Meeting Note

Author: Henry Lin
Date: 2024/08/19 – 2024/08/23

Weekly Summary
U-Net [1].

Plan for Next Week
Denoising Diffusion Probabilistic Models (arXiv: 2006.11239).

Details
To address two problems: (1) the lack of labeled data and (2) the lack of computational power (GPU is still not so powerful in 2015), on the task of image segmentation, Ronneberger et al. [1] proposed an efficient end-to-end architecture based on the "fully convolution networks" [3] which is used to generate the correspondingly-sized segmentation, called U-Net. Traditional convolution neural network (CNN) usually output a label instead of another image, although the SOTA model in 2012 [2] can perform well on this task, it still has the drawbacks (e.g. redundancy due to overlapping and loss of global context).

In Figure 1, the left side called contracting path is similar to the traditional CNN, consisting of 3x3 convolution, activation function (i.e. ReLU), and 2x2 max pooling; and the right side called expansive path uses 2x2 up-convolution which halves the channel size and doubles the x-y size, 3x3 convolution, and ReLU, finally, a 1x1 convolution layer is applied to convert the 64-channel output to a desired number of channels.

In the up-sampling progress, a feature map with same size will be copied from corresponded down-sampling convolution layer and concatenated to the up-scaled feature map (i.e. the "copy and crop" in Figure 1), by doing so, the information can be remained and used to generate higher resolution images.

During the training step, the pixel-wise softmax over the final feature map, cross entropy loss function, and weight loss map are applied. The weight loss map is used to force the network to learn the border pixels, and pre-computed using

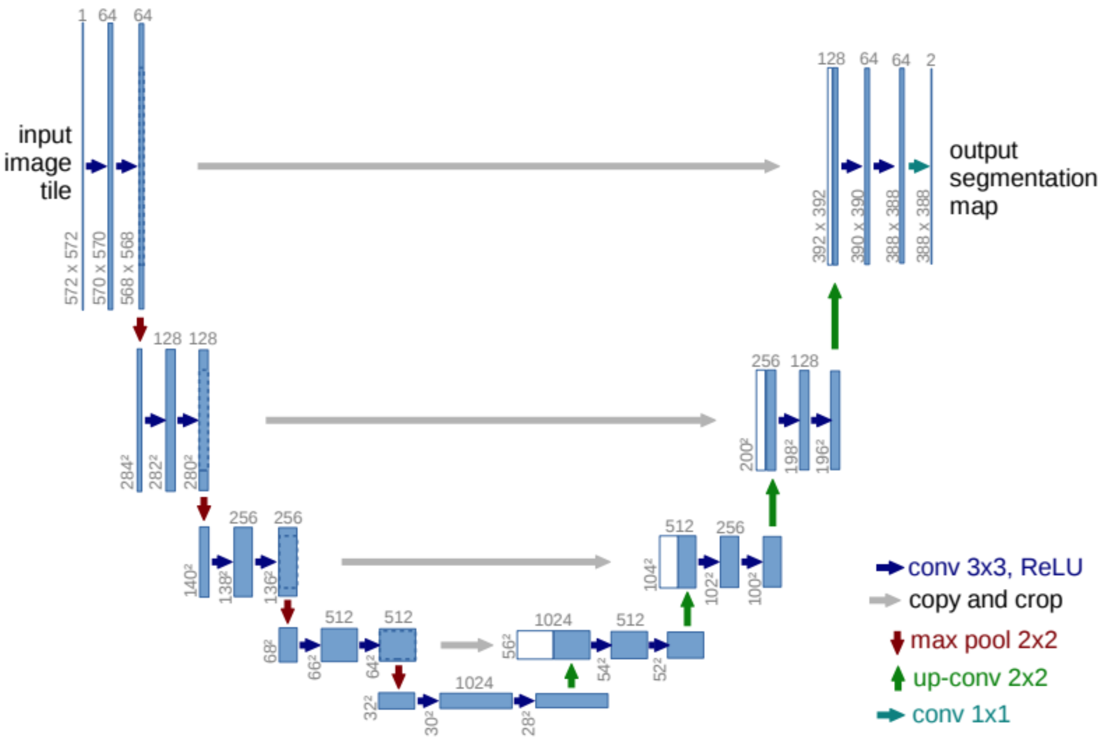

Figure 1. The overall architecture of U-Net. The number on the top of box denotes the number of channels.

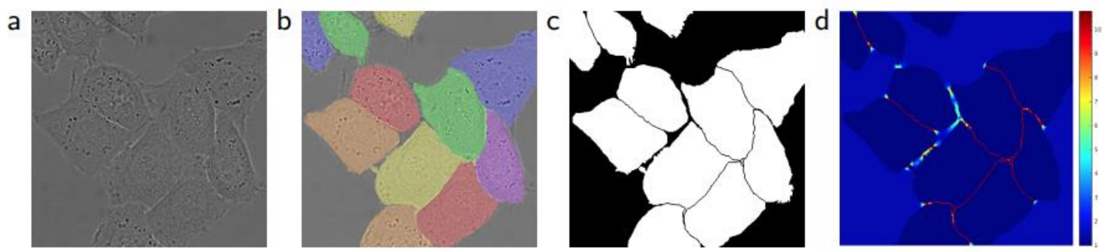

Figure 2. A sample of training data. (a) raw image. (b) ground truth segmentation (label). (c) segmentation mask generated by U-Net. (d) weight loss map.

morphological operations as follow:

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

where ω_c is the weight map to balance the class frequencies; d_1 and d_2 denote the distances to the nearest border and second nearest border, respectively; and $\omega_0 = 10,\ \sigma \approx 5$ pixels.

In addition to shifting and rotation, random elastic deformation is also used as the data augmentation method, making U-Net to achieve great performance with very few labeled data.

Table 1. Experimental results on EM segmentation challenge [5] sorted by warping error.

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 [1]	0.000420	0.0504	0.0613
4.	DIVE	0.000430	0.0545	**0.0582**
⋮				
10.	IDSIA-SCI	0.000653	**0.0189**	0.1027

Table 2. Experimental results on ISBI cell tracking challenge 2015 [6].

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**

The Table 1 shows that U-Net achieve the new best score of 0.0003529 on warping error (see remarks) without any preprocessing or postprocessing. The Table 2 shows the Intersection over Union (IoU) results of cell segmentation tasks, U-Net also get the best scores of 92% and 77.5%.

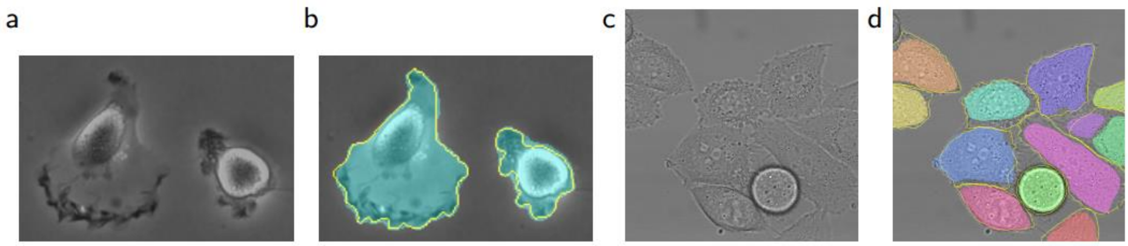

Figure 3. Visualized result of cell segmentation results. (a, b) a sample from "PhC-U373" data set. (c, d) a sample from "DIC-HeLa" data set.

Remarks

"Elastic deformation" is a technique that artificially distorts an image by applying random displacements to its pixels. This process simulates real-world deformations, creating new, varied training data. It's commonly used in image processing and machine learning to improve model performance and prevent overfitting.

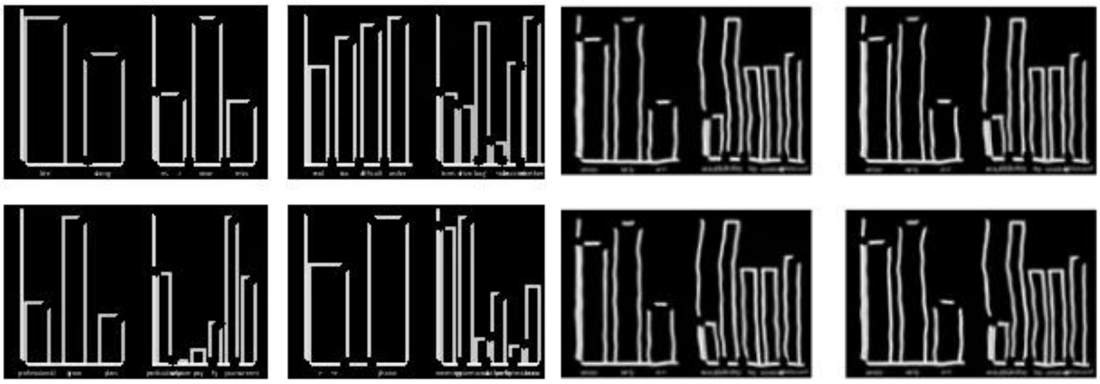

Figure 4. Images before (left) and after (right) elastic deformation [4].

The "DIVE-SCI" [7], "DIVE", and "IDSIA-SCI" in Table 1 are methods of segmentation, but the sources are not mentioned in this paper. In addition, you can see that "DIVE-SCI" performances almost as well as U-Net on warping error, and better than U-Net on the other two metrics.

Warping Error: This metric attempts to address a limitation of pixel error. Pixel error treats all misclassified pixels equally, regardless of their location. Warping error, however, focuses on discrepancies at the boundaries between objects. The idea is that minor errors along object borders might be less critical than large misclassifications within an object.

Rand Error (RE): This metric was originally proposed to measure the similarity of two data clusters, it takes a more holistic view of segmentation accuracy. It considers the entire image as a collection of data points (pixels) and compares how similar the predicted clustering of these points is to the ground truth clustering. A lower Rand Error signifies a better match between the predicted and actual segmentation.

Pixel Error: This is the most straightforward metric. It simply calculates the difference between each pixel's predicted class label and the corresponding label in the ground truth (actual data).

References

[1] RONNEBERGER, Olaf; FISCHER, Philipp; BROX, Thomas. U-net: Convolutional networks for biomedical image segmentation. In: *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*. Springer International Publishing, 2015. p. 234-241.

[2] CIRESAN, Dan, et al. Deep neural networks segment neuronal membranes in electron microscopy images. *Advances in neural information processing systems*, 2012, 25.

[3] LONG, Jonathan; SHELHAMER, Evan; DARRELL, Trevor. Fully convolutional networks for semantic segmentation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015. p. 3431-3440.

[4] V. YILMAZ, "Elastic Deformation on Images." medium.com. Accessed: Aug. 9, 2024. [Online.] Available: https://towardsdatascience.com/elastic-deformation-on-images-b00c21327372

[5] WWW: Web page of the em segmentation challenge, http://brainiac2.mit.edu/isbi_challenge/

[6] WWW: Web page of the cell tracking challenge, http://www.codesolorzano.com/ celltrackingchallenge/Cell_Tracking_Challenge/Welcome.html

[7] FAKHRY, Ahmed; PENG, Hanchuan; JI, Shuiwang. Deep models for brain EM image segmentation: novel insights and improved performance. *Bioinformatics*, 2016, 32.15: 2352-2358.

[8] Neville, "Warping Error, Rand Error and Pixel Error in Semantic Segmentation", medium.com. Accessed: Aug. 9, 2024. [Online.] Available: https://cvinvolution.medium.com/warping-error-rand-error-and-pixel-error-in-semantic-segmentation-5cd561ba99cb