U-Net: Convolutional Networks for Biomedical Image Segmentation

Summary written by Nicholas Hinke
April 05, 2022

Summary: As demonstrated throughout the years prior to 2015, the importance of network *depth* had been growing clearer and clearer [5, 10]. However, networks of this size and complexity typically required millions of training samples in order to perform at "state-of-the-art" levels. For many classes of problems–such as those within the field of biomedical image segmentation–attaining this enormous amount of labeled training data is either infeasible or impossible. In an attempt to address this problem, the authors of this paper presented a novel network architecture–which they denoted as *U-Net*–that built upon the motivating ideas behind other fully-convolutional networks [9] and leveraged data augmentation [2] in order to achieve success. Notably, these networks were able to perform successfully despite the severely constrained number of labeled training samples available for a given problem. Inspired by the then-recent work on other deep neural network architectures [3, 5, 10], sliding-window convolutional networks [1, 4, 8], and especially data augmentation [2] and fully-convolutional neural networks [9], the creators of U-Net later went on to win a 2015 International Symposium on Biomedical Imaging (ISBI) challenge, as well as post the best-ever achieved performance on a 2012 ISBI challenge (which was still accepting submissions at the time of publishing) [7].

Approach: As noted above, the motivation behind U-Net came from other fully-convolutional networks [9] which had shown great promise in the realm of data-constrained biomedical image segmentation. In that vein, the authors chose to structure their network architecture in the encoder/decoder style where the encoder side was known as "the contracting path" and the decoder side was known as "the expanding path". This was done deliberately, as the task of the contracting path was to capture local context within the image, while the expanding path enabled precise pixel-wise localization (*i.e.* it performed the segmentation). The key insight made by the authors was to utilize several skip connections from the contracting path to the expanding path; consequently, this provided valuable contextual information when performing localization. It is due to this insight as well as the authors' decision to include a high number of feature channels during upsampling within the expanding path that gave the network its characteristic symmetric (or "U-like") shape [7].

In order to combat the many problems associated with a lack of training data [2], the authors leveraged "excessive" data augmentation within U-Net [7]. As a result, the network could be trained end-to-end using far fewer labeled training samples. Indeed, data augmentation was performed explicitly by generating random elastic deformations on the few labeled samples that were actually available. Additionally, implicit data augmentation was realized through the use of dropout layers at the end of the contracting path [7].

In order to demonstrate the effectiveness of the U-Net architecture compared to other state-of-the-art methods at the time, the authors provided its performance statistics on two relevant ISBI challenges. First, U-Net achieved the best-ever performance on the then-open 2012 ISBI electron microscope segmentation challenge (when ranked by "warping error"). More impressively, U-Net very convincingly won the 2015 ISBI cell tracking challenge, winning both categories by more than 9% (PhC-U373) and more than 31% (DIC-HeLa), respectively [7].

Strengths: Most notably, the resulting U-Net as constructed by the authors outperformed all other previous best methods on the ISBI challenges, thus demonstrating the true potential of this fully-convolutional approach. Moreover, this paper yet again demonstrated the value and effectiveness of well-designed data augmentation, especially within data-constrained environments. Finally, the authors greatly influenced future related advances in the field, as they made their Caffe implementation and trained network parameters publicly available online [6].

Weaknesses: While the authors provide a relatively detailed overview of the architecture of U-Net, they provide little to no discussion on how or why the exact number or size of convolutional layers was chosen. In fact, there is very little discussion regarding any of the hyperparameter choices or tuning [7]. Additionally, there may be a potential for U-Net to miss some of the finer details during segmentation due to the nature of the utilized convolutional layers as feature extractors within the contracting path (since they will necessarily extract higher level features in the image).

Reflections: Since it was unclear in the paper how much hyperparameter tuning was done when designing U-Net, it may prove valuable to study the network's performance while varying some of those parameters (*e.g.* filter sizes, nonlinearities, *etc.*) [7]. Additionally, it may be quite advantageous to consider adding some other useful structural features of neural networks to U-Net such as residual blocks or inception modules. Finally, it would be very interesting to examine the runtime efficiency of U-Net on the modern hardware and GPUs of today.

References

[1] D. C. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber. Deep neural networks segment neuronal membranes in electron microscopy images. In *NIPS*, 2012.

[2] A. Dosovitskiy, J. T. Springenberg, M. A. Riedmiller, and T. Brox. Discriminative unsupervised feature learning with convolutional neural networks. In *NIPS*, 2014.

[3] R. B. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 580–587, 2014.

[4] B. Hariharan, P. Arbeláez, R. B. Girshick, and J. Malik. Hypercolumns for object segmentation and fine-grained localization. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 447–456, 2015.

[5] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60:84–90, 2012.

[6] O. Ronneberger. U-net: Convolutional networks for biomedical image segmentation. http://lmb.informatik.uni-freiburg.de/people/ronneber/u-net, 2015.

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

[8] M. Seyedhosseini, M. S. M. Sajjadi, and T. Tasdizen. Image segmentation with cascaded hierarchical models and logistic disjunctive normal networks. *2013 IEEE International Conference on Computer Vision*, pages 2168–2175, 2013.

[9] E. Shelhamer, J. Long, and T. Darrell. Fully convolutional networks for semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39:640–651, 2017.

[10] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556, 2015.