BiO-Net: Learning Recurrent Bi-directional Connections for Encoder-Decoder Architecture

Tiange Xiang[1], Chaoyi Zhang[1], Dongnan Liu[1], Yang Song[2], Heng Huang[3,4], and Weidong Cai[1(✉)]

[1] School of Computer Science, University of Sydney, Sydney, Australia
{txia7609,dliu5812}@uni.sydney.edu.au
{chaoyi.zhang,tom.cai}@sydney.edu.au
[2] School of Computer Science and Engineering, University of New South Wales, Sydney, Australia
yang.song1@unsw.edu.au
[3] Electrical and Computer Engineering, University of Pittsburgh, Pittsburgh, USA
henghuanghh@gmail.com
[4] JD Finance America Corporation, Mountain View, CA, USA

Abstract. U-Net has become one of the state-of-the-art deep learning-based approaches for modern computer vision tasks such as semantic segmentation, super resolution, image denoising, and inpainting. Previous extensions of U-Net have focused mainly on the modification of its existing building blocks or the development of new functional modules for performance gains. As a result, these variants usually lead to an unneglectable increase in model complexity. To tackle this issue in such U-Net variants, in this paper, we present a novel **Bi**-directional **O**-shape network (BiO-Net) that reuses the building blocks in a recurrent manner without introducing any extra parameters. Our proposed bi-directional skip connections can be directly adopted into any encoder-decoder architecture to further enhance its capabilities in various task domains. We evaluated our method on various medical image analysis tasks and the results show that our BiO-Net significantly outperforms the vanilla U-Net as well as other state-of-the-art methods. Our code is available at https://github.com/tiangexiang/BiO-Net.

Keywords: Semantic segmentation · Bi-directional connections · Recursive neural networks

1 Introduction

Deep learning based approaches have recently prevailed in assisting medical image analysis, such as whole slide image classification [27], brain lesion

Electronic supplementary material The online version of this chapter (https://doi.org/10.1007/978-3-030-59710-8_8) contains supplementary material, which is available to authorized users.

A. L. Martel et al. (Eds.): MICCAI 2020, LNCS 12261, pp. 74–84, 2020.
https://doi.org/10.1007/978-3-030-59710-8_8

segmentation [26], and medical image synthesis [10]. U-Net [22], as one of the most popular deep learning based models, has demonstrated its impressive representation capability in numerous medical image computing studies. U-Net introduces skip connections that aggregate the feature representations across multiple semantic scales and helps prevent information loss.

U-Net Variants. Recent works were proposed to extend the U-Net structure with varying module design and network construction, illustrating its potentials on various visual analysis tasks. V-Net [17] applies U-Net on higher dimension voxels and keeps the vanilla internal structures. W-Net [25] modifies U-Net to tackle unsupervised segmentation problem by concatenating two U-Nets via an autoencoder style model. Compared to U-Net, M-Net [16] appends different scales of input features to different levels, thus multi-level visual details can be captured by a series of downsampling and upsampling layers. Recently, U-Net++ [28] adopts nested and dense skip connections to represent the fine-grained object details more effectively. Moreover, attention U-Net [20] uses extra branches to adaptively apply attention mechanism on the fusion of skipped and decoded features. However, these proposals may involve additional building blocks, which lead to a greater number of network parameters and thus an increased GPU memory. Unlike above variants, our BiO-Net improves the performance of U-Net via a novel feature reusing mechanism where it builds bi-directional connections between the encoder and decoder to make inference from a recursive manner.

Recurrent Convolutional Networks. Using recurrent convolution to iteratively refine the features extracted at different times has been demonstrated to be feasible and effective for many computer vision problems [1,7,8,24]. Guo *et al.* [7] proposed to reuse residual blocks in ResNet so that available parameters would be fully utilized and model size could be reduced significantly. Such a mechanism also benefits the evolution of U-Net. As a result, Wang *et al.* [24] proposed R-U-Net, which recurrently connects multiple paired encoders and decoders of U-Net to enhance its discrimination power for semantic segmentation, though, extra learnable blocks are introduced as a trade-off. BiO-Net distinguishes R-U-Net from its design of backward skip connections, where latent features in every decoding levels are reused, enabling more intermediate information aggregations with gradients preserved among temporal steps. R2U-Net [1] adopts a similar approach that only recurses the last building block at each level of refinement. By contrast, our method learnt recurrent connections in the existing encoder and decoder rather than recursing the same level blocks without the involvement of refined decoded features.

To this end, we propose a recurrent U-Net with an bi-directional O-Shape inference trajectory (BiO-Net) that maps the decoded features back to the encoder through the backward skip connections, and recurses between the encoder and the decoder. Compared to previous works, our approach achieves a better feature refinement, as multiple encoding and decoding processes are triggered in our BiO-Net. We applied our BiO-Net to perform semantic segmentation on nuclei segmentation task and EM membrane segmentation task and our results show that the proposed BiO-Net outperforms other U-Net

variants, including the recurrent counterparts and many state-of-the-art approaches. Experiments on super resolution tasks also demonstrate the significance of our BiO-Net applied to different scenarios.

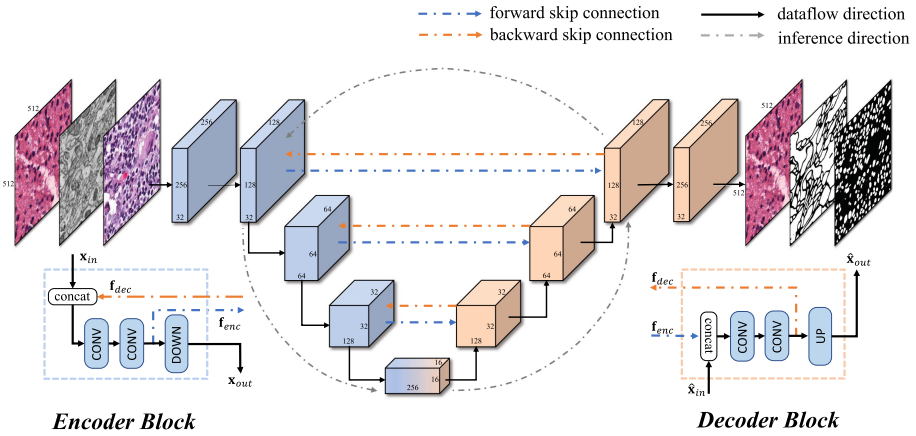

Fig. 1. Overview of our BiO-Net architecture. The network inferences recurrently in an O-shape manner. `CONV` represents the sequence of convolution, non-linearity and batch-norm layers. `DOWN` stands for downsampling (implemented by a `CONV` followed by max-pooling), while `UP` denotes upsampling (archived by transpose convolution).

2 Methods

As shown in Fig. 1, BiO-Net adopts the same network architecture as U-Net, without any dependencies on extra functional blocks but with paired bi-directional connections. It achieves better performance as t increases with no extra trainable parameters introduced during its unrolling process. Moreover, our method is not restricted to U-Net and can be integrated into other encoder-decoder architectures for various visual analysis tasks.

2.1 Recurrent Bi-directional Skip Connections

The main uniqueness of our BiO-Net model is the introduction of bi-directional skip connections, which facilitate the encoders to process the semantic features in the decoders and vice versa.

Forward Skip Connections. Forward skip connections linking encoders and decoders at the same level can preserve the encoded low-level visual features $\mathbf{f}_{enc}$, with their gradients well-preserved [9,22]. Hence, the l-th decoder block could fuse $\mathbf{f}_{enc}$ with its input $\hat{\mathbf{x}}_{in}$ generated from lower blocks, and propagate them

through the decoding convolutions DEC to generate $\mathbf{f}_{dec}$, which will be further restored to higher resolutions via UP block. This process can be defined as:

$$\mathbf{f}_{dec} = \texttt{DEC}([\mathbf{f}_{enc},\ \hat{\mathbf{x}}_{in}]), \tag{1}$$

where concatenation is employed as our fusion mechanism $[\cdot]$. The index notation (l-th) for encoders and decoders is omitted in this paper for simplicity purpose.

Backward Skip Connections. With the help of our novel backward skip connections, which pass the decoded high-level semantic features $\mathbf{f}_{dec}$ from the decoders to the encoders, our encoder can now combine $\mathbf{f}_{dec}$ with its original input $\mathbf{x}_{in}$ produced by previous blocks, and therefore achieves flexible aggregations of low-level visual features and high-level semantic features. Similar to the decoding path enhanced by forward skip connections, our encoding process can thus be formulated with its encoding convolutions ENC as:

$$\mathbf{f}_{enc} = \texttt{ENC}([\mathbf{f}_{dec},\ \mathbf{x}_{in}]). \tag{2}$$

The DOWN block feeds $\mathbf{f}_{enc}$ to subsequent encoders for deeper feature extraction.

Recursive Inferences. The above bi-directional skip connections create an O-shaped inference route for encoder-decoder architectures. Noticeably, this O-shaped inference route can be recursed multiple times to receive immediate performance gains, and more importantly, this recursive propagation policy would not introduce any extra trainable parameters. Hence, the outputs of encoders and decoders equipped with our proposed O-shaped connections can be demonstrated as follows, in terms of their current inference iteration i:

$$\begin{aligned}
\mathbf{x}_{out}^{i} &= \texttt{DOWN}(\texttt{ENC}([\texttt{DEC}([\mathbf{f}_{enc}^{i-1}, \hat{\mathbf{x}}_{in}^{i-1}]), \mathbf{x}_{in}^{i}])), \\
\hat{\mathbf{x}}_{out}^{i} &= \texttt{UP}(\texttt{DEC}([\texttt{ENC}([\mathbf{f}_{dec}^{i},\ \mathbf{x}_{in}^{i}]), \hat{\mathbf{x}}_{in}^{i}])).
\end{aligned} \tag{3}$$

Compared to the vanilla U-Net, our BiO-Net takes both encoded and decoded features into consideration and performs refinement according to features from the previous iterations.

2.2 BiO-Net Architecture

We use the plain convolutional layers, batch normalization [11] layers, and ReLU [18] layers in the network architecture. No batch normalization layer is reused.

The input image is first fed into a sequence of three convolutional blocks to extract low-level features. Note that there is no backward skip connection attached to the first stage block and, hence, the parameters in the first stage block will not be reused when recursing. The extracted features are then sent to a cascade of encode blocks that utilize max pooling for feature downsampling. During the encoding stage, the parameters are reused and the blocks are recursed through the paired forward and backward connections as shown in

Fig. 1. After the encoding phase, an intermediate stage of which contains convolutional blocks that are used to further refine the encoded features. Then, the features are subsequently passed into a series of decode blocks that recover encoded details using convolutional transpose operations. During the decoding stage, our proposed backward skip connections preserve retrieved features by concatenating them with the features from the same level encoders as depicted in Sect. 2.1. The recursion begins with the output generated from the last convolutional block in the decoding stage. After recursing the encoding and decoding stages, the updated output will be fed into the last stage block corresponding to the first stage blocks. Similarly, the last stage blocks will not be involved in the recurrence.

We define several symbols for better indication: 't' represents the total recurrence time; '$\times n$' represents the expansion multiplier times to all hidden output channel numbers; 'w' represents the number of backward skip connections used from the deepest encoding level; '`INT`' represents stacking decoded features from each iteration and feeding them into the last stage block as a whole and 'l' represents the most encoding depth. Details can be seen in Fig. 2.

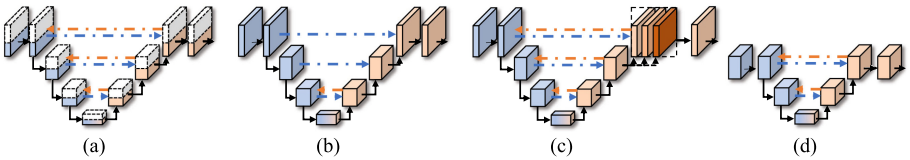

Fig. 2. Visualizations of setups in the ablation experiments. (a) BiO-Net with $\times 0.5$. (b) BiO-Net with $w = 2$. (c) BiO-Net with $t = 3$ and `INT`. (d) BiO-Net with $l = 3$.

3 Experiments

Datasets. Our method was evaluated on three common digital pathology image analysis tasks: nuclei segmentation, EM membrane segmentation, and image super resolution on a total of four different datasets. Two publicly available datasets, namely MoNuSeg [13] and TNBC [19], were selected to evaluate our method for nuclei semantic segmentation. The MoNuSeg dataset consists of a 30-image training set and a 14-image testing set, with images of size 1000^2 sampled from different whole slide images of multiple organs. We extract 512^2 patches from 4 corners of each image, which enlarges the dataset by 4 times. TNBC is comprised of 50 histopathology images of size 512^2 without any specific testing set. Both datasets include pixel-level annotation for the nuclei semantic segmentation problem. The second task we evaluated is EM membrane segmentation, where the piriform cortex EM dataset of the mice collected from [15], which contains four stacks of EM images with the slice image sizes of 255^2,

512^2, 512^2, and 256^2, respectively. Image super resolution is the last task we evaluated our method on, the dataset was constructed from a whole slide image collected by MICCAI15 CBTC. We sampled 2900 patches of size 512^2 at $40\times$ magnification level, with the 9 : 1 split ratio for the training and testing set, respectively.

Implementation Details. We used Adam [12] optimizer with an initial learning rate of 0.01 and a decay rate of 0.00003 to minimize cross entropy loss in segmentation tasks and mean square error in super resolution tasks. The training dataset was augmented by applying random rotation (within the range $[-15°, +15°]$), random shifting (in both x- and y-directions; within the range of $[-5\%, 5\%]$), random shearing, random zooming (within the range $[0, 0.2]$), and random flipping (both horizontally and vertically). The batch size is set to 2 in both training and testing phases. Unless explicitly specified, our BiO-Net is constructed with an encoding depth of 4 and a backward skip connection built at each stage of the network. Our BiO-Net was trained by 300 epochs in all experiments, which were conducted on a single NVIDIA GeForce GTX 1080 GPU with Keras. Given the GPU limitation, we explore the performance improvement to its maximum possible temporal step at $t = 3$.

Table 1. Comparison of segmentation methods on MoNuSeg testing set and TNBC.

	MoNuSeg		TNBC			
Methods	IoU	DICE	IoU	DICE	#params	Model size
U-Net [22] **w.** ResNet-18 [9]	0.684	0.810	0.459	0.603	15.56 M	62.9 MB
U-Net++ [28] **w.** ResNet-18 [9]	0.683	0.811	0.526	0.652	18.27 M	74.0 MB
U-Net++ [28] **w.** ResNet-50 [9]	0.695	0.818	0.542	0.674	37.70 M	151.9 MB
Micro-Net [21]	0.696	0.819	0.544	0.701	14.26 M	57.4 MB
Naylor *et al.* [19]	0.690	0.816	0.482	0.623	36.63 M	146.7 MB
M-Net [16]	0.686	0.813	0.450	0.569	0.6 M	2.7 MB
Att U-Net [20]	0.678	0.810	0.581	0.717	33.04 M	133.2 MB
R2U-Net, $t = 2$ [1]	0.678	0.807	0.532	0.650	37.02 M	149.2 MB
R2U-Net, $t = 3$ [1]	0.683	0.815	0.590	0.711	37.02 M	149.2 MB
LinkNet [3]	0.625	0.767	0.535	0.682	11.54 M	139.4 MB
BiO-LinkNet [3], $t = 2$ (Ours)	0.621	0.766	0.541	0.690	11.54 M	139.4 MB
BiO-LinkNet [3], $t = 3$ (Ours)	0.634	0.774	0.571	0.716	11.54 M	139.4 MB
BiO-Net, $t = 1$ (ours)	0.680	0.803	0.456	0.608	15.0 M	60.6 MB
BiO-Net, $t = 2$ (ours)	0.694	0.816	0.548	0.693	15.0 M	60.6 MB
BiO-Net, $t = 3$ (ours)	0.700	0.821	0.618	0.751	15.0 M	60.6 MB
BiO-Net, $t = 3$, `INT` (Ours)	**0.704**	**0.825**	**0.651**	**0.780**	15.0 M	60.6 MB

3.1 Semantic Segmentation

Nuclei Segmentation. In this task, our method is compared to the baseline U-Net [22] and other state-of-the-art methods [1,16,19–21,28]. Following [6],

models were trained on the MoNuSeg training set only and evaluated on the MoNuSeg testing set and TNBC dataset. Dice coefficient (DICE) and Intersection over Union (IoU) are evaluated. As shown in Table 1, our results are better than others on the MoNuSeg testing set. Our results on the TNBC dataset are also higher than the others by a large margin, which demonstrates a strong generalization ability. Additionally, compared with other extensions of U-Net, and, our BiO-Net is more memory efficient. Qualitative comparison of our method and the recurrent counterpart R2U-Net [1] is shown in Fig. 3. It can be seen that our model segments nuclei more accurately as the inference time increases. In our experiments, BiO-Net infers a batch of two predictions in 35, 52, and 70 ms when $t = 1$, $t = 2$, and $t = 3$, respectively. Further evaluation of incorporating our method into another encoder-decoder architecture LinkNet [3], which is also shown in the table. Our BiO-LinkNet adds the skipped features element-wisely and, hence, shares the same number of parameters as the vanilla LinkNet.

Table 2. Ablative results. The parameters are defined as depicted in Sect. 2.2. IoU(DICE), number of parameters, and, model size are reported.

	MoNuSeg			TNBC				
	$t = 1$	$t = 2$	$t = 3$	$t = 1$	$t = 2$	$t = 3$	#params	Model size
$\times 1.25$	0.685 (0.813)	0.698 (0.819)	0.695 (0.817)	0.490 (0.637)	0.557 (0.697)	0.623 (0.758)	23.5 M	94.3 MB
$\times 1.0$	0.680 (0.803)	0.694 (0.816)	0.700 (0.821)	0.456 (0.608)	0.548 (0.693)	0.618 (0.751)	15.0 M	60.6 MB
$\times 0.75$	0.676 (0.800)	0.678 (0.805)	0.691 (0.815)	0.516 (0.661)	0.571 (0.710)	0.598 (0.738)	8.5 M	34.3 MB
$\times 0.5$	0.668 (0.792)	0.680 (0.806)	0.691 (0.814)	0.491 (0.644)	0.543 (0.679)	0.611 (0.742)	3.8 M	15.8 MB
$\times 0.25$	0.667 (0.791)	0.678 (0.804)	0.677 (0.804)	0.524 (0.674)	0.535 (0.678)	0.575 (0.710)	0.9 M	4.0 MB
$w = 3$	0.680 (0.803)	0.694 (0.817)	0.688 (0.814)	0.456 (0.608)	0.510 (0.656)	0.620 (0.757)	15.0 M	60.3 MB
$w = 2$	0.680 (0.803)	0.672 (0.801)	0.686 (0.813)	0.456 (0.608)	0.527 (0.679)	0.601 (0.742)	14.9 M	60.1 MB
INT	0.680 (0.803)	0.689 (0.812)	**0.704 (0.825)**	0.456 (0.608)	0.588 (0.728)	**0.651(0.780)**	15.0 M	60.6 MB
$l = 3$	0.681 (0.806)	0.679 (0.805)	0.689 (0.812)	0.613 (0.742)	0.594 (0.733)	0.615 (0.741)	3.8 M	15.4 MB
$l = 2$	0.690 (0.810)	0.695 (0.817)	0.697 (0.818)	0.596 (0.734)	0.647 (0.775)	0.596 (0.735)	0.9 M	4.0 MB

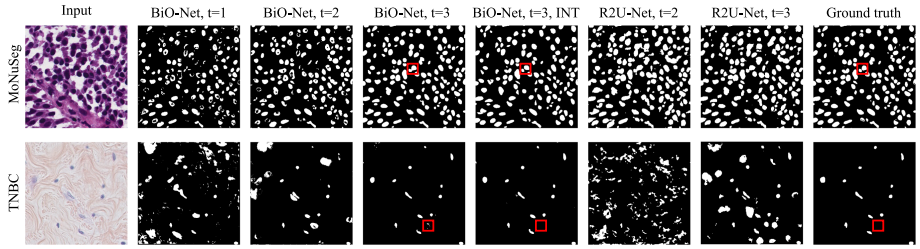

Fig. 3. Qualitative comparison between our models and the R2U-Net [1] on the MoNuSeg testing set and TNBC. Red boundary boxes indicate the effects of integrating features from each iteration at the last stage block. (Color figure online)

Table 2 demonstrates our ablation study by varying the setups as defined in Fig. 2. The results show that recursing through the encoder and the decoder with

the proposed bi-directional skip connections improves network performances generally. Integrating decoded features from all inference recurrences yields state-of-the-art performances in both datasets. Furthermore, we find that when there are insufficient parameters in the network, increasing inference recurrence has little improvement or even makes the results worse. It is also interesting to observe that when constructing BiO-Net with shallower encoding depth, our models perform better on the two datasets than those with deeper encoding depth.

EM Membrane Segmentation. We further evaluated our method by segmenting Mouse Piriform Cortex EM images [15], where the models are trained on stack1 and stack2, and validated on stack4 and stack3. The results were evaluated by Rand F-score [2]. As shown in Table 3, our method demonstrates better segmentation results with the proposed bi-directional O-shaped skip connections.

Table 3. Comparison of different U-Net variants in EM membrane segmentation.

Variants	Stack3	Stack4	#params	Ours	Stack3	Stack4	#params
U-Net	0.939	0.821	15.56 M	BiO-Net, t = 1	0.941	0.827	15.0 M
Att U-Net	0.937	0.833	33.04M	BiO-Net, t = 2	0.955	0.871	15.0 M
U-Net++	0.940	0.844	18.27 M	BiO-Net, t = 3	**0.958**	**0.887**	15.0 M

3.2 Super Resolution

In addition to segmentation tasks, we are also interested in experimenting with our BiO-Net on a significantly different task: image super resolution, which has been studied actively [4,23]. In the super resolution task, low-resolution (downsampled) images are used as inputs to train the networks toward their original high-resolution ground truth, which can assist medical imaging analysis by

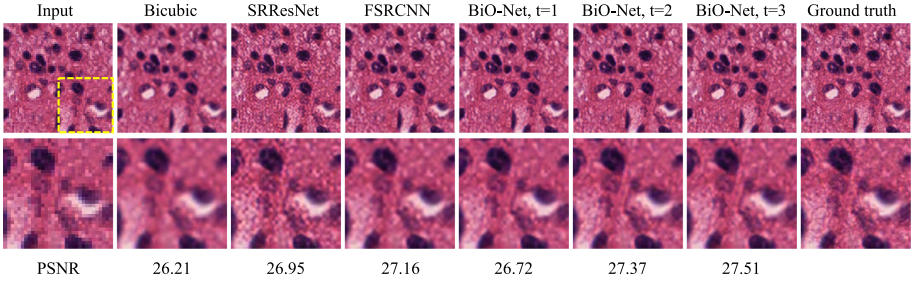

Fig. 4. Comparison of different methods in super resolution task. PSNR scores of the methods over the entire testing set are reported as well. The second row is projected from the yellow boundary box to have a closer view on the super resolution achieved in higher resolution. (Color figure online)

recovering missing details and generating high-resolution histopathology images based on the low-resolution ones. Two state-of-the-art methods, FSRCNN [5] and SRResNet [14], are adopted to compare with our BiO-Net. The qualitative results along with the Peak Signal to Noise Ratio (PSNR) score over the entire testing set are shown in Fig. 4. It can be seen that, our method outperforms the state-of-the-art methods by a safe margin, which validates the feasibility of applying our BiO-Net on different visual tasks.

4 Conclusion

In this paper, we introduced a novel recurrent variant of U-Net, named BiO-Net. BiO-Net is a compact substitute of U-Net with better performance and no extra trainable parameters, which utilizes paired forward and backward skip connections to compose a complex relation between the encoder and decoder. The model can be recursed to reuse the parameters during training and inference. Extensive experiments on semantic segmentation and super-resolution tasks indicate the effectiveness of our proposed model, which outperforms the U-Net and its extension methods without introducing auxiliary parameters.

References

1. Alom, M.Z., Yakopcic, C., Taha, T.M., Asari, V.K.: Nuclei segmentation with recurrent residual convolutional neural networks based U-Net (R2U-Net). In: IEEE National Aerospace and Electronics Conference, pp. 228–233. IEEE (2018)
2. Arganda-Carreras, I., et al.: Crowdsourcing the creation of image segmentation algorithms for connectomics. Front. Neuroanat. **9**, 142 (2015)
3. Chaurasia, A., Culurciello, E.: LinkNet: exploiting encoder representations for efficient semantic segmentation. In: IEEE Visual Communications and Image Processing (VCIP), pp. 1–4. IEEE (2017)
4. Chen, Y., Shi, F., Christodoulou, A.G., Xie, Y., Zhou, Z., Li, D.: Efficient and accurate mri super-resolution using a generative adversarial network and 3D multi-level densely connected network. In: Frangi, A., Schnabel, J., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11070, pp. 91–99. Springer, Heidelberg (2018). https://doi.org/10.1007/978-3-030-00928-1_11
5. Dong, C., Loy, C.C., Tang, X.: Accelerating the super-resolution convolutional neural network. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9906, pp. 391–407. Springer, Heidelberg (2016). https://doi.org/10.1007/978-3-319-46475-6_25
6. Graham, S., et al.: Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. Med. Image Anal. (MIA) **58**, 101563 (2019)
7. Guo, Q., Yu, Z., Wu, Y., Liang, D., Qin, H., Yan, J.: Dynamic recursive neural network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5147–5156 (2019)
8. Han, W., Chang, S., Liu, D., Yu, M., Witbrock, M., Huang, T.S.: Image super-resolution via dual-state recurrent networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1654–1663 (2018)

9. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778 (2016)
10. Hou, L., Agarwal, A., Samaras, D., Kurc, T.M., Gupta, R.R., Saltz, J.H.: Robust histopathology image analysis: to label or to synthesize? In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 8533–8542 (2019)
11. Ioffe, S., Szegedy, C.: Batch normalization: accelerating deep network training by reducing internal covariate shift. In: International Conference on Machine Learning (ICML), pp. 448–456 (2015)
12. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. In: International Conference on Learning Representations (ICLR) (2015)
13. Kumar, N., Verma, R., Sharma, S., Bhargava, S., Vahadane, A., Sethi, A.: A dataset and a technique for generalized nuclear segmentation for computational pathology. IEEE Trans. Med. Imaging (TMI) **36**(7), 1550–1560 (2017)
14. Ledig, C., et al.: Photo-realistic single image super-resolution using a generative adversarial network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4681–4690 (2017)
15. Lee, K., Zlateski, A., Ashwin, V., Seung, H.S.: Recursive training of 2D–3D convolutional networks for neuronal boundary prediction. In: Advances in Neural Information Processing Systems (NeurIPS), pp. 3573–3581 (2015)
16. Mehta, R., Sivaswamy, J.: M-net: A convolutional neural network for deep brain structure segmentation. In: 14th International Symposium on Biomedical Imaging (ISBI), pp. 437–440. IEEE (2017)
17. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: fully convolutional neural networks for volumetric medical image segmentation. In: 4th International Conference on 3D Vision (3DV), pp. 565–571. IEEE (2016)
18. Nair, V., Hinton, G.E.: Rectified linear units improve restricted boltzmann machines. In: Proceedings of the 27th International Conference on Machine Learning (ICML), pp. 807–814 (2010)
19. Naylor, P., Laé, M., Reyal, F., Walter, T.: Segmentation of nuclei in histopathology images by deep regression of the distance map. IEEE Trans. Med. Imaging (TMI) **38**(2), 448–459 (2018)
20. Oktay, O., et al.: Attention U-net: learning where to look for the pancreas. In: 1st Conference on Medical Imaging with Deep Learning (MIDL) (2018)
21. Raza, S.E.A., et al.: Micro-net: a unified model for segmentation of various objects in microscopy images. Med. Image Anal. (MIA) **52**, 160–173 (2019)
22. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
23. Sui, Y., Afacan, O., Gholipour, A., Warfield, S.K.: Isotropic MRI super-resolution reconstruction with multi-scale gradient field prior. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11766, pp. 3–11. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32248-9_1
24. Wang, W., Yu, K., Hugonot, J., Fua, P., Salzmann, M.: Recurrent U-net for resource-constrained segmentation. In: The IEEE International Conference on Computer Vision (ICCV) (2019)
25. Xia, X., Kulis, B.: W-net: a deep model for fully unsupervised image segmentation. arXiv preprint arXiv:1711.08506 (2017)

26. Zhang, C., et al.: Ms-GAN: GAN-based semantic segmentation of multiple sclerosis lesions in brain magnetic resonance imaging. In: 2018 Digital Image Computing: Techniques and Applications (DICTA), pp. 1–8. IEEE (2018)
27. Zhang, C., Song, Y., Zhang, D., Liu, S., Chen, M., Cai, W.: Whole slide image classification via iterative patch labelling. In: 25th IEEE International Conference on Image Processing (ICIP), pp. 1408–1412. IEEE (2018)
28. Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N., Liang, J.: UNet++: a nested U-Net architecture for medical image segmentation. In: Stoyanov, D., et al. (eds.) DLMIA/ML-CDS -2018. LNCS, vol. 11045, pp. 3–11. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00889-5_1