U-Net Preliminary Summary

Shishir Jakati
University of Massachusetts, Amherst
sjakati@umass.edu

Abstract

Traditional constricting neural networks gradually reduce the overall resolution of their input. While this property may be useful for tasks such as classification, localization tasks suffer from the loss of spatial structure. U-Net [1] attempts to combat this resolution problem by utilizing two symmetric network paths. These symmetric paths are the namesake of the U-Net and are able to propagate context information for use in semantic segmentation.

1. Network Structure

The network is composed of two distinct network paths; a constricting path and an expansive path. This architecture supplements the usual contracting path by successive upsampling layers, replacing the constricting pooling layer.

The constricting path of the network is a rather traditional fully convolutional network. The input size is fixed at 572 * 572, which allows for tiles of larger images to be used as inputs. The successive constriction is performed by traditional 3x3 covolutional layers, followed by the ReLU non-linearity. Immediately following 2 convolutional constricting operations is maximum pooling layer with kernel size 2x2 with stride two.

The expansive path of the network consists of upsampled features and copied crops from intermediate layers of the constricting path. Each block within the expansive path is composed of traditional convolutional layers followed by the ReLU non-linearity, and 2x2 upconvolutional layers with a stride of 2. These upconvolutional layers, when used with the crop copy operation increase the resolution of the output.

2. Training

As a segmentation model, U-Net was trained on medical microscopy image segmentation maps. Stochastic gradient descent was used as the optimizer on the loss function.

2.1. Loss Function

The original microscopy data is comprised of two components, segmentation and classification of the various cells within the images. Thus, a pixel-wise softmax loss function is applied over the final feature map, and a cross entropy loss penalizes at each position the deviation from the ground truth class.

The pixel-wise softmax is defined as follows:

$$p_k(\mathbf{x}) = \frac{exp(a_k(\mathbf{x}))}{\sum_{k'=1}^{K} exp(a_{k'}(\mathbf{x}))}$$

Here, $a_k(\mathbf{x})$ is the activation in channel k, given that the number of classes is K. $\mathbf{x}$ is the pixel location.

The cross entropy loss is defined as follows:

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

Here, l is the true label of the pixel value, and Ω is the set of pixel positions. w is a weight map introduced to give higher importance to border separation pixels. This weight map is important to normalize the variance in number of examples for each class. The weight map is computed as follows:

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

Here, w_c is the weight map used to balance the frequencies of different classes in the training set, d_1, d_2 denote the distance to border of the first and second nearest cells, respectively.

2.2. Data Augmentation

Due to the small size of the training dataset, data augmentation was used to achieve a more robust model. Specifically, the microscopy images were cropped, as well as shifted and rotated. Another data augmentation technique used was elastic deformations. Additionally, Dropout layers were used within the model architecture to obtain implicit data augmentation.

3. Relevance to Research

The U-Net architecture is proven to obtain favorable results on segmentation tasks, and fine-grained classification tasks. When considering scene text localization, these high resolution activation maps may be used to produce high fidelity bounding boxes surrounding characters. It may be highly attuned to the problem of small character localization due to the fact that context regarding higher-level features is provided directly to layers with access to lower-level features.

References

[1] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation, 2015. 1