A Brief Report of "U-net: Convolutional Networks for Biomedical Image Segmentation"

Senbai Kang

March 15, 2019

U-net is a winning deep learning algorithm on the ISBI cell tracking challenge 2015 with overwhelming performance and efficiency in image segmentation. It mainly proposes solutions to three major problems in the field of biomedical image processing.

It is agreeable that a large amount of images is essential to successful deep learning. In practice, however, only a limited, small number of images is usually available, which poses the first problem to the application of deep learning in biomedical image processing. U-net employs data augmentation by applying elastic deformations to the images, which not only enlarges the available data, but has practical meaning that deformation is the most common variation in tissue.

The second problem is that, although convolutional neural networks are commonly used in classification of images, there is a need of segmenting different areas in an image into corresponding classes. In other words, instead of labeling the whole image of a specific class in classification, each pixel in an image is labeled of its corresponding class in segmentation. In order to localize, U-net has a structure of two symmetrical processes, i.e., contraction and expansion. The former is a typical combination of convolutional layers and max pooling layers, which extracts features of the input image while ignoring the information of location. By applying unsampling operators, the latter recovers the information of location, and creates high-resolution segmentation maps. Note that in the contraction process, all the convolutional layers only use the valid part of input data, which means that they are not padded and going to shrink during convolution. Meanwhile, U-net employs a strategy of overlap-tile, where prediction of the segmentation in a specific area in the input image requires a larger area with context information. If there is missing data in the larger area, it will be extrapolated by mirroring. Finally, dropout layers appear at the end of the contraction process performing additional implicit data augmentation.

The last challenge is the separation of connecting objects from the same class, which is extensively required in biomedical segmentation problems. U-net suggests the application of a weighed loss $w(\boldsymbol{x})$ in 1, which gives more importance to pixels of separating background between connecting cells in the loss function.

In the training process, the loss function of U-net is computed by a pixel-wise softmax over the final feature map:

$$E = \sum_{\boldsymbol{x} \in \Omega} w(\boldsymbol{x}) log(p_{l(\boldsymbol{x})}(\boldsymbol{x})) \tag{1}$$

where $l : \Omega \rightarrow \{1, \ldots, K\}$ is the true label of each pixel and:

$$w(\boldsymbol{x}) = w_c \boldsymbol{x} + w_0 \cdot \exp(-\frac{(d_1(\boldsymbol{x}) + d_2(\boldsymbol{x}))^2}{2\sigma^2}) \tag{2}$$

where $w_c : \Omega \rightarrow \mathbb{R}$ is the weight map to balance the class frequencies, $d_1 : \Omega \rightarrow \mathbb{R}$ is the distance to the border of the closest cell and $d_2 : \Omega \rightarrow \mathbb{R}$ is the distance to the border of the second closest cell. Setting $w_0 = 10$ and $\sigma = 5$ pixels.