
U-RISC Cell Edge Detection

Xinran Dong¹
2301210603
Peking University

Abstract

Nowadays the demand of higher accuracy of cell edge detection is pressing for biological and medical research, especially for neurosciences. However, the performance of AI is far behind the human-level in cell edge detection task. Based on U-RISC dataset¹, we advance a new method which combines ensemble learning with morphological processing. We implement input preprocessing, ensemble learning and postprocessing in this work and eventually achieve excellent results: the Simple-Track score is **0.55667** and the Complex-Track score is **0.58794**. Our code is available at <https://github.com/007DXR/U-RISC-Cell-Edge-Detection>.

1. Introduction

U-RISC dataset is an annotated Ultra-high Resolution Image Segmentation dataset for cell membrane. U-RISC dataset is the largest cell membrane annotated Electron Microscopy (EM) dataset with a resolution of 2.18 nm/pixel, is given in the U-RISC Neuron Recognition Competition. The task of cell edge detection aims at both detecting salient edge and recognizing their categories, or more concretely, locating fine edges utilizing low-level features and identifying whether they are inside the neurons or on the boundary with abstracted high-level features. However, in the face of massive EM data, traditional approaches may introduce a lot of uncertainty, so robust detection of cell membrane boundary is very necessary for both medical and biological research, which means neither missing the boundary of the cell membrane, nor incorrectly detecting the boundary between the cell nucleus and biological tissue.

Biological cell membrane recognition is faced with two major difficulties. One is that in medical practice, cell photos taken by electron microscope are affected by illumination and shooting angle, resulting in great changes of image shape. Second, very little training data available will lead to over-fitting of the model. Consequently, we use excessive

data augmentation by applying elastic deformations to the available training images. This allows the network to learn invariance to such deformations, without the need to see these transformations in the annotated image corpus. This is particularly important in biomedical segmentation, since deformations used to be the most common variation in tissue and realistic deformations can be simulated efficiently. The value of data augmentation for learning invariance has been shown in Dosovitskiy et al. in the scope of unsupervised feature learning. We also adopt a series of physical and chemical transformations and denoising method, such as RandomResizedCrop and RandomBrightnessContrast, to eliminate the influence of large areas of mislabeling and small noise points and contrast difference.

Inspired by ensemble learning, we implement the ensemble model and compare it with the pure structure. The result has shown a significant improvement, with lower variance and higher accuracy. After the ensemble part, morphological processing methods such as opening and closing operation are also adopted to improve precision at the expense of slightly sacrificing recall, thus achieving a higher F-score. Sufficient experiments in both Simple-Track and Complex-Track demonstrate the effectiveness of our method

Our contributions are summarized as follows:

- We are the first to use a combination of neural networks and morphological processing to do EM cell edge detection.
- We use strong data augmentation to avoid overfitting in the small dataset, model ensemble to improve robustness which is very important in the medical field and morphological processing to achieve a higher F-score.
- We do experiments on both tracks and get the first place in Simple-Track and the second place in Complex-Track (Simpler-Track with less data is actually harder).

2. Related Work

The definition of boundary or edge detection has evolved over time from low-level to high-level features as simple edge filters, depth edges, object boundaries and semantic contours. In some sense, the evolution of edge detection

¹<https://www.biendata.xyz/competition/urisc/>

algorithms captures the progress in computer vision from simple convolutional filters such as Sobel or Canny(Canny, 1986) to fully developed deep neural networks.

2.1. Low-level Edges

Early edge detection methods used simple convolutional filters such as Sobel or Canny(Canny, 1986). Hough transform and morphological processing such as opening and closing operation are also commonly used as post-processing methods.

2.2. Deep Learning Method

Nowadays, with the improvement of computer computing power and rich training data, deep neural network has been applied to image segmentation and edge detection. A large number of advanced image segmentation and edge detection network structures have been proposed in recent years, such as U-Net(Ronneberger et al., 2015), LinkNet(Chaurasia & Culurciello, 2017), CASENet(Yu et al., 2017), DDS(Liu et al., 2018), DFF(Hu et al., 2019) and so on. Deep learning far exceeds traditional methods in semantic segmentation and edge detection.

2.2.1. U-NET

U-Net(Ronneberger et al., 2015) is built upon a more elegant architecture, the so-called *fully convolutional network*. In the unsampling part it has a large number of feature channels, which allows the network to propagate context information to higher resolution layers. It proposes the use of a weighted loss, where the separating background labels between touching cells obtain a large weight in the loss function. It is applicable to various biomedical segmentation problems.

2.2.2. CASENET

While classical edge detection is a challenging binary problem in itself, the category-aware semantic edge detection by nature is an even more challenging multi-label problem. CASENet(Yu et al., 2017) proposes a novel end-to-end deep semantic edge learning architecture based on ResNet(He et al., 2016) and a new skip-layer architecture where category-wise edge activation at the top convolution layer share and are fused with the same set of bottom layer features.

2.2.3. DFF

Features from multiple scales can greatly benefit the semantic edge detection task if they are well fused. DFF(Hu et al., 2019) proposes a novel dynamic feature fusion strategy that assigns different fusion weights for different input images and locations self-adaptively. It is superior to fixed weight fusion and also the naïve location-invariant weight fusion methods.

3. Methods

3.1. Pre-processing

3.1.1. LABEL DENOISING

Like other datasets, the raw data (label) in the U-RISC dataset contains a lot of noise, including large areas of distinguished mislabeling and small noise points. Since data quality determines the upper bound of model performance, we denoise the label images before training. The subsequent experiments confirm the view that denoising process is effective. Ablation experiments on the effect of label denoising are listed in the appendix.

We unfortunately observes that morphological method of denoising is limited because some ground-truth images are labelled mistakenly on purpose. We cope with these images manually. For instance, wrong labels and noise points with small local area, in which connected domain size is less than 40, are removed directly.

3.1.2. IMAGE AUGMENTATION

Due to the lack of training data of U-RISC dataset, especially in Simple-Track (only 30 training pictures in 1024*1024 size), we adopt a series of physical and chemical transformation methods for strong data augmentation in order to avoid overfitting. Transformation methods are showed in the Table 1.

Transformation	Method
Physical	RandomResizedCrop,HorizontalFlip, VerticalFlip,RandomRotate90
Chemical	RandomBrightnessContrast, MultiplicativeNoise
Medical	ElasticTransform,GridDistortion, OpticalDistortion,CoarseDropout

Table 1. Transformation methods for Image Augmentation

Ablation experiments on the effectiveness of strong image augmentation are listed in the appendix. The implementation details for all of the above transformations can be found in *albumentations*.

3.2. Network Architecture

We have tried different state-of-the-art image segmentation and edge detection network architectures, including U-Net(Ronneberger et al., 2015), ResNetUNet, LinkNet(Chaurasia & Culurciello, 2017), DDS(Liu et al., 2018) and so on. We find that DFF(Hu et al., 2019) and CASENet(Yu et al., 2017) perform best after comparison the effect among all of network models in U-RISC dataset. We also use model ensemble of the same architecture to improve robustness.

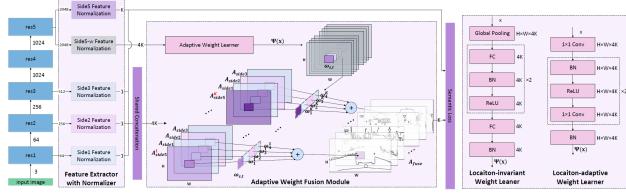


Figure 1. DFF Architecture

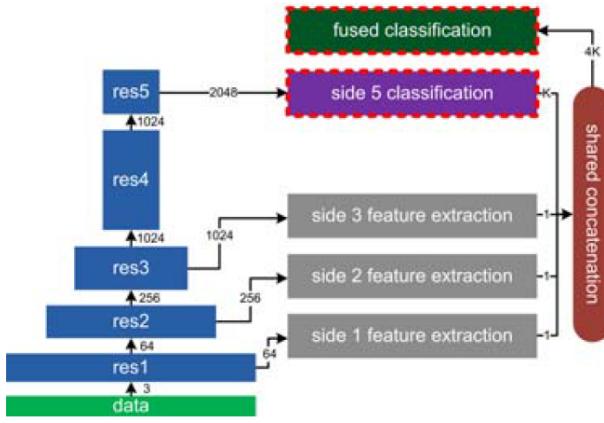


Figure 2. CASENet Architecture

3.2.1. SIMPLE-TRACK

In Simple-Track, the images contain fewer cells, but the labels on the cell edges are extremely fine, only 1-2 pixels. In order to obtain finer segmentation edges, we choose DFF (Fig. 1) which has an adaptive weighting mechanism.

3.2.2. COMPLEX-TRACK

Since the labels on the cell edges in Complex-Track are relatively thick (compared with Simple-Track), we choose CASENet (Fig. 2), a relatively simple structure, for faster training process without affecting the final performance.

3.3. Loss Function

Like other binary classification problems, we use Focal BCELoss to solve the imbalance between positive and negative samples. In order to get a higher F-score, we add Dice Loss (F-score Loss), the derivable version of F-score, directly to our loss function. Finally, we adopt Near Edge Loss proposed in U-Net to pay more attention to the edges of cells. Our total loss function can be expressed as:

$$\text{Focal BCELoss} + \text{Dice Loss} + \lambda \times \text{Near Edge Loss}$$

3.4. Post-processing

During our initial training process, we found that the precision of model's prediction was always lower than the recall.

According to the calculation method of F-score:

$$\text{F-score} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

a better balance between the two could improve the model's performance.

Therefore, after the model ensemble, morphological processing methods such as opening and closing operation are adopted to improve precision at the expense of slightly sacrificing recall, thus achieving a higher F-score.

However, with the improvement of our method, the precision of model's prediction has been greatly improved, and the precision and recall both exceed 0.6. In this case, the effect of morphological processing is not obvious, and may even bring negative benefits.

3.5. Training

During training, we use automatic mixed precision (AMP) to reduce GPU memory occupation and speed up. We train DFF with 2 GTX 3090 GPUs for about 5h in Simple-Track, and CASENet for about 40h in Complex-Track.

4. Experiments

4.1. Simple-Track

The data set of Simple track has a small number of cells, small image size, low resolution (1024×1024), and a small number of cell membrane pixels to be judged. The labels on the cell edges are very fine (only 1-2 pixel width).

So in Simple-Track, we choose DFF as our network and pretrained ResNet-50 as backbone, other hyperparameters are listed in the appendix. After 500 epochs of training, our method has achieved an F-score of 0.51182 on the validation dataset and 0.55667 on the test dataset (the result on test dataset is tested by BAAI), which has won the fourth place in the original U-RISC Competition (Table. 2).

Ranking	Research Institution	F-score
	Human 1st	0.96915 ± 0.014
	Human 2nd	0.99334 ± 0.008
1	UCAS	0.56932 ± 0.053
2	NJU	0.56213 ± 0.055
3	HDU	0.56136 ± 0.049
4	UCAS	0.55170 ± 0.046
5	THU	0.55103 ± 0.047
6	SCU	0.54847 ± 0.053
-	Ours	0.55667

Table 2. Simple-Track Result of U-RISC Competition

4.2. Complex-Track

The data set of Complex track contains a large number of cells, large image size, high resolution (original electron microscope size of 10000×10000 , processed to 9958×9959), and a large number of cell membrane pixels to be judged. But the cell edges are relatively coarse in this track.

In Complex-Track, we choose a relatively simple CASENet as our network but a more complex pretrained ResNet-152 as backbone, in order to extract more effective features from higher resolution images, other hyperparameters are also listed in the appendix. After 100 epochs of training, our method has achieved an F-score of 0.64880 on the validation dataset and 0.58794 on the test dataset (the result on test dataset is tested by BAAI).

4.3. Ablation

We also perform a series of ablation experiments to demonstrate the effectiveness of our method, details of which are listed in the appendix.

5. Conclusion

Although our neural network performs well so far, there is still a lot of room for improvement. First of all, considering the prior knowledge of nerve cell membrane, we can add some morphological methods in the stage of pretreatment to help neural work extract features better and reduce the learning difficulty. Secondly, we can use two-stage or cascade network structure to refine the segmentation at the boundary and produce more accurate cell membrane edge. Finally, because the picture of complex data set is so large, we can adopt the method of inflation prediction, namely overlapping sliding window sampling, and average the overlapping part of the predicted results to make the obtained boundary more continuous.

In conclusion, we have achieved good results on the neural network for biological cell edge detection, and achieved high accuracy and recall rate on the validation and test set. It has a high practical significance for future neuroscience and medical vision processing.

References

- Canny, J. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(6):679–698, 1986. doi: 10.1109/TPAMI.1986.4767851.
- Chaurasia, A. and Culurciello, E. Linknet: Exploiting encoder representations for efficient semantic segmentation. *CoRR*, abs/1707.03718, 2017.
- He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.
- Hu, Y., Chen, Y., Li, X., and Feng, J. Dynamic feature fusion for semantic edge detection. *CoRR*, abs/1902.09104, 2019.
- Liu, Y., Cheng, M.-M., Fan, D.-P., Zhang, L., Bian, J., and Tao, D. Semantic Edge Detection with Diverse Deep Supervision. *arXiv e-prints*, art. arXiv:1804.02864, April 2018.
- Ronneberger, O., Fischer, P., and Brox, T. U-net: Convolutional networks for biomedical image segmentation. *CoRR*, abs/1505.04597, 2015.
- Yu, Z., Feng, C., Liu, M., and Ramalingam, S. Casenet: Deep category-aware semantic edge detection. *CoRR*, abs/1705.09759, 2017.

A. Hyperparameters

In order to facilitate the reproduction of experimental results, the hyperparameters used in our experiments are listed in Table 3:

Hyperparameters	Simple	Complex
learning rate	backbone: 0.001, others: 0.01	
optimizer	Adam	
batch size	8	4
crop size	960	1280
training epoch	500	100
γ of Focal Loss	2	
α of Focal Loss	0.7	
λ	0.8	
pre-binariization thres	122	
post-binariization thres	127.5	

Table 3. Hyperparameters

B. Ablation Study

In this section, we conduct a series of ablation experiments to view the effects of label denoising, different data augmentations and different network architectures .

B.1. Label Denoising

As shown in Fig. 3, label denoising can accelerate the model convergence speed and improve the final performance of the model. The comparison of the results before and after denoising in Simple-Track is shown in Table 4.

B.2. Data Augmentation

Due to the extreme lack of training data (especially in Simple-Track), strong data augmentation is crucial to prevent overfitting. Here we try three different data augmenta-

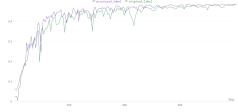
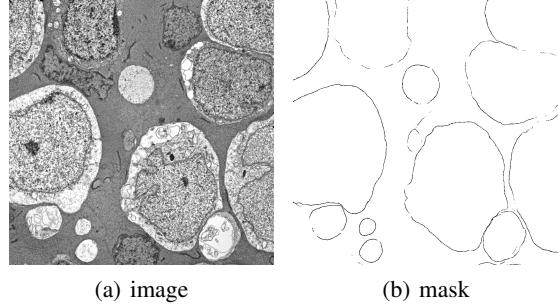


Figure 3. Training with labels before and after denoising

Method	F-score
before denoising	0.4856
after denoising	0.4869

Table 4. Comparison of the results before/after denoising

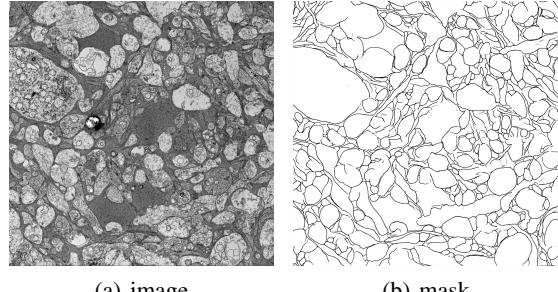


(a) image (b) mask

Figure 4. Simple-Track prediction

C.1. Simple-Track

C.2. Complex-Track



(a) image (b) mask

Figure 5. Complex-Track prediction

B.3. Network Architecture

We have also tried a variety of advanced network architectures for image segmentation and edge detection, including DDS, CASENet and DFF. Their performance differences are listed in Table 6, all experiments are conducted in Simple-Track and tested on the validation dataset.

Architecture	Backbone	F-score
DDS	ResNet-50	0.49058
DDS	ResNet-152	0.48914
CASENet	ResNet-50	0.49196
CASENet	ResNet-152	0.48699
DFF	ResNet-50	0.51182

Table 6. Comparison of different network architectures

C. Some Results

Here, we show some of the predicted results of our method on the test dataset in both Simple-Track and Complex-Track.