Review: U-Net (Biomedical Image Segmentation)

SH Tsang Follow

Nov 5, 2018 · 5 min read

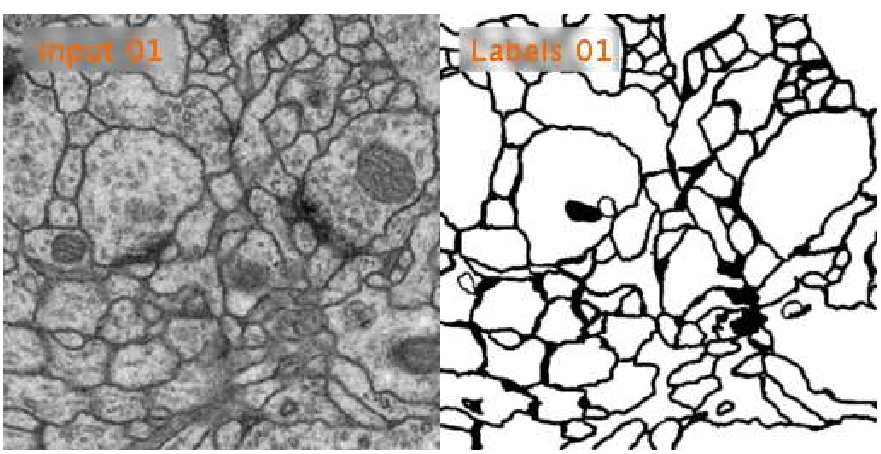
In this story, **U-Net** is reviewed. U-Net is one of the famous Fully Convolutional Networks (FCN) in biomedical image segmentation, which has been published in **2015 MICCAI** with more than **3000 citations** while I was writing this story. (SH Tsang @ Medium)

In the field of biomedical image annotation, we always **need experts**, who acquired the related knowledge, to annotate each image. And they also **consume large amount of time to annotate. If the annotation process becomes automatic, less human efforts and lower cost** can be achieved. Or it can be act as a assisted role to **reduce the human mistake**.

> *You may ask: "Is it too narrow to read about biomedical Image Segmentation?"*

> *However, we may learn the techniques of it, and apply it to different industries. Say for example, **quality control / automatic inspection / automatic robotics during construction / fabrication / manufacturing process**, or any other stuffs we may think of. **These activities involve quantitative diagnosis. If we can make it automatic, cost can be saved with even higher accuracy.***

In this paper, they **segment/annotate the Electron Microscopic (EM) Image**. They also modify the network a little bit to **segment/annotate the dental X-ray image in 2015 ISBI**.



EM Images

What Are Covered

A. EM Image Segmentation

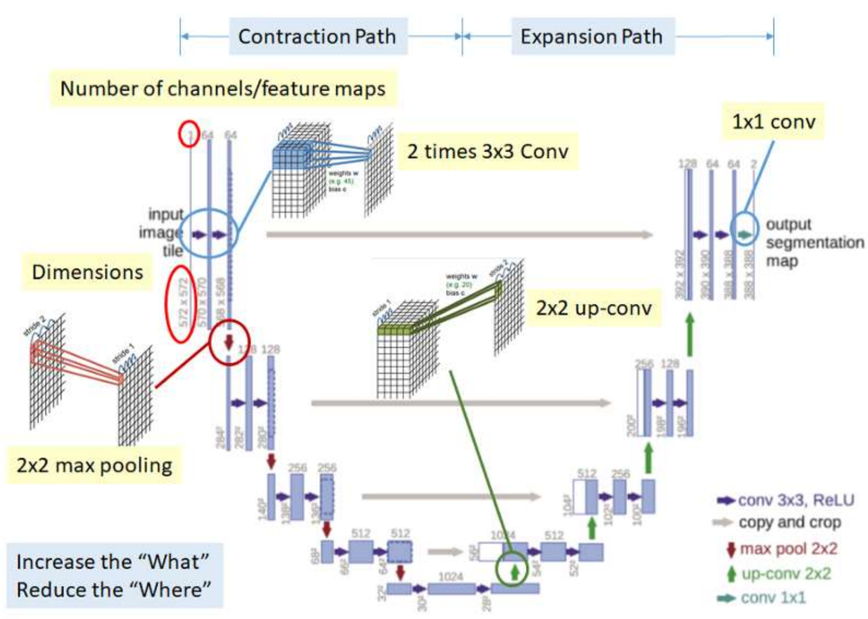
1. U-Net Network Architecture
2. Overlap Tile Strategy
3. Elastic Deformation for Data Augmentation
4. Separation of Touching Objects
5. Results

B. Dental X-Ray Image Segmentation

1. Some Modifications of U-Net
2. Results

. . .

A.1. U-Net Network Architecture



U-Net

The U-net architecture is as shown above. It consists of contraction path and expansion path.

Contraction path

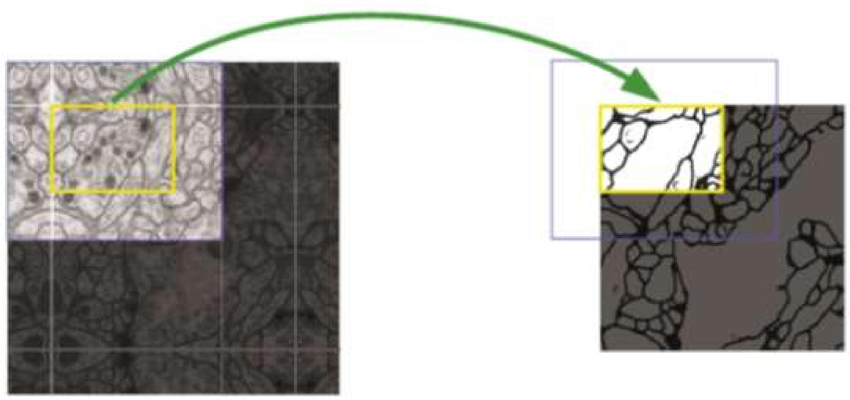
- Consecutive of **two times of 3×3 Conv** and **2×2 max pooling** is done. This can help to extract more advanced features but it also reduce the size of feature maps.

Expansion path

- Consecutive of **2×2 Up-conv** and **two times of 3×3 Conv** is done to recover the size of segmentation map. However, the above process **reduces the "where"** though it **increases the "what"**. That means, we can get advanced features, but we also loss the localization information.
- Thus, after each up-conv, we also have **concatenation of feature maps (gray arrows) that are with the same level**. This helps to **give the localization information from contraction path to expansion path**.
- At the end, **1×1 conv** to map the feature map size from 64 to 2 since the output feature map only have 2 classes, cell and membrane.

. . .

A.2. Overlap Tile Strategy

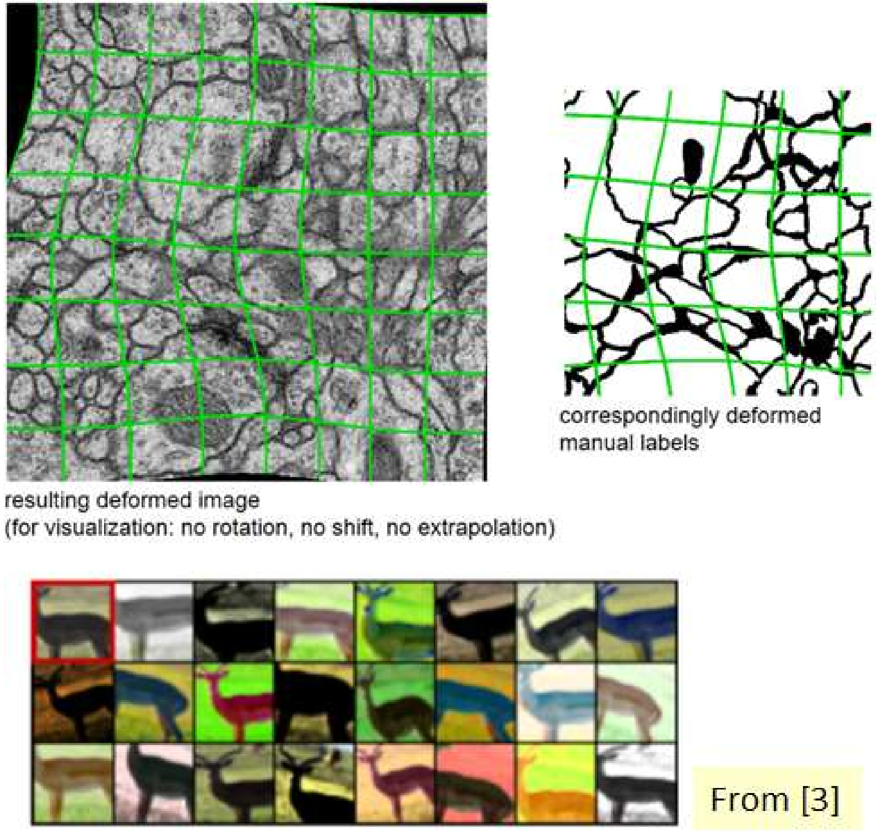


Overlap Tile Strategy

Since unpadded convolution is used, output size is smaller than input size. Instead of downsizing before network and upsampling after network, overlap tile strategy is used. Thereby, **the whole image is predicted part by part** as in the figure above. The yellow area in the image is predicted using the blue area. **At the image boundary, image is extrapolated by mirroring**.

. . .

A.3. Elastic Deformation for Data Augmentation

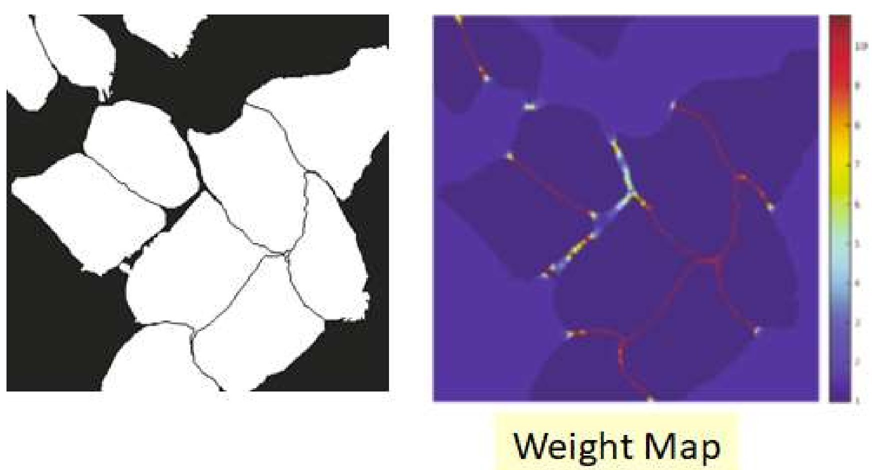


Elastic Deformation

Since the training set can only be annotated by experts, the training set is small. To increase the size of training set, data augmentation is done by randomly deformed the input image and output segmentation map.

. . .

A.4. Separation of Touching Objects



Segmentation Map (Left) and Weight Map (Right)

Since the touching objects are closely placed each other, they are easily merged by the network, to separate them, a weight map is applied to the output of network.

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

To compute the weight map as above, d1(x) is the distance to the nearest cell border at position x, d2(x) is the distance to the second nearest cell border. Thus, at the border, weight is much higher as in the figure.

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

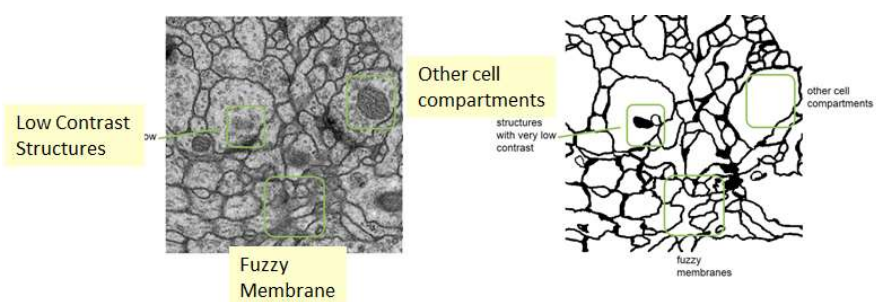
$$E = \sum_{\mathbf{x} \in \Omega} w(\mathbf{x}) \log(p_{\ell(\mathbf{x})}(\mathbf{x}))$$

Thus, the cross entropy function is penalized at each position by the weight map. And it help to **force the network to learn the small separation borders between touching cells.**

. . .

A.5. Results

A.5.1. ISBI 2012 Challenge



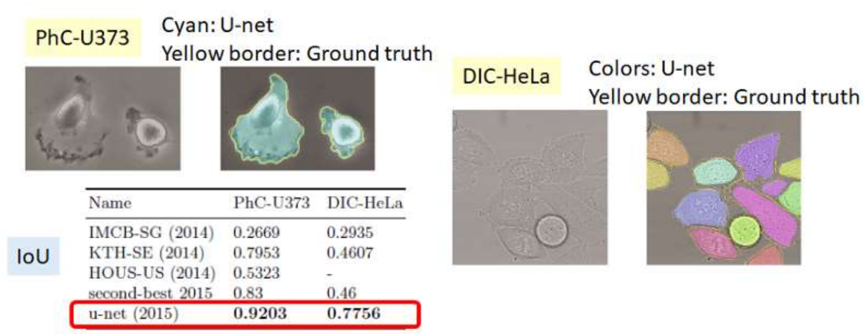
Some Difficult Parts in EM Images

Rank	Group name	Warping Error	Rand Error	Pixel Error
	** human values **	0.000005	0.0021	0.0010
1.	u-net	**0.000353**	0.0382	0.0611
2.	DIVE-SCI	0.000355	0.0305	0.0584
3.	IDSIA [2]	0.000420	0.0504	0.0613
4.	DIVE	0.000430	0.0545	**0.0582**
⋮				
10.	IDSIA-SCI	0.000653	**0.0189**	0.1027

U-Net has the Rank 1 Result at that moment

- **Warping Error**: A segmentation metric that penalizes topological disagreements.
- **Rand Error**: A measure of similarity between two clusters or segmentations.
- **Pixel Error**: A standard pixel-wise error.
- Training Hour: 10 Hours
- Testing speed: around 1s per image

A.5.2. PhC-U373 and DIC-HeLa Datasets



Name	PhC-U373	DIC-HeLa
IMCB-SG (2014)	0.2669	0.2935
KTH-SE (2014)	0.7953	0.4607
HOUS-US (2014)	0.5323	-
second-best 2015	0.83	0.46
u-net (2015)	**0.9203**	**0.7756**

PhC-U373 and DIC-HeLa Datasets

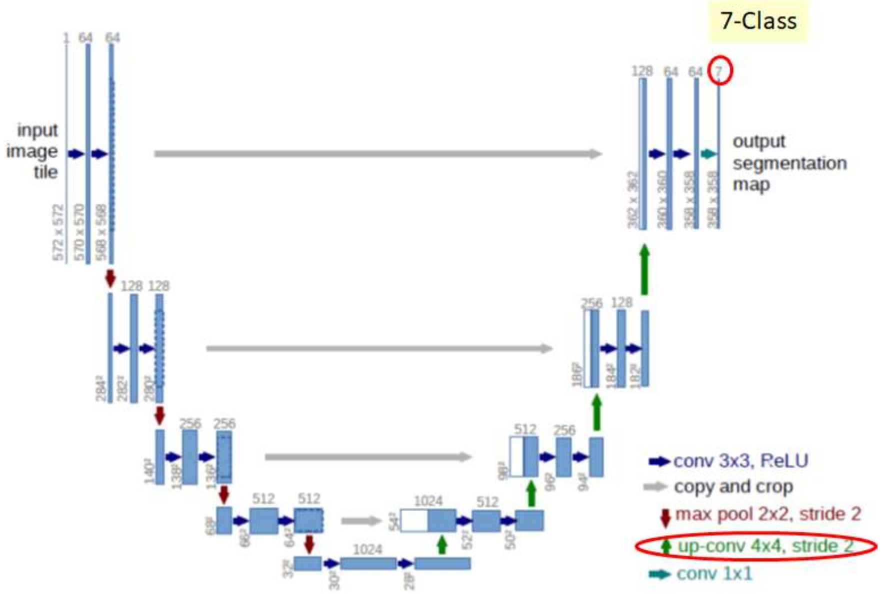
U-Net got the highest IoU for these two datasets.

. . .

B.1. Some Modifications of U-Net

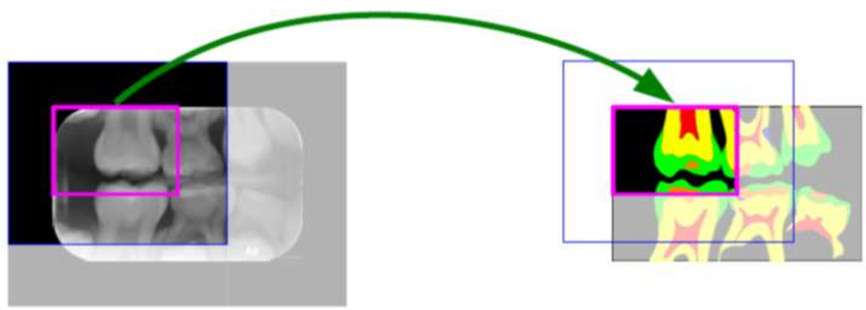
No.	Important Properties Parts
1	caries (blue color)
2	enamel (green color)
3	dentin (yellow color)
4	pulp (red color)
5	crown (skin color)
6	restoration (orange color)
7	root canal treatment (cyan color)

Dental X-Ray Image with 7 classes



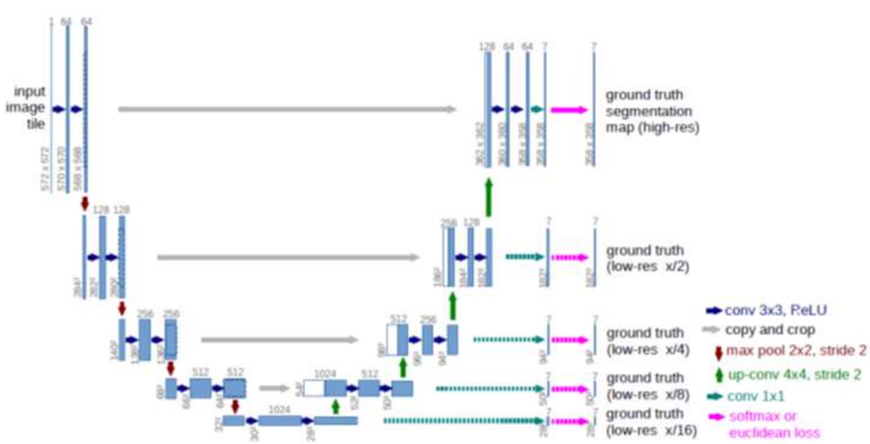
U-Net for Dental X-Ray Images

This time, **4×4 Up-conv** is used, and **1×1 Conv** to map feature maps from 64 to 7 because the output for each location has 7 classes.



Zero padding instead of mirroring at the image boundary

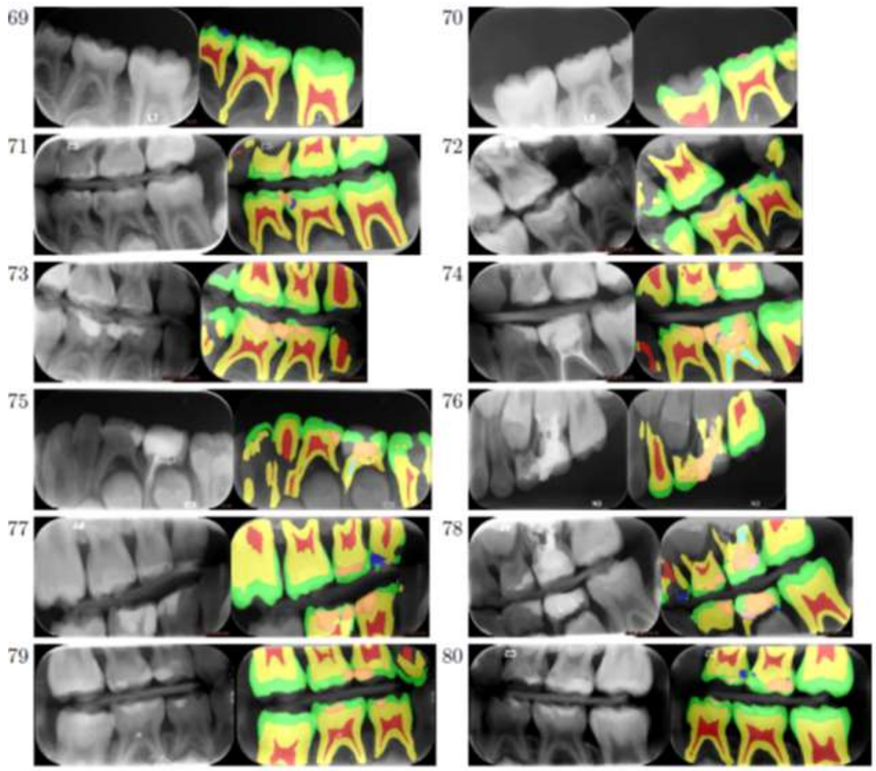
At the Overlap Tile Strategy, **zero padding** is used instead of mirroring at the image boundary. Because mirroring isn't making any sense for teeth.



Loss function at multiple levels

There are **additional loss layers** to the low-resolution feature maps using softmax loss, in order to guide the deep layers to directly learn the segmentation classes.

B.2. Results



Some Visualization Results

• • •

I have also reviewed **CUMedVision1** and **CUMedVision2**. Please feel free to visit if interested.

References

- [2015] [MICCAI]
  U-Net: Convolutional Networks for Biomedical Image Segmentation
- [2015] [ISBI]
  Dental X-ray Image Segmentation using a U-shaped Deep Convolutional Network

My Related Reviews

[CUMedVision1] [CUMedVision2] [FCN] [DeconvNet]