Automated annotation of cellular cryo-electron tomograms using U-Net

Yanyan Zhao (yanyanz) , Weijiang Zhou (weijianz)
Department of Bioengineering, School of Medicine, Stanford University

INTRODUCTION

Electron cryo-tomography(cryo-ET) is a popular tomography technique used by biologists in recent years to determine nanometer resolution 3-D sub-cellular structures. This technique is normally featured with imaging a rotating sample for 120 degrees, generating a series of 2D images that can be combined to produce a 3D reconstruction. The main method researchers use in reconstructing the 3D volume of the sample starts with manually annotating various features such as cytoskeletal filaments, cell wall elements and internal compartments in each 2D tilt image. Although manual annotation is still most accurate so far, the concern of its low efficiency becomes more serious especially when data collection is speeding up these days. Very few studies have been reported so far to address this issue. Here we propose to apply deep learning to cryo-ET tomogram annotation to provide users a good trained neural network for general cryo-ET cellular annotation and therefore help them accelerate cell biology study. Our goal is to generate an annotated tomogram by taking each slice of the 3D tomogram and passing it through our pre-trained neutral network. In this project, we are aiming at annotating six features, including background, microtubules, ribosome, double layer membrane, single layer membrane and carbon edge.

RESULTS AND DISCUSSION

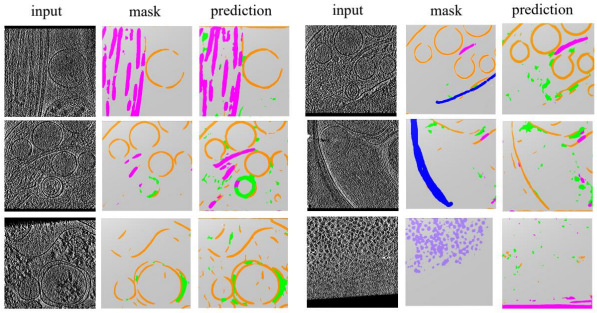

Figure 2. Neural network prediction.

Our model shows relatively good accuracy in predicting microtubule, double membrane and single membrane. Especially, it accurately predicts microtubules and double membranes in areas where these features are mislabeled as false negatives. However, the model performs poorly in terms of prediction on ribosome and carbon edge. Considering the relatively rare occurrence of these two features in the dataset, up-sampling or application of higher weights will be tried to solve the problem. In general, the metric F1 score for cell feature classes is relatively low compared to noise class. Given that noise accounts for more than 90% of the pixels in the dataset, this class imbalance poses a potentially challenging problem for better neural network training. In addition, for each feature class, especially for ribosome and carbon edge, the high variance of F1 score across batches in each epoch training suggests the bias in data loading during training and testing.

noise or other
ribosome
single membrane
microtubule
double membrane
carbon edge

DATASET

Raw dataset:

- One manually annotated PC-12 cell tomogram in size of (94, 864, 868).
- 38 neuron cell tomograms in size of (n, 960, 960) where n is between 113 and 630 with median 281.
- Four ribosome tomograms from EMPIAR 10064 in size of (256, 1024,1024).

Data preprocessing:

- Top and bottom slices in neuron and ribosome tomograms with only noise or mostly noise are excluded. 38 neuron cell tomograms are combined into 8 large tomograms for more accurate annotation.
- All neuron cell and ribosome tomograms are semi-automatically annotated using e2tomoseg_convnet.py developed by Muyuan Chen[1]. The annotated features includes microtubule, ribosome, double layer membrane, single layer membrane and carbon edge. False positives from semi-automatic annotation are largely excluded by thresholding and manual cleaning. Mask for noise and five features are further encoded with 0 <= number < 6(class number).
- All cleaned 3D tomograms and their corresponding 3D masks are extracted into 2D images with each image being cropped into four images of size 512 by 512.
- The final dataset contains 16,856 512x512 images. Dataset is randomly shuffled and divided into train set, dev set and test set(80/10/10).

DEEP LEARNING APPROACH

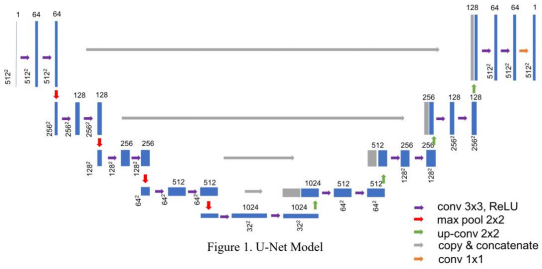

Figure 1. U-Net Model

Evaluation metric: F1 score for each feature.

Weighted cross-entropy loss:

$Loss(x, class) = \text{weight}[\text{class}] * (-x[class] + log(\sum_j exp(x[j])))$

Hyperparameters explored include batch size, learning rate, class weights and number of epochs.

F1 SCORE	NOISE	MICROTUBULE	RIBOSOME	DOUBLE MEMBRANE	SINGLE MEMBRANE	CARBON EDGE
TRAINING	0.94	0.57 (0.22)	0.47 (0.22)	0.48 (0.21)	0.71 (0.50)	0.60 (0.40)
TEST	0.98	0.77 (0.36)	0.01	0.66 (0.40)	0.79 (0.52)	0.40 (0.10)

(50 epochs, Adam optimizer, learning rate = 0.001)

CONCLUSION

The preliminary results of our U-net model shows that deep learning offers an relative accurate annotation for several cellular features including microtubule, double membrane and single membrane. Compared to shallow neural network reported previously which targets only one feature, the U-net model performs better in terms of low false positives. However, because of the class imbalance especially for ribosome and carbon edge, our model does not show perfect prediction on these features so far. Up-sampling of minority classes and introducing more data of these features is the next step we're aiming to try.

FUTURE WORK

For our current 2D U-Net model,

- Manually labeling the dataset to exclude false positives and false negatives.
- Clean the dataset and add more data with ribosome and carbon edge to balance classes.
- Balance all classes in training set, development set and test set.

In the future, we plan to try 3D U-Net on tomogram annotation considering some features e.g. microtubules has little information on 2D slices when they go in Z direction.

More GPUs to try larger batch size and also for faster training.

REFERENCE

1. Chen, M. *et al*. Convolutional neural networks for automated annotation of cellular cryo- electron tomograms. Nat. Methods 14, 983–985 (2017).

2. Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. in Medical Image Computing and Computer-Assisted Intervention -- MICCAI 2015.