

Flexible-CLmser: Regularized Feedback Connections for Biomedical Image Segmentation

Boheng Cao¹, Shikui Tu^{1,*}, Lei Xu^{1,2,*}

¹Department of Computer Science and Engineering

Center for Cognitive Machines and Computational Health (CMaCH)

Shanghai Jiao Tong University, Shanghai, China. 200240

²Guangdong Institute of Intelligence Science and Technology, Hengqin, Zhuhai, China.

{c451735634,tushikui,leixu}@sjtu.edu.cn; *Corresponding Author

Abstract—The skip connections in U-Net pass features from the levels of encoder to the ones of decoder in a symmetrical way, which makes U-Net and its variants become state-of-the-art approaches for biomedical image segmentation. However, the U-Net skip connections are unidirectional without considering feedback from the decoder, while this paper exploits the feedback information to refine the segmentation. We develop a deep bidirectional network based on the least mean square error reconstruction (Lmser) self-organizing network, an early model that folds an autoencoder along the central hidden layer such that the neurons on the paired layers between encoder and decoder merge into one, equivalently forming bidirectional skip connections between encoder and decoder. We find that the feedback links indeed increase the segmentation accuracy, but may also bring certain noise into the segmentation. To tackle this problem, we present a gating and masking mechanism on the feedback connections to filter the irrelevant information. Experimental results on MoNuSeg, TNBC, and EM membrane datasets demonstrate that our method are robust and outperforms state-of-the-art methods.

Index Terms—image segmentation, recurrent/feedback connections, Lmser, U-Net

I. INTRODUCTION

Biomedical image segmentation has wide applications in biomedical research, clinical diagnosis, health care, and so on. For example, nuclei segmentation is an important first step of analyzing digitized microscopy images. However, it is time-consuming and error-prone to segment the biomedical images by human, driving the demands of methods for automatic segmentation.

Existing state-of-the-art models are mostly built from Convolutional Neural Networks (CNN) which have been used to tackle the problem of automatic biomedical image segmentation, and they outperformed the traditional methods. One of the most popular methods is U-Net [1]. Its architecture was U-shaped, and it was built on fully convolutional networks (FCN) [2]. The U-Net consists of a contracting path as encoder to extract features and an expansive path as decoder to perform segmentation. The feature map from each of the layer of the contracting path is concatenated with the symmetrically corresponding layer in the expansive path. Such skip connections enable U-Net to reserve the hierarchical features in encoder

and enhance the ability of decoder to segment the images. Later, many works were built upon a U-shape architecture with different network backbones and modules, such as U-Net++ [3] and Attention U-Net [4].

The U-Net and its variants benefit from the forward skip connections from encoder to decoder, but these methods still fall into the feed-forward principle of computation, i.e., extracting image features by a feed-forward pass through stacks of convolutional layers. They ignore the feedback information from decoder to encoder, which may be used to enhance the representation learning and refine the segmentation results. It is noted that the brain achieves robust visual perception by using not only feed-forward procedure but also feedback mechanisms. In recent years, research efforts were made from this aspect, and feedback mechanisms played an important role in tasks such as image recognition [5], salient object detection [6] and image segmentation [7], [8].

Recently, one early method called Least mean square error reconstruction (Lmser) self-organizing network [9], [10] was revisited and developed into a deep CNN based Lmser (CLmser) for image related tasks [11]. Lmser was a further development of autoencoder (AE) by folding along the central hidden layer. Such folding makes the neurons on the paired layers between encoder and decoder merge into one, which equivalently induces bidirectional skip connections between them. It was confirmed in [11] that deep Lmser learning works well on several potential functions previously speculated and addressed in [9], such as image reconstruction, association recall, and so on. It is also demonstrated to be effective in image inpainting [12] and super-resolution [13]. Moreover, gating mechanism was imposed on the forward skip connections of CLmser in [14], showing better performance in biomedical image segmentation than U-Net models. But, how the backward connections in CLmser affect the segmentation results is still not investigated.

In this paper, we proceed with CLmser by considering not only forward skip connections from encoder to decoder but also backward connections from decoder to encoder. Following the recurrence induced by the feedback links, we unroll CLmser as multiple copies of the same network, each passing feature maps to a successor in a sequential manner. We demonstrate that the feedback connections are able to

improve the segmentation accuracy by effectively guiding the representation learning in encoder. Moreover, we find that the feedback links may also bring unwanted, noisy patterns into the segmentation results after multiple recurrence times. To tackle this issue, we propose a gating and masking mechanism on feedback connections to filter the irrelevant features. Experimental results on MoNuSeg, TNBC and EM membrane dataset show that gating mechanisms and masks can further improve the segmentation performance.

II. RELATED WORKS

A. FCN for Biomedical Image Segmentation

Fully convolutional networks (FCN) [2] was proposed by replacing all fully-connected layers by convolutional layers. FCN could preserve spatial information and accept any size image as input. To deal with biomedical image segmentation, U-Net [1] further expanded FCN by using an encoder-decoder structure and adding skip connections from encoder to decoder. The skip connections can preserve various levels of features in encoder and are then combined with ones in decoder to get precise segmentation results. Later, many works were proposed to deal with biomedical image segmentation based on U-Net. U-Net++ [3] expanded the skip connections in U-Net structure, using dense skip connections to obtain finer-grained segmentation details. Attention U-Net [4] applied attention mechanism on the features of skip connection and decoder. Besides, F-CLmser-S network [14] considered a gating mechanism on skip connections from encoder to decoder, based on the flexible Lmser network [9], [15].

B. Feedback Connections in Neural Networks

Motivated by the feedback mechanisms in the perception process of the brain, many works about feedback connections in neural networks have been developed. Predictive coding network [5] used bidirectional connections to build its model. It took images as input and ran recursive cycles of bottom-up and top-down computation to update the internal representations. In [6], attentive feedback modules were built to guide the message passing among encoders and corresponding decoders.

Moreover, BiO-Net [7] extended U-Net by deploying paired forward and backward skip connections between encoder and decoder blocks, and achieved better segmentation performance than U-Net and its variants without introducing extra trainable parameters. BiX-Net [8] further used two-phase neural architecture search algorithm to search for bidirectional multi-scale skip connections between encoder and decoder, which did not improve the accuracy significantly but shrank the network size. Our method is closely related to BiO-Net, but different in the following aspects. First, we fuse the features that are passed through the skip connections with the target layers in a weighted summation, instead of concatenation in BiO-Net. Our fusion strategy is more efficient to merge the features with 14.6% fewer parameters than BiO-Net. Second, we filter the noisy patterns in the feedback connections by a gating and masking mechanism, while BiO-Net may suffer from the noisy intervention brought by the feedback links.

III. METHODS

A. Overview of the Proposed Network

An overview of the proposed model is given in Fig. 1. It consists of a *pre-transform* block, *encoder-decoder* module, *post-transform* block. The pre-transform block extracts the raw input of $3 \times 512 \times 512$ to a feature map of $32 \times 256 \times 256$, while the post-transform concatenates all the recurrent output and calculates a segmentation result of 512×512 . The encoder-decoder module adopts an CLmser architecture [12], with bidirectional connections between encoders and decoders. The network backbones are similar to the ones in BiO-Net [7]. Different from BiO-Net, we abandon the fusion strategy of concatenating the features transferred through the skip connections with the ones in the receiver layer, and compute the fusion in a weighted summation. The weighted-sum fusion is able to control the information transfer from the two sides, keep the fusion output of the same size as the summands, and increase the learning efficiency using a fewer network parameters than BiO-Net. Moreover, we propose a channel gating and pixel-level masking to control the information passed by feedback connections. In channel level, we use attention mechanism to compute the importance vector of channels, and thus give weights to each channel. In pixel level, the segmentation result of last recurrence time is reshaped and used as a mask to filter the feature maps. With the gating and masking, the irrelevant information are blocked and the feedback is guided to positively refine the segmentation result.

In practice, all the convolutional kernels are set to 3×3 with stride size as 1 and padding size as 1. All the deconvolutional kernels are set to 3×3 with stride size as 2 and padding size as 1. Each convolutional layer or deconvolutional layer is followed with a ReLU activation layer and a batch normalization layer, except for the final convolutional layer in post transform block, which is only followed with a sigmoid layer to obtain the output.

B. Unroll the Recurrent Structure

The backward connections from decoder to encoder in Fig. 1 introduce recurrence to the learning process. It has been theoretically proved that the process will converge into an equilibrium state [10]. Here, we follow [7], [12] to unroll the encoder-decoder module as the same copies of the block in a sequential manner, each passing the feature maps to a successor. The unrolled structure is given in Fig. 2. Feature maps will be first extracted from the raw image by the pre-transform block and then sent to the encoder-decoder recurrent loop. The bidirectional connections are decoupled into forward links first and then backward links later with a time-delay. The convolutional kernels are reused in the loop. After several recurrence times, the final segmentation result is calculated by the post-transform block, which takes as input the concatenation of all decoder outputs during the recurrent process. Not like the U-Net based networks such as BiO-Net,

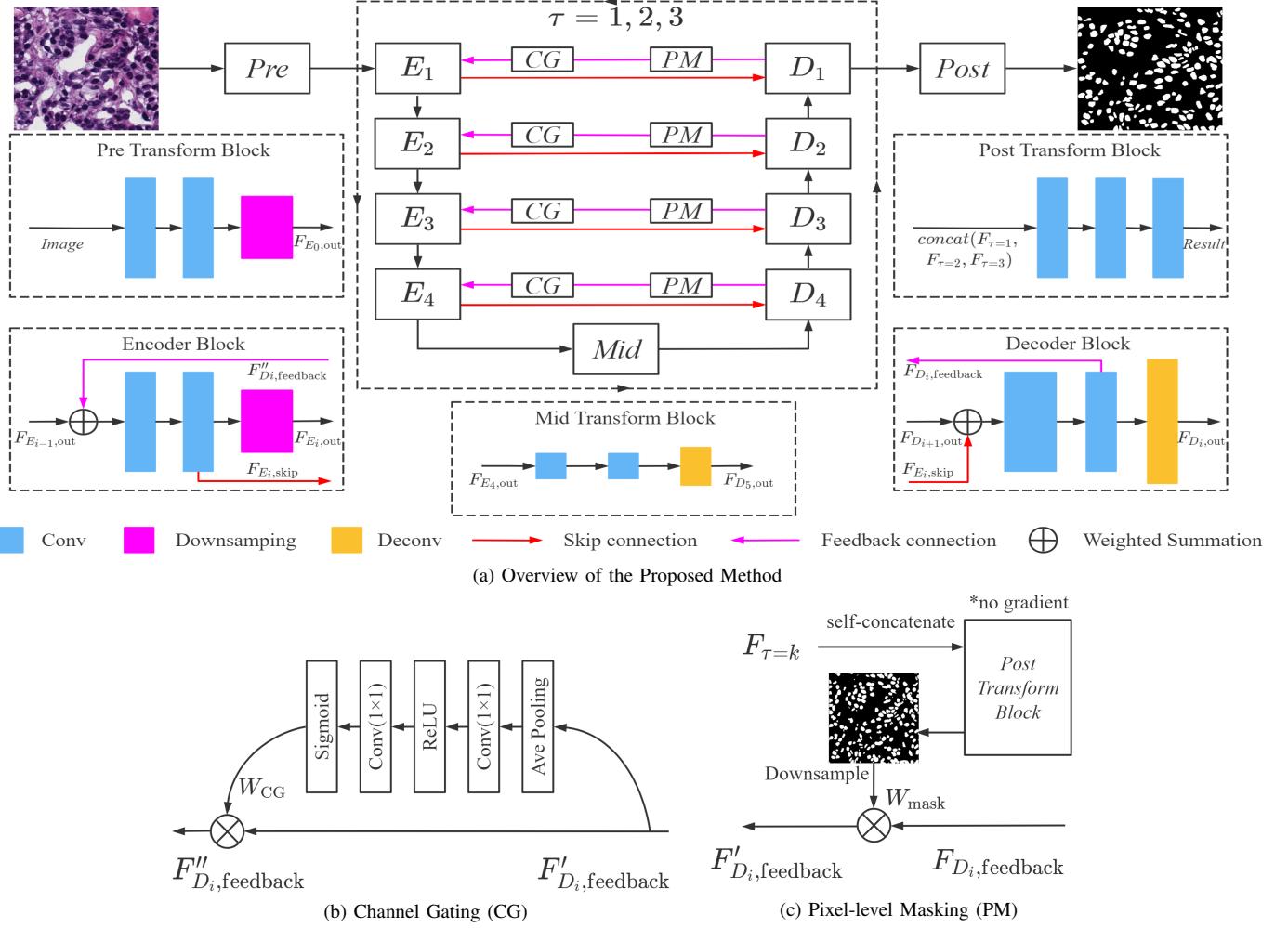


Fig. 1: Overview of the proposed model. (a) The overall network structure. A Downsampling block contains a convolutional layer and a max-pooling layer. Each skip or feedback connection comes from the second convolutional layer of a block and is finally fused by a weighted summation. Activation layers and batch normalization layers are not shown. (b) The computation of Channel Gating (CG). (c) The process of Pixel-level Masking (PM).

we compute the feature fusion in a weighted summation as follows:

$$F_{D_i,in} = \alpha F_{D_{i+1},out} + (1 - \alpha) concat(F_{E_i,skip}, F_{E_i,skip}), \quad (1)$$

$$F_{E_i,in} = \alpha F_{E_{i-1},out} + (1 - \alpha) F''_{D_i,feedback}, \quad (2)$$

where E_i denotes the i_{th} encoder block and D_i denotes the i_{th} decoder block. The details of E_i and D_i are provided in Fig. 1a. $F_{E_i,skip}$ denotes the feature map passed through the skip connection. It appears two times in Eq.(1) because it needs to be self-concatenated to align the same channel number to that of $F_{D_{i+1},out}$. $F''_{D_i,feedback}$ represents the feedback connection after gating and masking. Note that $F_{E_0,out}$ and $F_{D_5,out}$ denote the output of pre-transform block and mid-transform block respectively. α is a hyper-parameter to be set according to experience. Compared with traditional concatenation, our fusion method can save parameters of model and control the weight of connections.

C. Regularizing Feedback Connections by Channels and by Pixels

The feedback information is able to guide the model to pay attention to the key features that are relevant to the segmentation. However, the recurrent loop may also introduce noisy patterns that would disturb the representation learning by the encoder. Thus, the feedback connections need to be properly regularized. We propose a pixel-level masking and channel gating to filter the feature maps and suppress the noise. Pixel-level masking aims to filter essential spatial patterns while channel gating is to capture the inter-dependencies of different channels.

The pixel-level masking and channel gating are sequentially computed as below:

$$F'_{D_i,feedback} = W_{mask} \otimes F_{D_i,feedback}, \quad (3)$$

$$F''_{D_i,feedback} = W_{CG} \otimes F'_{D_i,feedback}, \quad (4)$$

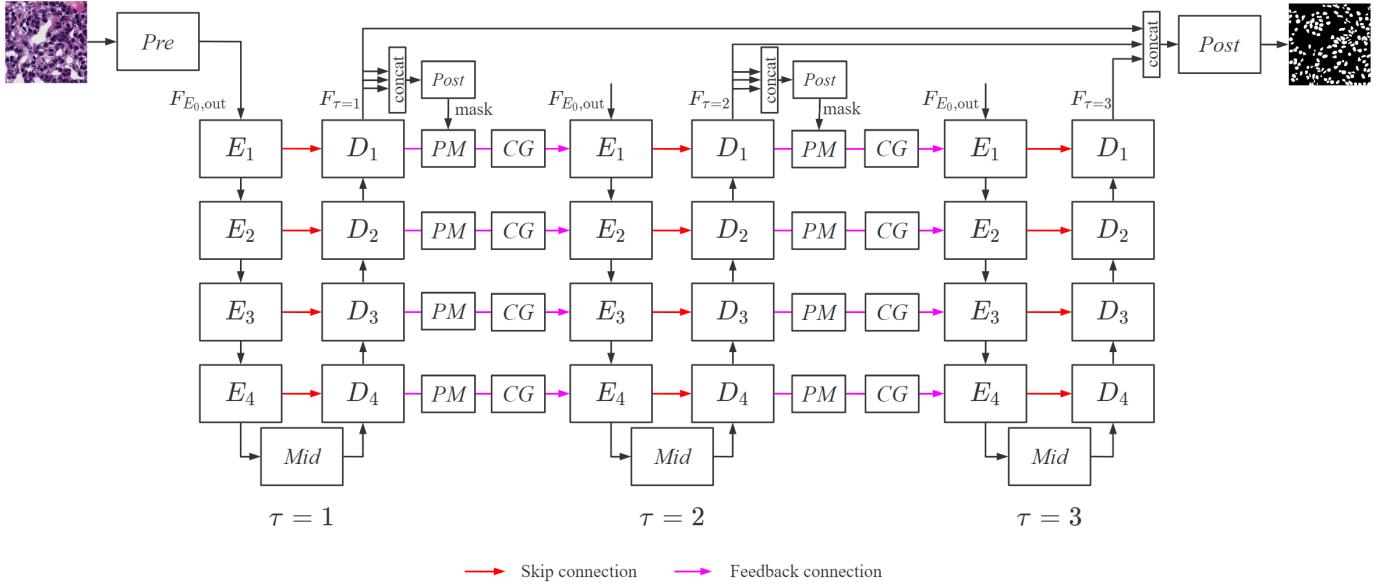


Fig. 2: Unfolded view of the proposed model, where $\tau = 1, \dots, t$ denotes the recurrence time, t represents the total number of recurrence times. Here, we set $t = 3$.

where the feature map $F_{D_i, \text{feedback}} \in \mathbb{R}^{C \times H \times W}$ is the output of the second layer from a decoder block, \otimes indicates the scaling operator along the pixel or channel coordinates, C, H, W denote channel number, height, and width, respectively. The mask matrix $W_{\text{mask}} \in \mathbb{R}^{1 \times H \times W}$ is calculated by downsampling the segmentation result of the last recurrence time. The channel gating matrix $W_{\text{CG}} \in \mathbb{R}^{C \times 1 \times 1}$ is computed by the feature map $F'_{D_i, \text{feedback}}$ itself.

Pixel-level masking aims to filter the feature maps by pixels. Original feature maps contain features of all regions in the image. However, features of the background regions might become noises when sent back to encoders. As given in Fig. 1c, the pixel-level masking is generated by downsampling the segmentation result of last recurrence time, which is calculated by a post-transform block. Here, the block uses duplication of the decoding result as input to fit the kernel size. The mask indicates the regions that contains the object to be segmented and filters the regions of background where might induce noisy patterns.

The channel gating module is constructed in the same way as [16]. As illustrated in Fig. 1b, we first apply average pooling on the feature map $F'_{D_i, \text{feedback}}$ to get a feature vector. Each scalar in vector can be a static representation of the corresponding channel. Then, two 1×1 convolutional layers are used to compute the weights for filtering information along the channel dimension. After channel gating, the number of channels is kept unchanged.

D. Loss Function

We adopt the binary cross entropy loss as the loss function:

$$L = -\frac{1}{h \times w} \sum_{i,j} E_{i,j}, \quad (5)$$

$$E_{i,j} = Y_{i,j} \log(X_{i,j}) + (1 - Y_{i,j}) \log(1 - X_{i,j}), \quad (6)$$

where $X_{i,j} \in \mathbb{R}^{h \times w}$ and $Y_{i,j} \in \mathbb{R}^{h \times w}$ are network prediction and ground truth respectively, h and w are the height and width of the images.

IV. EXPERIMENTS AND RESULTS

A. Datasets

Our method was evaluated on two nuclei segmentation datasets, MoNuSeg [17] and TNBC [18], and one dataset of Mouse Piriform Cortex EM images. The original MoNuSeg dataset contains a training set of 30 images and a testing set of 14 images. The image size is 1000^2 , which is sampled from different slides of organs. TNBC dataset consists of 50 histopathology images without specifying any training or testing set. The image size is 512^2 . Following [7], we used MoNuSeg training set as training set, while evaluating on both MoNuSeg testing set and the whole TNBC dataset. The MoNuSeg dataset was first enlarged by extracting 512^2 patches from 4 corners of each image to make the image size consisted with TNBC dataset. The dataset of Mouse Piriform Cortex EM images was collected by [19]. We used stack1 with image size of 255^2 as training set and stack4 with image size of 256^2 as testing set. The training set was first resized to 256^2 by zero-padding.

B. Implementation Details

To enrich the training data, data augmentation was applied on the training set by random rotation, shifting, shearing, zooming and flipping. In the training process, Adam [20] optimizer was used to optimize cross entropy loss. Initial learning rate and decay rate were set to 0.01 and 0.00003. The maximum epoch number was set to 300 and the batch

TABLE I: Experiment Results on MoNuSeg and TNBC dataset

Methods	MoNuSeg		TNBC		#params	model size
	IoU	DICE	IoU	DICE		
U-Net [1] w. ResNet-18 [26]	0.684	0.810	0.459	0.603	15.56 M	62.9 MB
U-Net++ [3] w. ResNet-18 [26]	0.683	0.811	0.526	0.652	18.27 M	74.0 MB
U-Net++ [3] w. ResNet-50 [26]	0.695	0.818	0.542	0.674	37.70 M	151.9 MB
Micro-Net [22]	0.696	0.819	0.544	0.701	14.26 M	57.4 MB
Naylor <i>et al.</i> [18]	0.690	0.816	0.482	0.623	36.63 M	146.7 MB
M-Net [21]	0.686	0.813	0.450	0.569	0.6 M	2.7 MB
Att U-Net [4]	0.678	0.810	0.581	0.717	33.04 M	133.2 MB
R2U-Net, t=2 [23]	0.678	0.807	0.532	0.650	37.02 M	149.2 MB
R2U-Net, t=3 [23]	0.683	0.815	0.590	0.711	37.02 M	149.2 MB
LinkNet [24]	0.625	0.767	0.535	0.682	11.54 M	139.4 MB
BiO-LinkNet, t=2 [7], [24]	0.621	0.766	0.541	0.690	11.54 M	139.4 MB
BiO-LinkNet, t=3 [7], [24]	0.634	0.774	0.571	0.716	11.54 M	139.4 MB
BiO-Net, t=1 [7]	0.680	0.803	0.456	0.608	15.0 M	60.6 MB
BiO-Net, t=2 [7]	0.694	0.816	0.548	0.693	15.0 M	60.6 MB
BiO-Net, t=3 [7]	0.700	0.821	0.618	0.751	15.0 M	60.6 MB
BiO-Net, t=3, INT [7]	0.704	0.825	0.651	0.780	15.0 M	60.6 MB
BiX-Net [8]	0.699	0.822	0.680	0.808	0.38 M	1.69 MB
F-CLmser-S [14]	0.682	0.810	0.645	0.781	8.35 M	31.93 MB
Flexible-CLmser (Ours)	0.712	0.831	0.693	0.816	13.47 M	51.7 MB

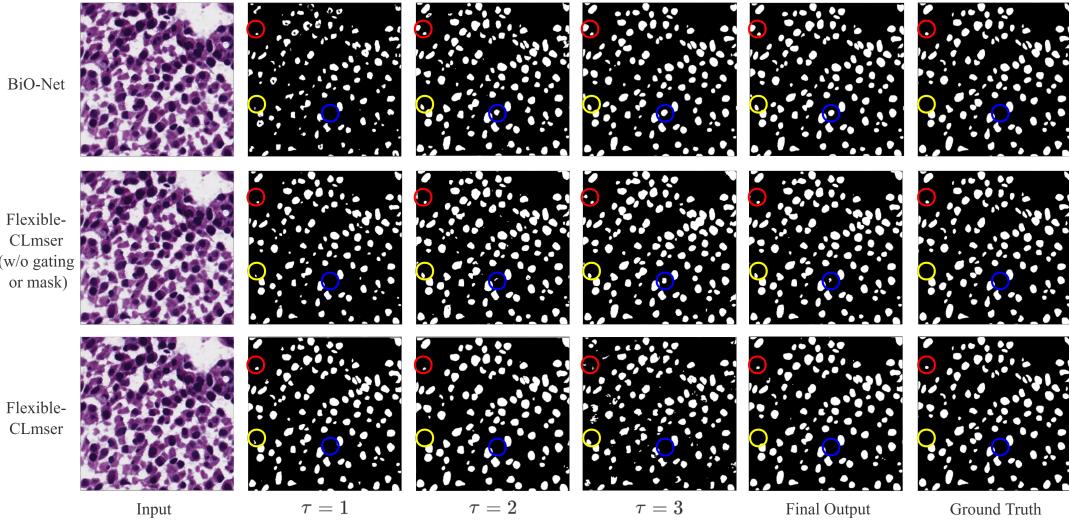


Fig. 3: Segmentation results of an example by three models. $\tau = 1, 2, 3$ depicts the segmentation results calculated by self-concatenation of the output in each recurrence time, which is the same with our pixel-level masking calculation.

size was set to 2. The hyper-parameter α was set to 0.4. All experiments of our model were conducted on a single Titan XP GPU with pytorch.

C. Results and Analysis

We compare the proposed method with state-of-the-art models on MoNuSeg and TNBC dataset, including U-Net models [1], [3], [4], [18], [21], [22], recurrent methods [23], and bidirectional methods [7], [8], [24]. We use IoU (Intersection over Union) and DICE (Dice coefficient) to evaluate the methods. The results are reported in Table I, where the results of baselines are from [7] and [8]. Our model outperforms the others on both MoNuSeg and TNBC testing set. Moreover, due to the fusion strategy by weighted summation, the channels of feature maps are reduced, thus the size of convolutional kernels

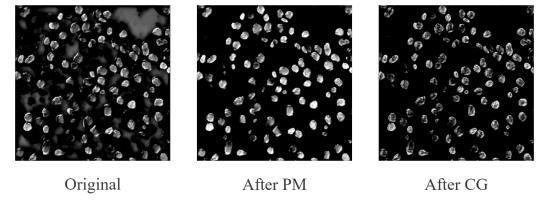


Fig. 4: Visualizations of a feedback connection in our full model, which are obtained by average pooling through the channel dimension. The images are contrast enhanced by rescaling the gray value from (30, 80) to (0, 255).

are decreased. As a result, number of parameters and size of our model are also smaller than most of the previous works.

TABLE II: Experiment results in EM membrane segmentation

Methods	Rand F-score	#params
U-Net	0.821	15.56 M
Att U-Net	0.833	33.04 M
U-Net++	0.844	18.27 M
BiO-Net, t=1	0.827	15.0 M
BiO-Net, t=2	0.871	15.0 M
BiO-Net, t=3	0.887	15.0 M
F-CLmser-S	0.858	8.35 M
Flexible-CLmser	0.903	13.47 M

We demonstrate by examples how the gating and masking mechanism works to reduce the noise which may be introduced through the feedback connections in Fig. 3. BiO-Net and the no-gating version of the proposed flexible CLmser are included as baselines for comparisons. All three models use recurrent structure and feedback connections. We visualize not only the final segmentation results but also the ones that are calculated by self-concatenation of the output for each recurrence time. We use circles to highlight the differences between the results of the models. Feedback connections can indeed help increase the accuracy, but it will bring some noise. In the circled part, the BiO-Net and the no-gating CLmser give the correct segmentation in the first recurrence time. Then, they are affected by the noise from the feedback connections, and finally output wrong segmentation. In the red and yellow circles, even the full CLmser model suffers from the noise and output incorrect labels at some pixels, especially in the third recurrence time. However, our gating mechanisms and masks can filter the noises and irrelevant information in feedback connections and maintain the loss of accuracy at a very low level. Therefore, our model is able to obtain the most accurate segmentation results.

Furthermore, we visualizes a feedback connection in our full model in Fig. 4. The original feedback connection passes many noisy patterns from the regions of background. After the pixel-level masking, the noises are filtered and only the useful features are retained. The channel gating makes the segmentation boundary more clear.

To verify the robustness of our model, we also evaluate the performance on segmentation task of Mouse Piriform Cortex EM images. We use Rand F-score [25] to evaluate the segmentation results. The results are reported in Table II, where the results of baselines are from [7]. Again, our model performs better than the existing methods.

D. Ablation Study

We study the effects of the pixel-level masking and the channel gating. The results are reported in Table III. First, even without any gating or mask, our model still achieve better or at least comparable results with previous works, which shows the superiority of Lmser structure. Second, the correct use of gating mechanisms and masks can further improve the performance of the model. Moreover, the channel gating only adds a few parameters to the model while pixel-level masking does not bring any new parameters. We also evaluate the effect

TABLE III: Investigations on different gating and masking in feedback connections. PM represents the pixel-level masking; CG represents the channel gating; w/o means "without"; w/PM (w/CG) means PM (CG) are applied on feedback connections; w/CG+PM means computing CG first and PM later.

Gating and masking	MoNuSeg		TNBC		#params
	IoU	DICE	IoU	DICE	
BiO-Net, t=3, INT [7]	0.704	0.825	0.651	0.780	15.0 M
w/o gating or mask	0.707	0.828	0.660	0.793	13.43 M
w/PM	0.704	0.826	0.665	0.797	13.43 M
w/CG	0.707	0.827	0.668	0.797	13.47 M
w/CG + PM	0.705	0.826	0.676	0.805	13.47 M
w/PM + CG	0.712	0.831	0.693	0.816	13.47 M

TABLE IV: Segmentation performance under different number of total recurrence times

#total recurrence time	MoNuSeg		TNBC		#params
	IoU	DICE	IoU	DICE	
t=1 ^a	0.697	0.821	0.654	0.787	13.44 M
t=2	0.700	0.823	0.689	0.814	13.46 M
t=3	0.712	0.831	0.693	0.816	13.47 M
t=4	0.707	0.828	0.690	0.815	13.49 M
t=5	0.674	0.804	0.676	0.805	13.51 M

^aWhen t=1, there is no feedback connections and no gating or masking.

of further regularizing feed-forward connections like [14], but there is no significant improvement on segmentation results.

We also investigate the role of the total number of recurrence times. Previous works usually set the maximum number of recurrence time as 3 [7], [23]. In [7], the reason was attributed to the GPU limitation. However, as in Fig. 3, feedback connections may have side effects when the recurrent loop grows. As a result, increasing the number of recurrence times may not improve segmentation results. We report the experimental results of the proposed network for different number of total recurrence times on MoNuSeg and TNBC dataset in Table IV. When the number of total recurrence times increases from 3 to 4, 5, the results on both two datasets get worse. An example of the segmentation results with the total recurrence time being 5 is provided in Fig. 5, where the noise is obvious at $\tau = 5$. The observation indicates that the gating and masking mechanism may not work well if the recurrence loop continues by too many cycles. It suggest that there is still a limitation of the proposed regulation on the feedback connections, which needs further investigations in the future. We also evaluate BiO-Net [7] with total recurrence time as 4 and 5, and the results get even worse. In conclusion, the noises in feedback connections become out of control when recurrence loop keeps continuing, thus the total loop number should be determined carefully according to experiments.

V. CONCLUSION

We have proposed a flexible-CLmser model for the problem of biomedical image segmentation. We exploit the feedback information to improve the segmentation in a recursive manner. The features passed by the forward and backward connections

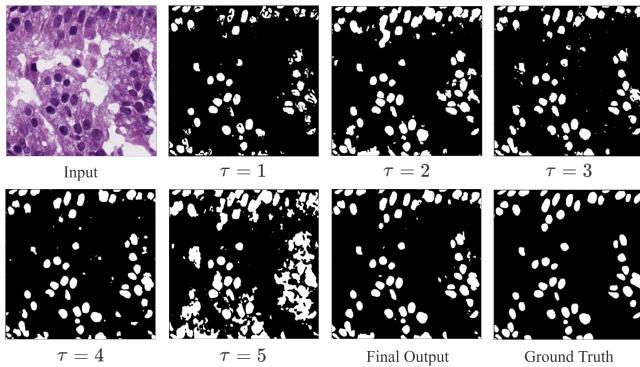


Fig. 5: Segmentation results of an example image by Flexible-CLmser with total recurrence times as 5. $\tau = 1, 2, 3, 4, 5$ depicts the segmentation results calculated by self-concatenation of the output at the τ -th recurrence time, which is the same with our pixel-level masking calculation.

between encoder and decoder are fused with the target layer in a weighted summation, making the network to transfer the information in an efficient and economic way. We find that the feedback information is able to improve the segmentation results but it may also bring noisy patterns if the recurrence times grow large. We impose channel gating and pixel-level masking on feedback connections to filter the irrelevant patterns. Experimental results on nuclei segmentation task and EM membrane segmentation task show that our method outperforms the state-of-the-art models. The visualization results and ablation study indicate that the channel gating and pixel-level masking can indeed reduce the side effects of feedback connections and increase the segmentation accuracy.

ACKNOWLEDGMENT

The work is supported by National Key R&D Program of China (2018AAA0100702), Shanghai Municipal Science and Technology Major Project (2021SHZDZX0102), and NSFC(61802256).

REFERENCES

- [1] Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham.
- [2] Long, J., Shelhamer, E., and Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3431-3440).
- [3] Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., and Liang, J. (2018). Unet++: A nested u-net architecture for medical image segmentation. In Deep learning in medical image analysis and multimodal learning for clinical decision support (pp. 3-11). Springer, Cham.
- [4] Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., ... and Rueckert, D. (2018). Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999.
- [5] Wen, H., Han, K., Shi, J., Zhang, Y., Culurciello, E., and Liu, Z. (2018). Deep predictive coding network for object recognition. In International Conference on Machine Learning (pp. 5266-5275). PMLR.
- [6] Feng, M., Lu, H., and Ding, E. (2019). Attentive feedback network for boundary-aware salient object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 1623-1632).
- [7] Xiang, T., Zhang, C., Liu, D., Song, Y., Huang, H., and Cai, W. (2020). Bio-net: Learning recurrent bi-directional connections for encoder-decoder architecture. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 74-84). Springer, Cham.
- [8] Wang, X., Xiang, T., Zhang, C., Song, Y., Liu, D., Huang, H., and Cai, W. (2021). BiX-NAS: Searching Efficient Bi-directional Architecture for Medical Image Segmentation. arXiv preprint arXiv:2106.14033.
- [9] Xu, L. (1991). Least MSE reconstruction for self-organization. II. Further theoretical and experimental studies on one-layer nets. In [Proceedings] 1991 IEEE International Joint Conference on Neural Networks (pp. 2368-2373). IEEE.
- [10] Xu, L. (1993). Least mean square error reconstruction principle for self-organizing neural-nets. Neural networks, 6(5), 627-648.
- [11] Huang, W., Tu, S., and Xu, L. (2019). Revisit Lmser and its further development based on convolutional layers. arXiv preprint arXiv:1904.06307.
- [12] Huang, W., Tu, S., and Xu, L. (2020). Deep CNN Based Lmser and Strengths of Two Built-In Dualities. Neural Processing Letters, vol. 8, pp. 108418-108428, 2020
- [13] Li, P., Tu, S., and Xu, L. (2019). GAN Flexible Lmser for Super-resolution. In Proceedings of the 27th ACM International Conference on Multimedia (MM2019), pp. 756-764. ACM, 2019
- [14] Guo, Y., Huang, W., Chen, Y., and Tu, S. (2019). Regularize network skip connections by gating mechanisms for electron microscopy image segmentation. In 2019 IEEE International Conference on Multimedia and Expo (ICME) (pp. 868-873). IEEE.
- [15] Xu, L. (2019). An overview and perspectives on bidirectional intelligence: Lmser duality, double IA harmony, and causal computation. IEEE CAA J. Autom. Sinica, 6(4), 865-893.
- [16] Hu, J., Shen, L., and Sun, G. (2018). Squeeze-and-excitation networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 7132-7141).
- [17] Kumar, N., Verma, R., Sharma, S., Bhargava, S., Vahadane, A., and Sethi, A. (2017). A dataset and a technique for generalized nuclear segmentation for computational pathology. IEEE transactions on medical imaging, 36(7), 1550-1560.
- [18] Naylor, P., Laé, M., Reyal, F., and Walter, T. (2018). Segmentation of nuclei in histopathology images by deep regression of the distance map. IEEE transactions on medical imaging, 38(2), 448-459.
- [19] Lee, K., Zlateski, A., Vishwanathan, A., and Seung, H. S. (2015). Recursive training of 2D-3D convolutional networks for neuronal boundary detection. Advances in Neural Information Processing Systems.
- [20] Kingma, D. P., and Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.
- [21] Mehta, R., and Sivaswamy, J. (2017). M-net: A convolutional neural network for deep brain structure segmentation. In 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017) (pp. 437-440). IEEE.
- [22] Raza, S. E. A., Cheung, L., Shaban, M., Graham, S., Epstein, D., Pelengaris, S., ... and Rajpoot, N. M. (2019). Micro-Net: A unified model for segmentation of various objects in microscopy images. Medical image analysis, 52, 160-173.
- [23] Alom, M. Z., Yakopcic, C., Taha, T. M., and Asari, V. K. (2018.). Nuclei segmentation with recurrent residual convolutional neural networks based U-Net (R2U-Net). In NAECON 2018-IEEE National Aerospace and Electronics Conference (pp. 228-233). IEEE.
- [24] Chaurasia, A., and Culurciello, E. (2017). Linknet: Exploiting encoder representations for efficient semantic segmentation. In 2017 IEEE Visual Communications and Image Processing (VCIP) (pp. 1-4). IEEE.
- [25] Arganda-Carreras, I., Turaga, S. C., Berger, D. R., Cireşan, D., Giusti, A., Gambardella, L. M., ... and Seung, H. S. (2015). Crowdsourcing the creation of image segmentation algorithms for connectomics. Frontiers in neuroanatomy, 9, 142.
- [26] He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).