# REGULARIZE NETWORK SKIP CONNECTIONS BY GATING MECHANISMS FOR ELECTRON MICROSCOPY IMAGE SEGMENTATION

*Yuze Guo, Wenjing Huang, Yajing Chen, Shikui Tu*

Department of Computer Science and Engineering, Shanghai Jiao Tong University
{guoyuze, huangwenjing, cyj907, tushikui}@sjtu.edu.cn

## ABSTRACT

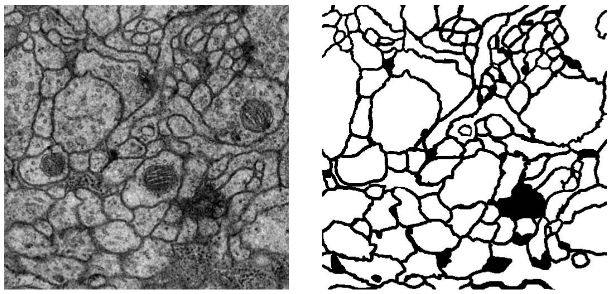
Recently, one earliest skip connected networks named Lmser was revisited and its convolutional layer based version named CLmser was proposed. This paper studies CLmser for segmentation (shortly CLmser-S) of Electron Microscopy (EM) images and also one further development. First, we experimentally show that CLmser-S outperforms the popular U-Net and save many free parameters. Second, we combine one newest formulation named Flexible Lmser (F-Lmser) and CLmser-S into a version called F-CLmser-S, together with learned masks replacing the similarity based one used in F-Lmser for implementing fast-lane skip connections. Experimental results on the ISBI 2012 EM dataset show that F-CLmser-S improves CLmser and achieves competitive performance with state-of-the-art results.

***Index Terms***— electron microscopy, image segmentation, flexible Lmser, CLmser, gated skip connections

## 1. INTRODUCTION

High-resolution Electron Microscopy (EM) image has been used in biomedical research to investigate the detailed structure of tissues, cells, organelles and so on. For example, EM images were used to study Drosophila brain structure [1] which required segmentation of neural structures from the images. Manual labeling of each element in the image requires by an expert human neuroanatomist. However, due to the visual complexity of the EM images, it can be time-consuming for human experts to interpret them one-by-one, which drives the demand for automated approaches.

Recently, deep learning methods have been used to solve the task of EM image segmentation based on Convolutional Neural Networks (CNN) [2, 3, 4, 5, 6, 7, 8, 9]. One of the early attempts is U-Net [3]. It consists of a contracting path as encoder and an expansive path as decoder, and both paths form a U-shaped architecture. The feature map from each of the layer of the contracting path was copied and concatenated with the symmetrically corresponding layer in the expansive path. Such skip connections enabled U-Net to work very well for biomedical image segmentation [3].



(a) (b)

**Fig. 1**. Examples in the ISBI 2012 EM dataset: an EM image (a) and its ground truth segmentation (b).

Many following works were built upon a U-Net-like structure[4, 7]. FusionNet[4] adopts residual blocks in the U-Net structure so that the coexistence of long skip connections and short-cuts enable the model to have a deeper depth and achieve higher performance. Based on conditional Generative Adversarial Network, ADDN[7] uses a U-net-like architecture with dilated convolutions in its generator, called densely dilated network.

The skip connections used in U-net [3] can actually be backtracked to one earliest skip connected networks called Lmser proposed in 1991 [10, 11]. The Lmser architecture is obtained by folding AE along the central hidden layer, and thus the same architecture takes a dual role for both encoder and decoder. Such folding also make the neurons on the paired layers between encoder and decoder merge into one, equivalently got skip connections in forward and backward directions jointly. Though Lmser learning was proposed in [10, 11] as one multiple layer deep learning approach, its advantages were only demonstrated in a one-hidden-layer implementation due to the lack of computing resources and big data at that time.

Recently, Huang et al. in [12] has revisited Lmser and has confirmed that deep Lmser learning works well on several potential functions addressed in [10, 11], demonstrated by experiments on image recognition, reconstruction, association recall, and so on. Moreover, Lmser is developed into

978-1-5386-9552-4/19/$31.00 ©2019 IEEE
DOI 10.1109/ICME.2019.00154

a multiple convolutional layers based version named CLmser for image related tasks.

In this paper, the fast-lane CLmser is applied on segmentation of EM images, i.e., only skip connections from encoder to decoder are considered, without feedback from decoder to encoder. Not only we experimentally show that CLmser outperforms U-Net and save many free parameters, but also we proceed to a version called F-CLmser-S that integrates CLmser and one newest formulation named Flexible Lmser (F-Lmser) [13], featured with fast-lane skip connections that are regularised by learned templates instead of by similarity between bottom-up pattern and top-down pattern as suggested in F-Lmser. Experimental results on the ISBI 2012 EM dataset show that F-CLmser-S improves CLmser and achieves competitive performance with state-of-the-art results.

We summarize our contribution as follows:

- We present a F-CLmser-S network for EM image segmentation. We consider the fast-lane version of CLmser, to have skip connections from the neurons of encoder to the ones in symmetrically paired layers of decoder, without feedback from decoder to encoder.

- We propose to gate the patterns transferring through the skip connections at different levels of layers, in order to reduce the noisy, redundant, uninformative patterns for the task of segmentation. At high-level layers close to the central hidden layer, we compute the gating mask from the layer output in encoder to filter the feature maps by channels and by pixels, while at low-level layers close to input and output, we compute the gating mask from the layer output in decoder to identify the uncertain prediction to be supplied with more details from encoder.

- Experimental results show that the proposed method can have comparable performance with state-of-the-art methods in the ISBI 2012 EM challenge. Ablation study and qualitative evaluation further demonstrate that the gating masks at low- or high-level layers can enhance and refine the segmentation outputs.

## 2. RELATED WORK

### 2.1. Deep Model for EM Image Segmentation

One of the earliest work by Ciresan *et al.*[2] simply used a succession of convolutional and max-pooling layers to perform the prediction. Their pioneer work won the ISBI 2012 challenge. Long *et al.*[14] proposed to replace fully connected layers with fully convolutional layer for preserving spatial information and allowing arbitrary input size. Since then, many variants of FCN have been proposed for EM image segmentation. Chen *et al.*[6] adopts a concatenation of multi-level feature maps to integrate different contextual information. Shen *et al.*[5] present a multi-stage and multi-recursive-input FCN. In each stage, the model learns to predict outputs at different levels. Then, all the predictions are combined with the original images to form the input for the next stage. Ronneberger *et al.*[3] proposed a U-net architecture consisted of a contracting path and a symmetric expanding path. They replace pooling operations by upsampling, and use skip connections to preserve low-level information. The skip connections are then combined with high resolution features from the contracting paths. However, the model still suffers from the vanishing gradient problem, which limits the depth of U-net. He *et al.*[15] proposed the residual blocks and demonstrated that shortcut connections and direction summations can reduce the influence of vanishing gradients. Combining short and long skip connections, Quan *et al.*[4] presented FusionNet, which leverages the U-net with residual blocks.

### 2.2. Attention Mechanism

Inspired by the human perception process, numerous studies have been proposed to apply attention mechanism into neural networks. Recently, several approaches attempt to integrate attention modules with state-of-the-art deep model architecture to improve the performance of networks[16, 17, 18]. Residual Attention Network[16] was built by stacking attention modules, the network performs very well on several benchmarks and is proven to be robust to noisy inputs. Zhang *et al.*incorporate Residual network with channel attention which is able to re-scale channel-wise features to solve image super-resolution problems[17]. A lightweight and general module called Convolutional Block Attention Module(CBAM) [18] combines channel attention and spatial attention. It can be incorporated into existing models to improve the results in classification and detection problems.

## 3. METHODS

### 3.1. Overview of the proposed network

The overall architecture of the proposed model is similar to the CNN based Lmser in [12], as shown in Figure 2. Different from [12], we use residual blocks as the basic building modules, and we propose a gating strategy to make the skip connections focus on the important features and ignore the redundant ones. Specifically, we compute attention masks based on the output of layers in encoder, for filtering pixel-level and channel features transferred from encoder at high-level layers close to the central hidden layer, while we compute confidence masks based on the output of layers in decoder, to allow the uncertain segmentation regions to receive more details from the encoder, for low-level layers close to the input and final segmentation output. With the gating masks at different levels, the irrelevant patterns are blocked, the missing details are enhanced, and then the final segmentation results are refined and improved.

In practice, all the convolutional layers adopt $3\times3$ kernels with stride size as 1. For all the deconvolutional or transposed convolutional layers, we use $3 \times 3$ kernels with stride size as 2. Activation functions are set as ReLUs.

### 3.2. Gating Feature Maps by Channels and by Pixels

Channel gating aims to capture the inter-dependencies of different channels by first squeezing and then expanding the channel size. Pixel-level gating is to filter essential spatial patterns by the computed weights.

Figure 2(b) shows the details of two gating modules in high-level layers. Given the feature map $F\in\mathbb{R}^{H\times W\times C}$ which will be transferred through the skip-connections, the gating weight matrix $W_{CG} \in \mathbb{R}^{1\times1\times C}$ to gate the channels are computed from $F$ itself, and the weight matrix $W_{PG} \in \mathbb{R}^{H\times W\times 1}$ are calculated from the output feature map $F_1$ of the gated channels,

$$F_1 = W_{CG} \otimes F, \quad F_2 = W_{PG} \otimes F_1, \tag{1}$$

where $\otimes$ indicates the scaling operator along the channel or pixel coordinates.

**Channel Gating.** We construct the channel gating module in the same way as [17]. As shown in Figure 2(b), we first apply average pooling on the feature maps $F$ to get a feature vector $F_1 \in \mathbb{R}^{1\times1\times C}$. Then, two $1 \times 1$ convolutional layers are used to compute the weights for filtering information along the channel dimension. The number of channels is kept unchanged.

**Pixel-level Gating.** As in Figure 2(b), average pooling operation is used to aggregate the channel information of a feature map, then a convolutional layer and a sigmoid activation function are employed to compute pixel-level gating map.

### 3.3. Gating Low-level Layers

Different from pixel-level and channel gating weights, the gating masks for low-level layers are computed on the outputs of the decoder layers close to the final segmentation output, to select the uncertain segmentation regions, and they allow the skip connections to pass more details from the encoder to refine the segmentation on such uncertain regions.

Specifically, the mask is computed by:

$$f(x) = (1 - x^2)^\gamma, \tag{2}$$

where $x$ denotes an entry of the feature maps, $\gamma$ is a hyperparameter to control the shape of the function curve. The whole process of generating a gating mask is illustrated in Figure 2(c).

The segmentation task can be viewed as a binary classification problem on pixels, where the membrane is $-1$ and cell is $+1$. The prediction is considered of high confidence with pixel values near $-1$ or $+1$, while values near 0 indicate that the network cannot tell whether the corresponding pixels are within the membrane or non-membrane region. The function by Eq.(2) has high values with inputs near 0 and low values with inputs near $-1$ and $+1$, which allows the skip connections to focus more on uncertain prediction. Therefore, we can refine the outputs by fusing the decoder layers with filtered patterns passed through skip connections. In practice, other functions with similar characteristics might also be used.

### 3.4. Loss Function

Since the cross entropy loss might induce gradient vanishing problem in modern deep-learning frameworks, we adopt smooth L1 loss [19] as our loss function:

$$\mathcal{L}(X,Y) = \frac{1}{w \times h} \sum_{i,j} E_{i,j} \tag{3}$$

$$E_{i,j} = \begin{cases} 0.5(X_{i,j} - Y_{i,j})^2, & \text{if } |X_{i,j} - Y_{i,j}| < 1 \\ |X_{i,j} - Y_{i,j}| - 0.5, & \text{otherwise} \end{cases}, \tag{4}$$

where $X \in \mathcal{R}^{h\times w}$ and $Y \in \mathcal{R}^{h\times w}$ are network prediction and ground truth respectively, $h$ and $w$ are the height and width of the test images.

## 4. EXPERIMENTS AND RESULTS

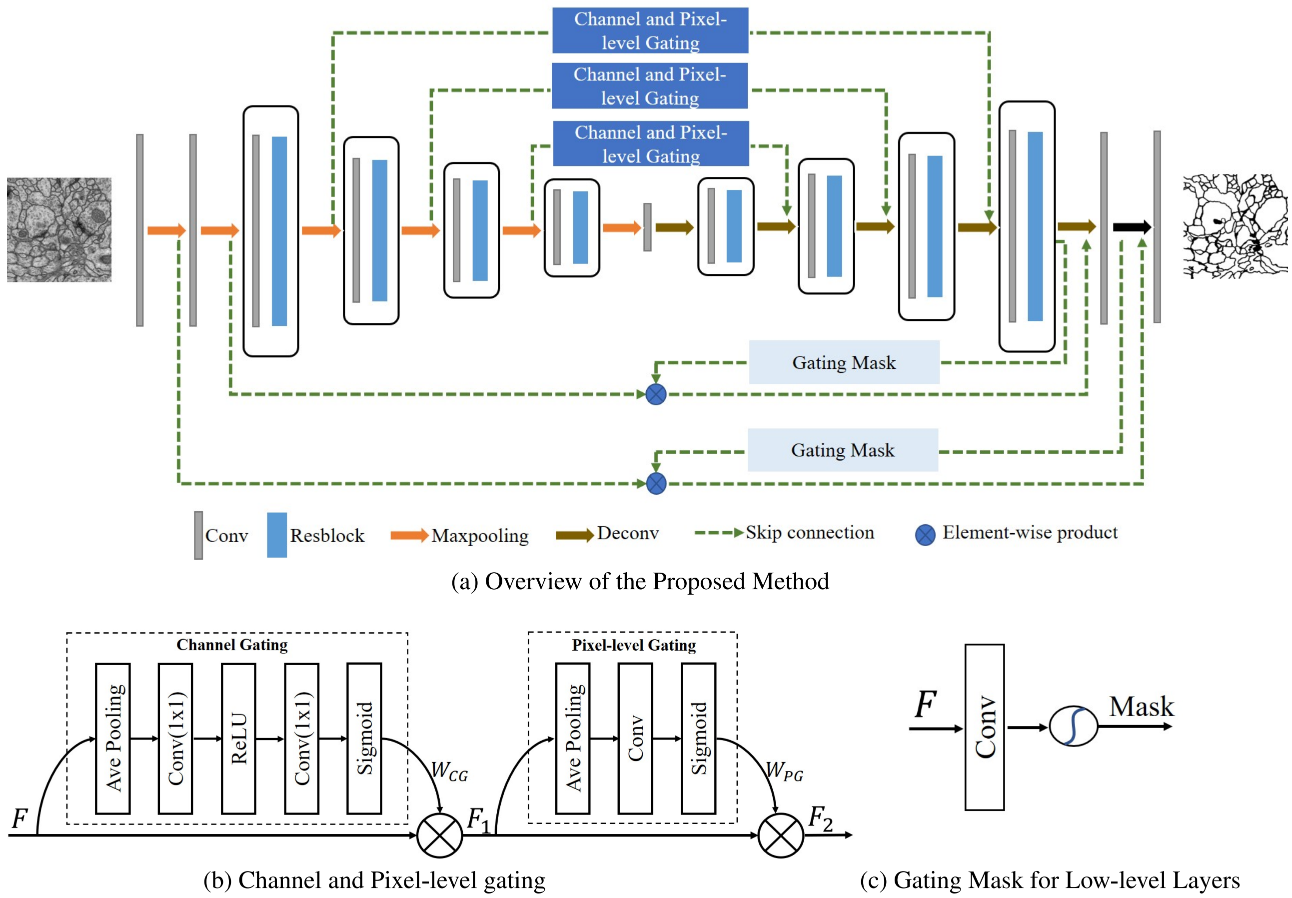
### 4.1. ISBI 2012 EM Segmentation Dataset

The training data are 30 pairs of EM images and ground truth labels obtained from ISBI 2012 EM Segmentation Challenge [20]. Figure 1 shows an example of the dataset, where the ground truth is a binary image with membranes in white and non-membrane area in black. The testing data for public also contain 30 EM images, while the ground truths are not provided.

### 4.2. Experimental Setup

In the training phase, we randomly split the dataset into 25 training pairs and 5 validation pairs. As the training dataset is small, we apply several data augmentation techniques to enrich our training data, including rotation, horizontal flip, elastic transformation, random crop, and mirror reflections on the boundary. We set the parameter $\gamma = 3$ in Eq.(2) to compute the confidence mask [1]. We train the model using Adam optimizer. The learning rate is set to be $2 \times 10^{-4}$ initially and decays by a factor of 10 every 300 epochs. We use weight decay policy to prevent the network from overfitting with weighting

[1]We perform experiments on $\gamma = 1, 2, 3, 4$, and $\gamma = 3$ shows the best results.




(a) Overview of the Proposed Method

(b) Channel and Pixel-level gating

(c) Gating Mask for Low-level Layers

**Fig. 2**. Overview of the proposed model. (a) shows the overall structure. We adopt channel and pixel-level gating in high-level layers to filter the corresponding skip connections. And at low level layers, a two-stage cascading usage of masks is applied to refine the temporary outputs step by step to obtain the final outputs. (b) describes the computation of the channel-level and pixel-level gating. (c) depicts the computation of the gating mask in low-level layers.

parameter as $10^{-4}$. Post-processing[21] is performed to obtain the final segmentation results.

In the testing phase, we evaluate the performance of the proposed model using two different metrics, *i.e.*, Foreground-restricted Rand Scoring after border thinning($V^{Rand}$) and Foreground-restricted Information Theoretic Scoring after border thinning($V^{Info}$). The details of these two metrics can be found in [20]. The results are sorted by $V^{Rand}$ since it is more robust.

### 4.3. Ablation Study

We conduct ablation study to evaluate the effectiveness of different functional modules in our model, and the numerical evaluation results are given in Table 1. We use the fast-lane CLmser without duality on connection weights(CLmser-w) as the baseline model, which have residual blocks as the basic building blocks within the CLmser structure. *CG* and *PG* represent channel-level and pixel-level gating respectively, and they are all applied in skip connections at deep layers. *GM* means the gating mask applied in skip connections at low-

**Table 1**. Roles of different ingredients. Here, the fast-lane CLmser without duality on connection weights(CLmser-w) is used as a baseline.

| Model | $V^{rand}$ | $V^{info}$ |
|---|---|---|
| CLmser-w | 0.97310 | 0.98712 |
| CLmser-w + CG | 0.97359 | 0.98726 |
| CLmser-w + CG + PG | 0.97438 | 0.98868 |
| CLmser-w + CG + FullPG | 0.97709 | 0.98710 |
| CLmser-w + CG + PG + GM | **0.98223** | **0.98919** |

**Table 2**. Comparison between Different Models

| Method | $V^{rand}$ | $V^{info}$ |
|---|---|---|
| SFCNNs [22] | **0.98680** | **0.99144** |
| ADDN [7] | 0.98317 | 0.99088 |
| **Our approach** | 0.98223 | 0.98919 |
| PolyMtl[8] | 0.98058 | 0.98816 |
| M2FCN [5] | 0.97805 | 0.98919 |
| FusionNet [4] | 0.97804 | 0.98893 |
| CUMedVision [6] | 0.97682 | 0.98865 |
| FCN+LSTM [9] | 0.97537 | 0.98743 |
| Unet [3] | 0.97276 | 0.98662 |

**Table 3**. Parameters comparisons between Different Models

| Method | Parameters(M) |
|---|---|
| **Our approach** | **8.3** |
| ADDN [7] | 8.9 |
| PolyMtl[8] | 13 |
| FusionNet [4] | 31 |
| Unet [3] | 33 |

leval layers. We can see a mild increase in performance with the addition of channel-level gating on the baseline model. After applying pixel-level gating, the model sees a larger improvement. The performance is further significantly improved when adding the masks to refine the predicted output in successive steps.

As both pixel-level gating and the masks for low-level layers provide gating policies on the spatial dimension for skip connections, we compare two alternatives in model choice by applying either one in the low-level layers. Comparison between the results of the last tow lines in Table 1 also demonstrate that that using gating mask in low-level layers outperforms the usage of pixel-level gating, where *FullPG* means pixel-level gating is applied in all skip connections. The reason might be that the gating masks generated at low-level layers help the skip connections extract more useful low-level features to predict a better segmentation.

### 4.4. Comparisons with state-of-the-art approaches

We compare the proposed method with other models on this benchmark dataset. For fair comparisons, we only list published results for models whose main contribution lie in a new model architecture. From Table 2, we can see that the proposed model can achieve competitive performance with state-of-the-art. Table 3 records the results of parameters comparison between several methods. Note that FusionNet [4] is a combination of U-net and residual blocks. Thus, it shares a similar architecture with our baseline model but with much deeper layers. With the help of gating mechanisms on skip connections, the proposed model outperforms FusionNet with fewer parameters.

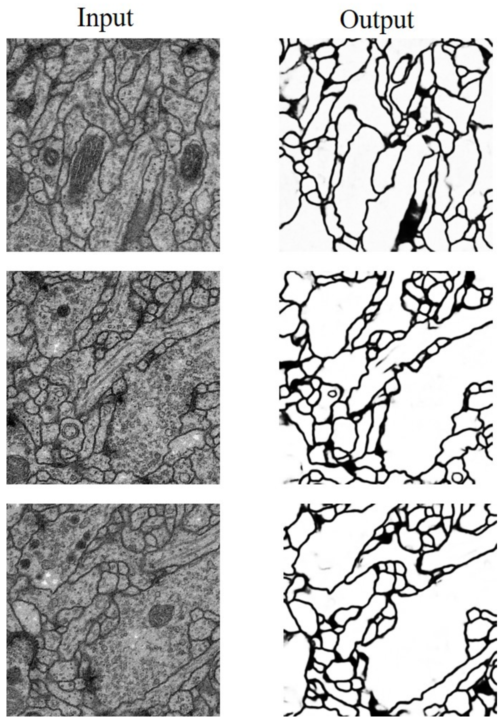



**Fig. 3**. Visualization of original EM images and the corresponding segmentation results by our method. The darker color of pixels represent higher probability of being membrane

Figure 3 shows examples of testing input-prediction pairs by the proposed method. We can see that the predictions by the proposed model remove the nucleus and other tiny elements within cells while maintaining the boundaries between neurons.

## 5. CONCLUSION

In this paper, we present a F-CLmser-S network for biomedical image segmentation, which integrates the fast-lane CLmser and F-Lmser. Based on the built-in dualities of Lmser, the encoder and decoder of the proposed network share the same architecture, and skip connections have been added symmetrically from encoder to decoder. We leverage feature levels of different layers to compute the gating policies for the feature maps, to improve the efficiency of the skip connections. At high-level layers close to the central coding layer, we gate the skip connections by weighting the channels and pixels, while at low-level layers, we exploit the masks to filter the skip feature maps. Experimental results on the ISBI 2012 EM dataset show that the proposed model can achieve competitive performance with state-of-the-art.

## Acknowledgement

This work was supported by a start-up grant (WF220403029) from Shanghai Jiao Tong University.

## 6. REFERENCES

[1] Albert Cardona, Stephan Saalfeld, Stephan Preibisch, Benjamin Schmid, Anchi Cheng, Jim Pulokas, Pavel Tomancak, and Volker Hartenstein, "An integrated micro- and macroarchitectural analysis of the drosophila brain by computer-assisted serial section electron microscopy," *PLOS Biology*, 2010.

[2] Dan Ciresan, Alessandro Giusti, Luca M. Gambardella, and Jürgen Schmidhuber, "Deep neural networks segment neuronal membranes in electron microscopy images," in *NIPS*. 2012.

[3] O. Ronneberger, P.Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *MICCAI*, 2015.

[4] Tran Minh Quan, David G. C. Hildebrand, and Won-Ki Jeong, "Fusionnet: A deep fully residual convolutional neural network for image segmentation in connectomics," *CoRR*, vol. abs/1612.05360, 2016.

[5] Wei Shen, Bin Wang, Yuan Jiang, Yan Wang, and Alan L. Yuille, "Multi-stage multi-recursive-input fully convolutional networks for neuronal boundary detection," in *ICCV*, 2017.

[6] Hao Chen, Xiaojuan Qi, Jie-Zhi Cheng, and Pheng-Ann Heng, "Deep contextual networks for neuronal structure segmentation," in *AAAI*, 2016.

[7] Ke Chen, Dandan Zhu, Jianwei Lu, and Ye Luo, "An adversarial and densely dilated network for connectomes segmentation," *Symmetry*, 2018.

[8] Michal Drozdzal, Gabriel Chartrand, Eugene Vorontsov, Mahsa Shakeri, Lisa Di-Jorio, An Tang, Adriana Romero, Yoshua Bengio, Chris Pal, and Samuel Kadoury, "Learning normalized inputs for iterative estimation in medical image segmentation," *Medical Image Analysis*, 2018.

[9] Jianxu Chen, Lin Yang, Yizhe Zhang, Mark Alber, and Danny Z Chen, "Combining fully convolutional and recurrent neural networks for 3d biomedical image segmentation," in *NIPS*. 2016.

[10] Lei Xu, "Least mse reconstruction for self-organization: (i)&(ii)," in *Proc. of 1991 International Joint Conference on Neural Networks*, 1991, pp. 2363–2373.

[11] Lei Xu, "Least mean square error reconstruction principle for self-organizing neural-nets," *Neural networks*, vol. 6, no. 5, pp. 627–648, 1993.

[12] Wenjing Huang, Shikui Tu, and Lei Xu, "Revisit lmser and its further development based on convolutional layers," *arXiv preprint 1904.06307*, 2019.

[13] Lei Xu, "An overview and perspectives on bidirectional intelligence: Lmser duality, double ia harmony, and causal computation," *IEEE/CAA Journal of Automatica S (to appear)*, vol. 6, no. 2-3, 2019.

[14] Evan Shelhamer, Jonathan Long, and Trevor Darrell, "Fully convolutional networks for semantic segmentation," *TPAMI*, 2017.

[15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *CVPR*, 2016.

[16] Fei Wang, Mengqing Jiang, Chen Qian, Shuo Yang, Cheng Li, Honggang Zhang, Xiaogang Wang, and Xiaoou Tang, "Residual attention network for image classification," in *CVPR*, 2017.

[17] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu, "Image super-resolution using very deep residual channel attention networks," in *ECCV*, 2018.

[18] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon, "CBAM: convolutional block attention module," in *ECCV*, 2018.

[19] Ross Girshick, "Fast r-cnn," in *International Conference on Computer Vision (ICCV)*, 2015.

[20] Ignacio Arganda-Carreras, Srinivas C. Turaga, Daniel R. Berger, Cirean Dan, Alessandro Giusti, Luca M. Gambardella, Jrgen Schmidhuber, Dmitry Laptev, Sarvesh Dwivedi, and Joachim M. Buhmann, "Crowdsourcing the creation of image segmentation algorithms for connectomics," *Frontiers in Neuroanatomy*, 2015.

[21] T Beier, B Andres, Ullrich Köthe, and F. A. Hamprecht, "An efficient fusion move algorithm for the minimum cost lifted multicut problem," in *ECCV*, 2016.

[22] Maurice Weiler, Fred A. Hamprecht, and Martin Storath, "Learning steerable filters for rotation equivariant cnns," in *CVPR*, 2018.