Quasi-Replica of U-Net for Biomedical Image Segmentation

Minxin Zhang

University of Sydney, Sydney, Australia

November 7, 2023

U-Net model relying on data augmentation with a small training sample is effective to perform nuclei segmentation with a cost of 400ms for each 256×256 pixel U2OS image using an NVIDIA T4 graphic card.

1 Introduction

This project is to evaluate the performance of a U-Net model, which is a variant convolutional network trained model, for BBBC039 dataset – a subset of BBBC022 chemical screen on U2OS cells. By tuning parameters of the training model, we could compare the performances under different costs(mini batched stochastic gradient descent sample feeders, optimizer convergence behaviours and data augmentation) so that we can get a model with satisfying prediction accuracy to segment further U2OS chemical screen with less computational resource overhead. The implementation framework is the same as that in the U-Net paper given (**1**). By enabling data augmentation, the U-Net model is tested whether it is effective even only fed 2 raw images, which is consistent with the claim of the U-Net paper - "it only needs very few annotated images and has a very reasonable training time" (**1**)

2 Method

Here is the recipe how I trained the U-Net model and evaluate its performance. Firstly, raw data was processed so that it could be a valid input fed to the convolution network and also valid to be fed to the trained model to generate valid output which are to be validated and tested. The encoder part of the network architecture is the same as a CNN while the decoder part gradually increasing spatial resolution. In this case, images of $256 \times 256 \Rightarrow 128 \times 128 \Rightarrow 64 \times 64$.

2.1 Raw Image preprocessing

1. Normalize the raw images(cropped into 256×256)
2. Label the raw (ground truth) annotations
3. Filter noises in the annotations(remove tiny nuclei)
4. Find the labeled boundaries and apply dilation (by a 2×2 kernel) to make the boundaries a bit larger
5. Convert the processed raw annotation images to greyscale images(binary)
6. Apply elastic deformation both to images and annotations.(1 raw entry is deformed into 10 augmented entries

Split a subset of BBBC039 into 3 partitions which are for training (40%), validating(30%) and testing (30%).

2.2 Training

1. use a Root Mean Squared Propagation optimizer as the minibatch SGD learning method with an heuristic objective of a weighted 3-class categorical cross entropy, labels of which are 'background', 'foreground(nuclei interior)' and 'boundary' to segment the background and the foreground.

2. feed mini batches of the training set(each entry is cropped to a 256×256 pixels image to the optimizer so that the model parameters are updated

2.3 Predict and Evaluate

1. use the trained model to covert the validation raw images to a label mapping(predicted model to label matrix for each pixel) and then we expand the pixels of 'boundary' class a little thicker for a better visibility at the cost of shrinking 'foreground' convexly.
2. compare the predicted images with the corresponding ground truth annotations by calculating the rate of the intersection of the prediction data and the test data over the union of them('background' area excluded) to generate a Jaccard index. In addition, F1 score is calculated to measure the balance of the classification.

3 Experiment Setup

3.1 Dataset

The original data consists of 200 raw images, which are 200 fields of view of nuclei captured with fluorescence microscopy using the Hoechst stain, (**2**) and 200 raw corresponding labeled annotations. The raw images are stored as TIFF files with 520x696 pixels at 16 bits. (**2**). Label matrices require decoding the raw annotations since the ground truth annotations are stored as PNG files encoding masks of independent nuclei: if two nuclei touch, they are labeled with a different color. (**2**)

3.2 Training params

Please check **Table 1** for the setting.

Table 1: *Training setting*

min nucleus size	25 px
nucleus boundary width	2 px
augmentation ratio	10
learning rate	0.02
# of epochs	5
# of steps per epoch	200
heuristic fn	softmax loss
SGD batch size	8 images
dilation kernel size	3×3 pxs

3.3 Evaluation metrics

- Jaccard index(intersection over union ratio) to measure the model prediction accuracy, which is correct predicted area over the whole image.
- the mean of F1 scores for Jaccard values greater than such partition $\{0.5, 0.55, ..., 0.95\}$. This is to measure the balance between precision and callback in classification procedure.

4 Results & Discussion

Let us firstly check how the model fine tuned(0.0002 learning rate, 15 epochs, 500 steps per epoch and batches of 10 per step) performs. **Figure 1** is the ground truth image of IXMtest_A02_s1_w1051DAA7C-7042-435F-99F0-1E847D9B42CB. Now we feed validating partition of raw images to the trained model (saved at `/data/experiments/02/model.hdf5`). **Figure 2** is the corresponding predicted image for **Figure 1**. We can see that the prediction is decent enough only missing 7 nuclei, which are highlighted in **Figure 3**. **Figure 4** shows the intersection over union of **Figure 1** and **Figure 2**.

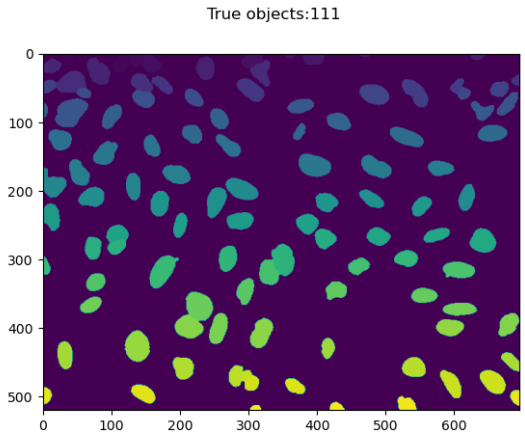

Figure 1: *Test image*

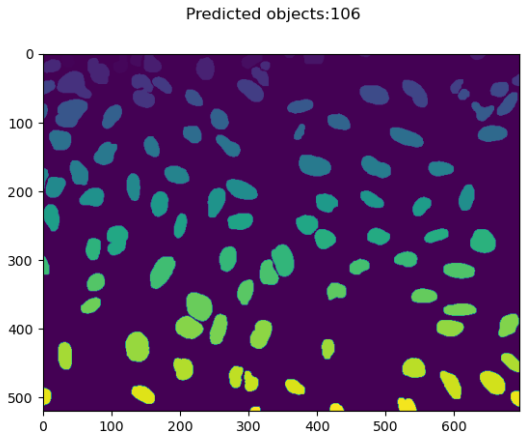

Figure 2: *Predicted image*

Now you have a general sense of how the evaluation for 1 raw image is performed. Because data augmentation is adopted, the actual training size is enlarged. In

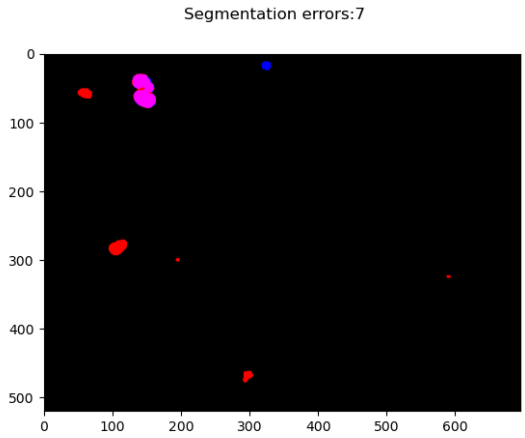

Figure 3: *Difference between the predicted and the test*

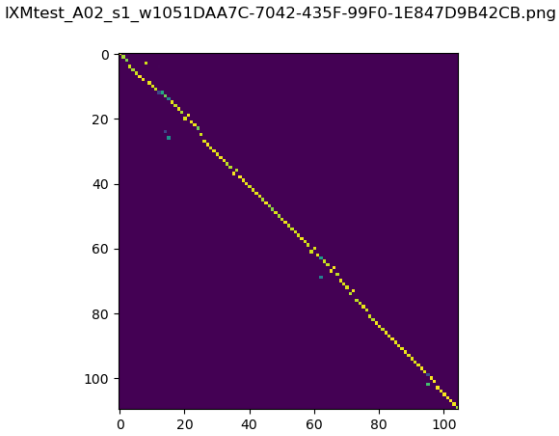

Figure 4: *Intersection over union*

this instance, 1 raw image is mapped to 10 elastically deformed images(normalized). Therefore, we have 1000 images to use as our dataset after the original 100 raw images are preprocessed. So we need to measure the statistics information. Feeding different number of raw images, I got **Table 2**, which is part of the experiment output. You can check the original csv file at `/data/experiments/impact_of_augmented_dataset_size.csv` We can see that the similarity between predicted images and test images(Jaccard index) is stable except for 6 raw images entry. This is due to an iteration of simulation had a bad random number generator. The 2nd training simulation for 6 raw images(Experiment # 11) had 4000 missed nuclei segmentation.

5 Conclusion

In summary, I have verified that the U-Net model trained by 2 raw image performs almost the same as those trained by much more images. However, there's still some limitations on this study. The data augmentation method is not tested and I kept the default amplifying factor(10). If more augmented data is required given the same size of raw data, the deforming algorithm needs to be calibrated with respect to the amplifying constant. Besides, the claim of "seamless segmentation" (**1**) for overlap-tile strategy is not verified as the neuronal structures in EM stacks within the CNN model builder are not tested.

Table 2: *Training setting*

# of raw images	F1 mean	Jaccard mean
2	0.73	0.85
4	0.77	0.87
6	0.62	0.74
8	0.77	0.87
10	0.76	0.87
20	0.76	0.86
40	0.73	0.85
60	0.77	0.87
80	0.75	0.86
Benchmark 100	0.83	0.89

References

[1] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, *U-Net: Convolutional Networks for Biomedical Image Segmentation*, Computer Science Department and BIOSS Centre for Biological Signalling Studies, University of Freiburg, Germany.

[2] Broad Bioimage Benchmark Collection, *Nuclei of U2OS cells in a chemical screen*, Broad Institute.