the features extracted by GCSBA-Net are shown in Fig. 8. We notice that the GCSBA-Net features help the convolutional network to obtain texture and shape information. Both GCSBA-Net and typical CNN kernel adopt 3×3 convolutions. The typical CNN has a strong fitting ability, but it also has a fatal shortcoming, whose kernels are randomly generated and there is no relationship between them. It hardly learns robust features. The Gabor-based convolution kernels are generated from a meaningful Gabor function. The strategy of using Gabor wavelet to modulate CNN kernels is useful for extracting more texture features and enhances their interpretability and effectiveness. Gabor wavelet is a useful mathematical tool to extract the texture details of histology images, which are important in gland segmentation. It is the reason why we use Gabor-based kernel to extract features. In short, these visualizations further demonstrate that Gabor-based convolution can extract better boundaries and more texture features simultaneously.

B. The Effect of GCSBA-Net Architecture

Table IV shows the quantitative comparison of the different parts of GCSBA-Net. The first row is the performance of the U-Net [3]. The proposed baseline uses a designed encoder-decoder structure with dense module. Specifically, we gradually added the Gabor-based encoder module, SPA module, SCA module and ACSP module based on baseline. The second row of Table IV shows that the Gabor-based encoder module makes the segmentation more accurate. Compared with baseline+Gabor, the F_1 score and Dice index have increased about 2% after adding the SPA module. Baseline+Gabor+SCA module is slightly inferior to the baseline+Gabor+SPA module, but it also has an improvement effect compared with baseline+Gabor. Using baseline+Gabor+ACSP+SPA modules, the network achieves a performance improvement of around 18.7% in Dice and 19.3% in F1 score comparing with U-Net. For the baseline+Gabor+SCA+ACSP module, the F1 score, the Dice and Hausdorff all increase a lot comparing with the performance of the U-Net. Moreover, when we integrate the two attention modules and the ACSP module, the performance of the three indicators are 82.9%, 88.1% and 152.37, respectively. To further illustrate the effectiveness of our proposed network, we have added two indicators, sensitivity (TPR) and specificity (SPC). All modules are combined together, which obtains the best performance with mean segmentation sensitivity of 0.853 and specificity of 0.239, because it successfully fuses and learns feature representation from Gabor-based convolution and CSBA module.

TABLE III
THE COMPARATIVE ANALYSIS OF MODELS ON THE GLAS CHALLENGE DATASET. S AND R REPRESENT SCORES AND RANK, RESPECTIVELY

Methods	F1 score				Obj. Dice				Obj. Hausdorff				Rank
	Test A		Test B		Test A		Test B		Test A		Test B		
	S	R	S	R	S	R	S	R	S	R	S	R	
CUMedVision2	0.912	4	0.716	9	0.897	5	0.781	11	45.2	4	160.35	13	5
ExB1	0.891	8	0.703	10	0.882	8	0.786	7	57.41	10	145.58	7	7
ExB3	0.896	5	0.719	8	0.886	6	0.765	13	57.36	9	159.87	12	9
Freidburg2	0.87	9	0.695	11	0.876	9	0.786	7	57.09	7	148.47	10	9
CUMedVision1	0.868	9	0.769	3	0.867	11	0.8	5	74.6	12	153.65	10	7
ExB2	0.892	7	0.686	13	0.884	7	0.754	14	54.79	6	187.44	15	12
Freidburg1	0.834	13	0.605	14	0.875	10	0.783	10	57.19	8	146.61	8	13
CVML	0.652	16	0.541	15	0.664	17	0.654	15	155.43	17	176.24	14	15
LiB	0.777	15	0.306	17	0.781	15	0.617	16	112.71	16	190.45	16	16
Vision4Glas	0.635	17	0.527	16	0.737	16	0.61	17	107.49	15	210.1	17	17
FCN-8 [13]	0.783	14	0.692	12	0.795	14	0.767	12	105.04	14	147.28	9	14
SegResNet [47]	0.862	11	0.764	5	0.873	11	0.814	4	61.37	11	117.54	4	5
DeepLabv3+ [35]	0.857	12	0.762	6	0.863	13	0.808	5	64.29	12	122.57	5	9
Xu et al. (2017) [48]	0.893	6	0.843	2	0.908	3	0.833	3	44.13	3	116.82	3	3
Micro-Net [21]	0.913	3	0.724	7	0.906	4	0.785	9	49.15	5	133.98	6	4
MILD-Net [22]	0.914	2	**0.844**	1	0.913	2	**0.836**	1	41.54	2	105.89	2	2
GCSBA-Net	**0.916**	1	0.832	3	**0.914**	1	0.834	2	**41.49**	1	**102.88**	1	1

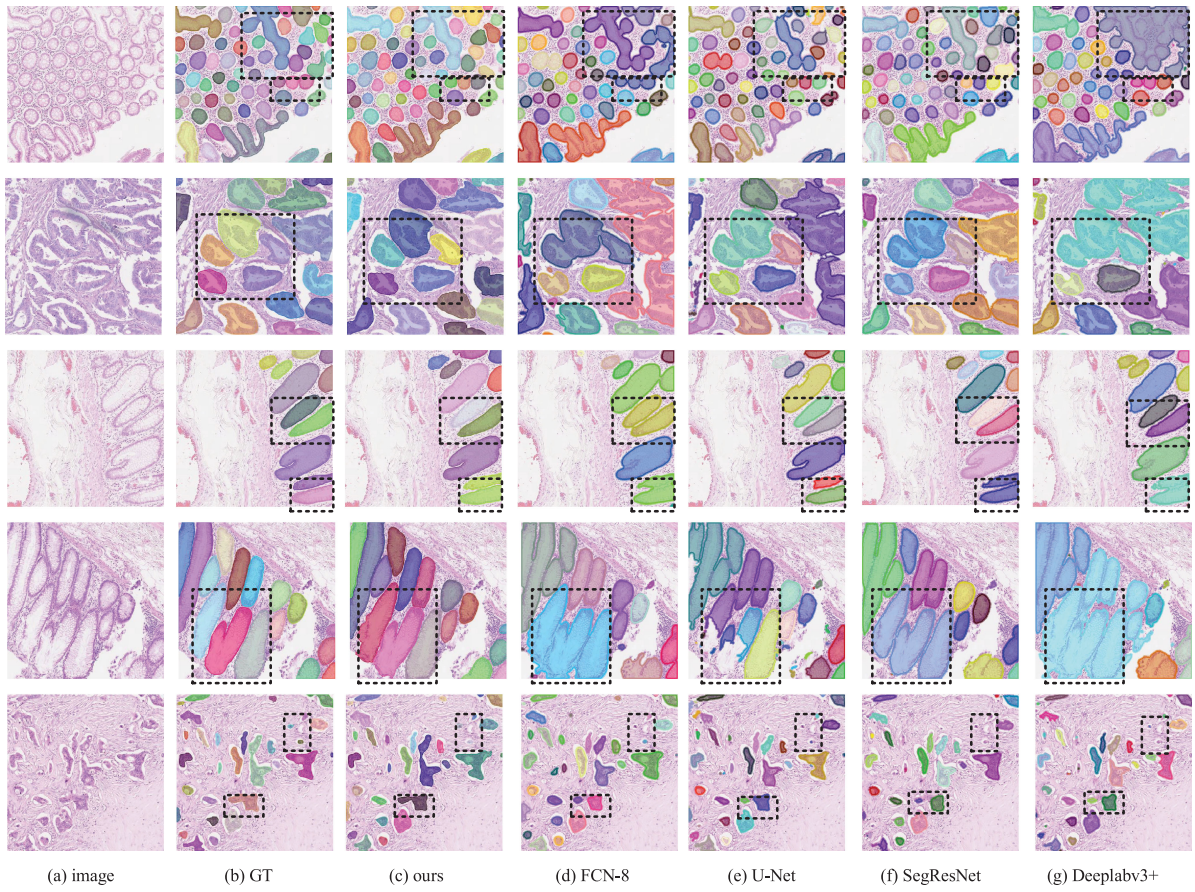


Fig. 6. The qualitative comparison of the proposed method with the four methods on CRAG dataset. Different colored glands represent different instances.

These experiments illustrate the importance of capturing dense and global contextual information. Table IV shows that the proposed architecture achieves the best performance.

C. The Ablation Study of Loss Function

In this section, we discuss the impact of different combinations of SSIM loss function, BCE loss function and $F_\beta - focal$ loss function on network performance. The first rows of Table V shows that the results of the method only based on $F_\beta - focal$ loss function is higher those of the methods only based on L_{ssim} or L_{bce}, which brings 0.9% and 0.02% improvement in F1 score, respectively. Dice index is increasing from 87.2% to 88.6% and from 88.1% to 88.6% comparing with the results of methods based on L_{ssim} and L_{bce}. Three kinds of loss functions are adopted alone, whose results are lower than the results based

TABLE IV
THE ABLATION STUDIES OF DIFFERENT MODULES ON THE CRAG DATASET ONLY WITH BCE LOSS FUNCTION

Methods	Configurations				Evaluation Metrics				
	Gabor	SPA	SCA	ACSP	$F_1 score$	Obj. Dice	Obj. Hausdorff	TPR	SPC
U-Net [3]					0.6	0.654	354.09	0.689	0.304
baseline	√				0.721	0.78	284.85	0.774	0.303
	√	√			0.741	0.806	177.65	0.768	0.309
	√		√		0.748	0.794	186.78	0.787	0.311
	√	√		√	0.793	0.841	169.24	0.813	0.295
	√		√	√	0.782	0.846	164.33	0.804	0.286
	√	√	√		**0.831**	0.861	172.41	0.837	0.257
		√	√	√	0.826	0.858	166.25	0. 843	**0.236**
	√	√	√	√	0.829	**0.881**	**152.37**	**0.853**	0.239

TABLE V
THE ABLATION STUDIES ON DIFFERENT LOSS FUNCTIONS

Configurations			Evaluation Metrics				
L_{ssim}	L_{bce}	$L_{F_\beta - focal}$	$F_1 score$	Obj. Dice	Obj. Hausdorff	TPR	SPC
√			0.822	0.872	148.43	0.842	0.231
	√		0.829	0.881	152.37	0.853	0.239
		√	0.831	0.886	150.16	0.857	0.228
√	√		0.830	0.889	151.36	0.861	0.223
	√	√	0.834	0.891	150.51	0.863	0.214
√	√	√	**0.836**	**0.894**	**146.77**	**0.866**	**0.212**

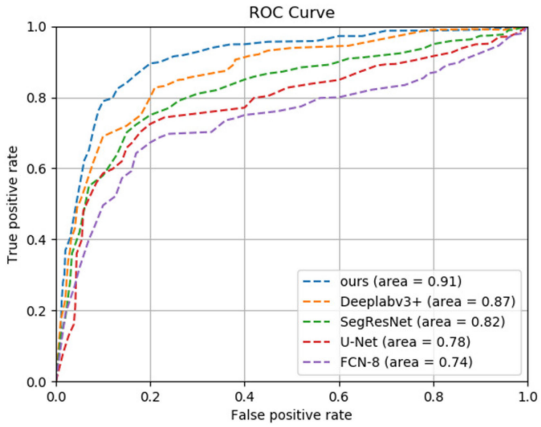


Fig. 7. ROC curves of different algorithms with the corresponding AUC values.

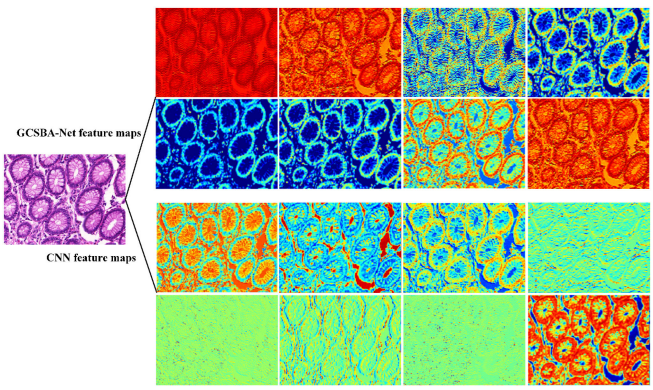


Fig. 8. The visualization of features extracted by the first convolution layer.

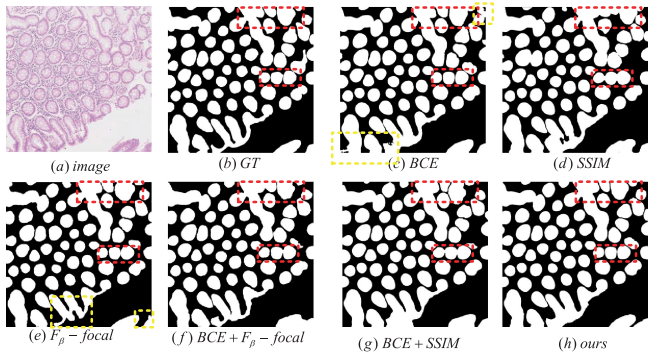


Fig. 9. The visualization of the ablation study on different loss functions.

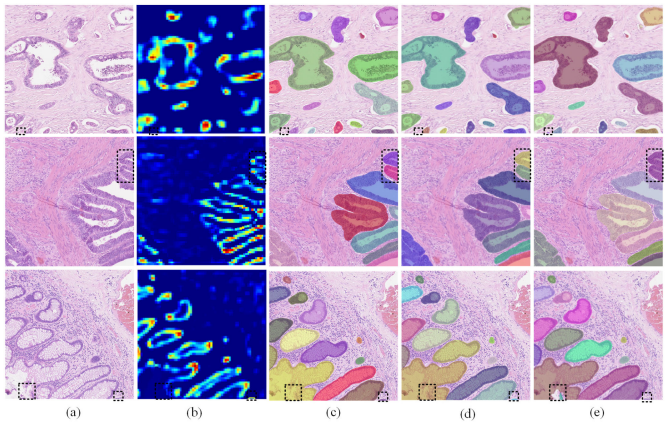


Fig. 10. The performance of the cascade bi-attention mechanism. (a) is the original image. (b) represents the cascade squeeze bi-attention map. (c) is the ground truth. (d) shows the segmentation results with the cascade squeeze bi-attention module. (e) illustrates the segmentation results without cascade squeeze bi-attention module.

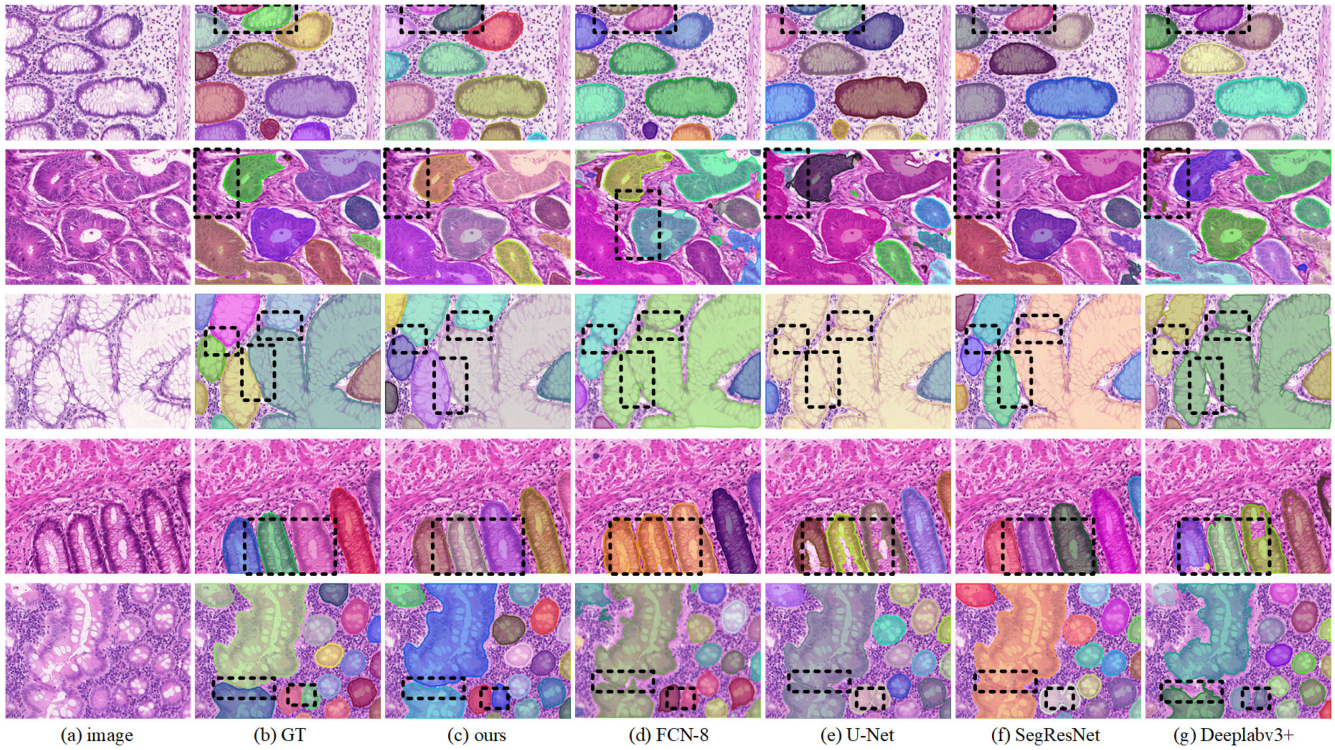


Fig. 11. The qualitative comparison of the proposed method with the four methods on GlaS dataset. Different colored glands represent different instances.

on the methods combining two of three loss functions. The performance adopting the novel hybrid loss function significantly improved F1 score 0.7%, dice index 1.3% and hausgorff 5.6 comparing with the performance of the L_{bce}-based method. In addition, TPR and SPC show the hybrid loss function still outperforms all other loss functions. The best TPR and SPC are 0.866, 0.212, respectively. As shown in Table V, this novel hybrid loss function greatly improves the performance of the network.

To further illustrate the effect of the proposed hybrid loss function, we show the qualitative segmentation results of several different loss function combinations in Fig. 9. In the yellow box in Fig. 9(c), L_{bce}-based method results in holes and mis-divisions. Although the L_{ssim}-based method improves its structural similarity, the cell boundaries are linked together in Fig. 9(d). Fig. 9(e) balances the recall and the accuracy, focusing on difficult samples, which leads to some degree of over-segmentation. This paper uses a combination of three loss functions. We can see that the proposed loss function achieves the best result.

D. Visualization of Cascade Squeeze Bi-Attention

To gain a deeper understanding of the CSBA module, the visualization results are given in Figure 10. As shown in Fig. 10(b), our cascade squeeze bi-attention map can capture semantic similarity and long-term dependencies, focusing on most of the lesion area. More importantly, it also pays attention to the cell boundary. For each input image, we mark the representative area with a black box. For example, in the first row, the black box in the original image is marked on a non-cancer area, and its map without cascade squeeze bi-attention identifies it as a cancer area shown in the fifth column. The second row guides attention focusing on the boundary of two cancer areas (Fig. 10(b)). GCSBA-Net captures the boundary features and separates them correctly. In the third row of Fig. 10(b), cascade squeeze bi-attention map finds small objects and suppresses false predictions. The cascade bi-attention module can finally gather richer and more intensive contextual information.

VI. Conclusion

In this paper, we propose a Gabor-based feature extraction and cascade squeeze bi-attention network for gland segmentation. The proposed GCSBA-Net uses the Gabor-based kernel to extract the texture information of the image to improve system performance. In addition, the proposed cascade squeeze bi-attention module enhances the ability to learn cell information at different scales. We also design a novel loss function to solve the problem of imbalanced data distribution and capture the structural similarity of gland segmentation. The experiments show that the proposed method achieves state-of-the-art result on the challenging CRAG dataset and the widely used MICCAI gland dataset. The performance indicates that our method has great potential for gland segmentation. In future work, we will optimize this approach and study its performance on a large-scale histopathology dataset.

APPENDIX
THE SYMBOLIC EXPLANATION.

Symbol	Symbolic explanation
$g_{\theta,v}(x,y)$	The Gabor wavelet transform function.
v	v is the scale.
θ	θ represents the orientations.
ψ	ψ is the phase offset.
σ	σ is the standard deviation of the Gaussian envelope.
C_i	C_i is a convolution filter.
L	L is a loss function.
$F_k(i,j)$	$F_k(i,j)$ is the $k-th$ layer feature.
$Q_k(i,j)$	$Q_k(i,j)$ is the Max pooling layer.
$K_k(i,j)$	$K_k(i,j)$ is the $k-th$ layer feature.
$k_{1\times1}$	The 1×1 convolution.
S_k, $S(i,j)$	S_k, $S(i,j)$ are position squeeze weights and channel squeeze weights, respectively.
λ_p, λ_c	The scale parameters.
TP_c	The true positives.
FN_c	The false negatives.
FP_c	The false positives.
g_{ic}	the ground truth for pixel i being class c.
g_i	The ground-truth label.
p_{ic}	The predicted probability for pixel i being class c.
p_i	The probability map.

REFERENCES

[1] J. Xu *et al.*, "Multi-tissue partitioning for whole slide images of colorectal cancer histopathology images with deeptissue net," in *Proc. 15th Eur. Congr. Digit. Pathol.*, 2019, pp. 100–108.

[2] M. Fleming, S. Ravula, S. F. Tatishchev, and H. L. Wang, "Colorectal carcinoma: Pathologic aspects," *J. Gastrointest. Oncol.*, vol. 3, no. 3, pp. 153–173, 2012.

[3] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*, 2015, pp. 234–241.

[4] F. Jun *et al.*, "Dual attention network for scene segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 3146–3154.

[5] U. Mohd, A. Belal, X. Wenjing, H. M. Shamim, and M. Ghulam, "Self-attention based recurrent convolutional neural network for disease prediction using healthcare data," *Comput. Methods Programs Biomed.*, vol. 190, 2019, Art. no. 105191.

[6] G. Huang, Z. Liu, L. Maaten, and K. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2261–2269.

[7] J. Jacobs, E. Panagiotaki, and D. Alexander, "Gleason grading of prostate tumours with max-margin conditional random fields," *Mach. Learn. Med. Imag.*, vol. 8679, pp. 85–92, 2014.

[8] K. Sirinukunwattana, D. Snead, and N. RajpootJacobs, "A novel texture descriptor for detection of glandular structures in colon histology images," *SPIE Med. Imag.*, vol. 9420, pp. 94200S–94200S2, 2015.

[9] L. T.-M., P. B., and M. G-H., "Computational texture features of dermoscopic images and their link to the descriptive terminology: A survey," *Comput. Methods Programs Biomed.*, vol. 182, 2019, Art. no. 105049.

[10] N. Ostu, "A threshold selection method from gray-histogram," *IEEE Trans. Syst., Man, Cybern.*, vol. 9, no. 1, pp. 62–66, Jan. 1979.

[11] K. Sirinukunwattana, D. Snead, and N. Rajpoot, "A stochastic polygons model for glandular structures in colon histology images," *IEEE Trans. Med. Imag.*, vol. 34, no. 11, pp. 2366–2378, Nov. 2015.

[12] A. Fakhrzadeh, E. Sporndly-Nees, L. Holm, and C. Hendriks, "Analyzing tubular tissue in histopathological thin sections," *Digit. Image Comput. Techn. Appl.*, pp. 1–6, 2012.

[13] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2017.

[14] F. Xing, Y. Xie, and L. Yang, "An automatic learning-based framework for robust nucleus segmentation," *IEEE Trans. Med. Imag.*, vol. 35, no. 2, pp. 550–566, Feb. 2016.

[15] K. Sirinukunwattana, S. Raza, Y.-W. Tsang, D. Snead, I. Cree, and N. Rajpoot, "Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1196–1206, May 2016.

[16] N. Kumar, R. Verma, S. Sharma, S. Bhargava, A. Vahadane, and A. Sethi, "A dataset and a technique for generalized nuclear segmentation for computational pathology," *IEEE Trans. Med. Imag.*, vol. 36, no. 7, pp. 1550–1560, Jul. 2017.

[17] M. Sapkota, X. Shi, F. Xing, and L. Yang, "Deep convolutional hashing for low dimensional binary embedding of histopathological images," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 2, pp. 805–816, Mar. 2019.

[18] C. Hao, X. Qi, J.-Z. Cheng, and P.-A. Heng, "Deep contextual networks for neuronal structure segmentation," in *Proc. 30th AAAI Conf. Artif. Intell.*, 2016, pp. 1167–1173.

[19] H. Chen, X. Qi, L. Yu, and P.-A. Heng, "DCAN: Deep contour-aware networks for accurate gland segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 2487–2496.

[20] K. Sirinukunwattana, J. Pluim, C. Hao, X. Qi, and N. Rajpoot, "Gland segmentation in colon histology images: The glas challenge contest," *Med. Image Anal.*, vol. 35, pp. 489–502, 2017.

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

[22] S. Graham, H. Chen, Q. Dou, P. Heng, and N. Rajpoot, "Mild-Net: Minimal information loss dilated network for gland instance segmentation in colon histology images," *Med. Image Anal.*, vol. 52, pp. 199–211, 2019.

[23] D. Gabor, "Theory of communication. part 1: The analysis of information," *J. Institution Elect. Engineers - Part III: Radio Commun. Eng.*, vol. 93, pp. 429–441, 1946.

[24] A. Kinnikar, M. Husain, and S. Meena, "Face recognition using gabor filter and convolutional neural network," in *Proc. Int. Conf. Informat. Anal.*, 2016, pp. 1–4.

[25] A. Calderón, S. R. Ovalle, and J. Victorino, "Handwritten digit recognition using convolutional neural networks and gabor filters," in *Proc. Int. Congr. Comput. Intell.*, 2003, pp. 429–441.

[26] S. Luan, C. Chen, B. Zhang, J. Han, and J. Liu, "Gabor convolutional networks," *IEEE Trans. Image Process.*, vol. 27, no. 9, pp. 4357–4366, Sep. 2018.
[27] J. Yoo, Y. Uh, S. Chun, B. Kang, and J.-W. Ha, "Photorealistic style transfer via wavelet transforms," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2019, pp. 9035–9044.
[28] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7794–7803.
[29] Z. Huang, X. Wang, L. Huang, C. Huang, Y. Wei, and W. Liu, "Ccnet: Criss-cross attention for semantic segmentation," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2019, pp. 603–612.
[30] B. Chen, J. Li, G. Lu, and D. Zhang, "Lesion location attention guided network for multi-label thoracic disease classification in chest x-rays," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 7, pp. 2016–2027, Jul. 2020.
[31] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7132–7141.
[32] H. Zhao *et al.*, "Psanet: Point-wise spatial attention network for scene parsing," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 270–286.
[33] Q. Wang, Y. Zheng, G. yang, W. Jin, X. Chen, and Y. yin, "Multi-scale rotation-invariant convolutional neural networks for lung texture classification," *IEEE J. Biomed. Health Informat.*, vol. 22, no. 1, pp. 184–195, Jan. 2018.
[34] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2881–2890.
[35] L. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Adam. encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 801–818.
[36] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-resNet and the impact of residual connections on learning," in *Proc. 31st AAAI Conf. Artif. Intell.*, 2017, pp. 4278–4284.
[37] Z. Gu *et al.*, "CE-Net: Context encoder network for 2 D medical image segmentation," *IEEE Trans. Med. Imag.*, vol. 38, no. 10, pp. 2281–2292, Oct. 2019.
[38] S. Zhao, Y. Dong, E. I.-C. Chang, and Y. Xu, "Recursive cascaded networks for unsupervised medical image registration," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2019, pp. 10599–10609.
[39] P. Boerr, D. Kroese, and S. Mannor, "A tutorial on the cross-entropy method," *Ann. Operations Res.*, vol. 134, no. 1, pp. 19–67, 2005.
[40] Z. Wang, E. Simoncelli, and A. Bovik, "Multi-scale structural similarity for image quality assessment," in *Proc. Asilomar Conf. Signals Syst. Comput.*, 2002, pp. 1398–1402.
[41] T. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 318–327, Feb. 2020.
[42] W. Zhu *et al.*, "AnatomyNet: Deep 3 D squeeze-and-excitationu-nets for fast and fully automated whole-volume anatomical segmentation," *bioRxiv*, pp. 1–14, 2018.
[43] K. Wong, M. Moradi, H. Tang, and M. Syeda, "3D segmentation with exponential logarithmic loss for highly unbalanced object sizes," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*, 2018, pp. 612–619.
[44] N. Abraham and N. M. Khan, "A novel focal tversky loss function with improved attention U-net for lesion segmentation," in *Proc. IEEE 16th Int. Symp. Biomed. Imag.*, 2019, pp. 1–4.
[45] Q. Xuebin, Z. Zichen, H. Chenyang, G. Chao, D. Masood, and J. Martin, "BasNet: Boundary-aware salient object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 7471–7481.
[46] P. Hashemi and R. Abbaspour, "Assessment of logical consistency in OpenStreetMap based on the spatial similarity concept, in *OpenStreetMap in GIScience*, 2015, pp. 19–36.
[47] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for scene segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.
[48] Y. Xu *et al.*, "Gland instance segmentation using deepmultichannel neural networks," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 12, pp. 2901–2912, Dec. 2017.
[49] A. Paszke *et al.*, "Automatic differentiation in pytorch," in *Proc. 31st Conf. Neural Informat. Process. Syst.*, 2017, pp. 1–4.
[50] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Representations*, 2015, pp. 1–15.