Deep Learning for Partial Annotated Lung CT Analysis

Jiahao Li[1], Xingzheng Lyu[2], Li Cheng[1]

[1]Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Canada, [2]College of Computer Science and Technology, Zhejiang University, Hangzhou, China

Introduction

Multi structures segmentation of area of interests is one of the main tasks in biomedical image analysis. Previously, the area of interests in biomedical image are annotated manually by experts in medical area. But the booming of deep learning technology has helped to achieve a great performance on biomedical image segmentation in various kinds of medical images from computed tomography to MRI for one or more structures.

Especially, the deep learning network, U-Net, has greatly improved the quality of multi-structures biomedical image segmentation for the areas of interests with using annotated mask data. However, there will not be annotation for every segment in a medical image. It is important to train a network which can combine fully annotated image with partially annotated image to yield a precise prediction in those partially annotated medical image. In this project, a modified version of multi-class segmentation U-Net framework, CU-Net[1], is extended to multi-label segmentation. And the original U-Net framework Is served as a reference in experiment. The developed partial supervised learning model will be used on image segmentation in partially annotated biomedical image.

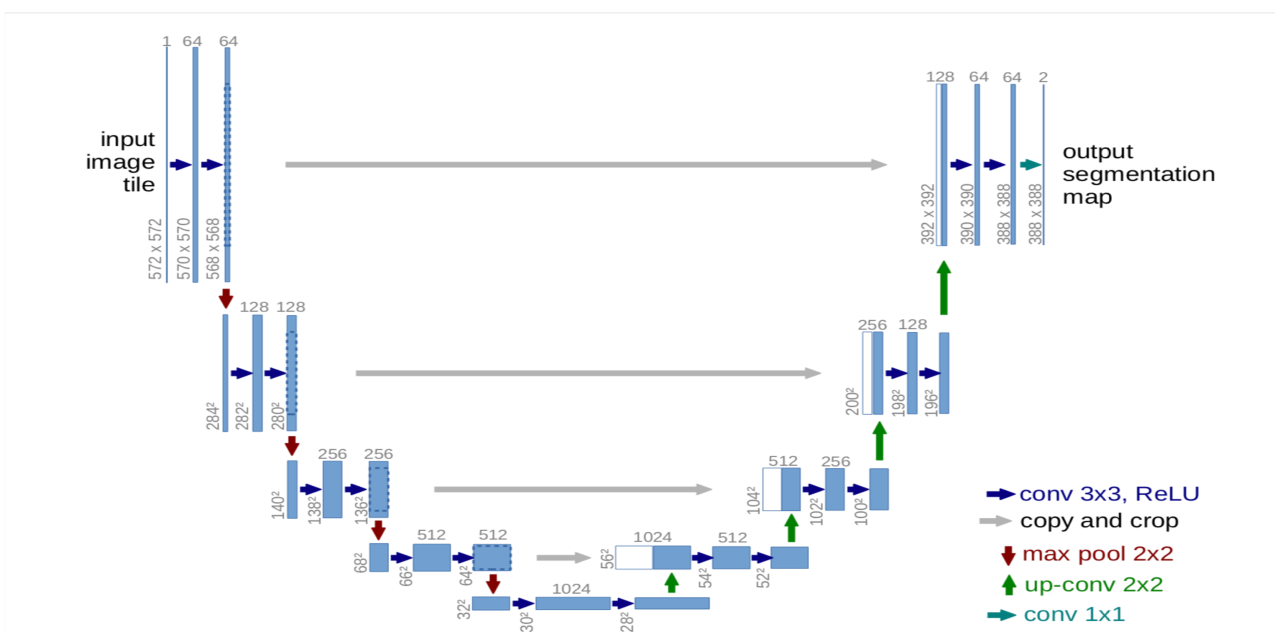

Figure 1: Architecture of U-Net

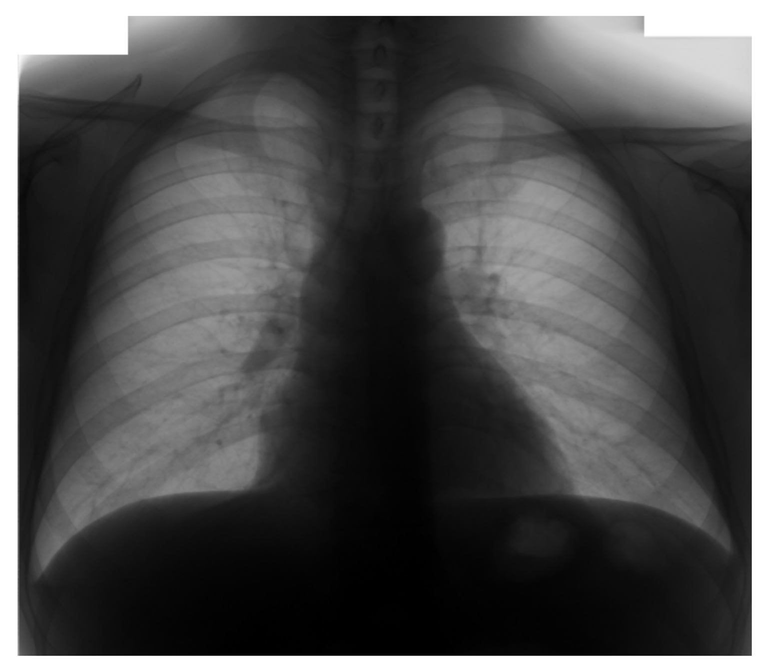

Figure 2: Image of lung CT in JSRT

Algorithm

The multi-structure segmentation network is based on the convolutional neural network, CUNet.[1] In order to adapt the network into multi-label image segmentation, the last layer of the network is in one-hot coding schema, with sigmoid activation function for one-vs-rest logistic regression.

The cost function of this multi-label network is defined as

$$f(y,\hat{y}) = \frac{\sum_{i=1}^{n} \delta(id_i = i) dice(y,\hat{y})}{\sum_{i=1}^{n} \delta(id_i = i)},$$

where δ is an identifier which is equal to 1 when mask of an organ exist in training data. And the loss function is optimized by *Adam* with step size $5 * 10^{-5}$.

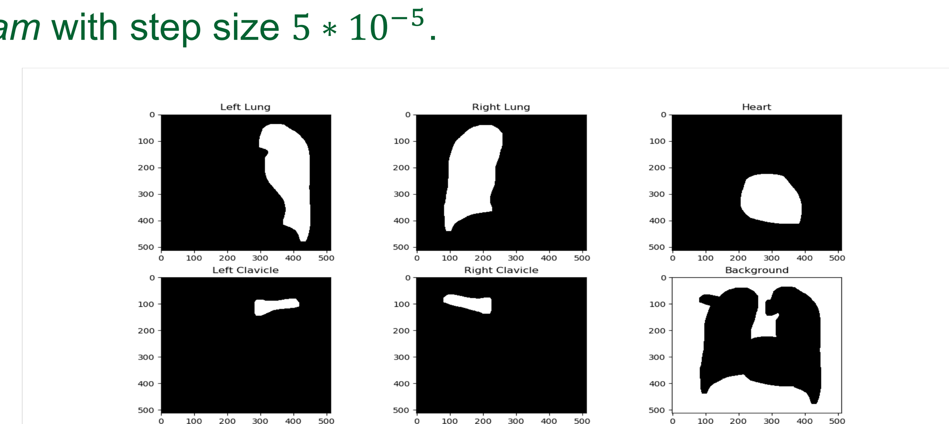

Figure 3: One-hot coding Schema

Why U-Net?

U-Net is a deep learning framework can provide high quality segmentation result within a small dataset with performing image deformation. The localization of segmentation is improved by combining high resolution features with output from up-sampling. And the U-Net can also achieved high accuracy in the context in segmentation by replacing pooling operator by up-sampling operator.[2]

The modification in CUNet is there is a convolution layer in the skip connections of the network. Skip connection can effectively prevent the gradient vanishing in Deep Neural Network. And the convolution in the skip connections can contribute to the up-sampling and learn from the up-sampling directly.

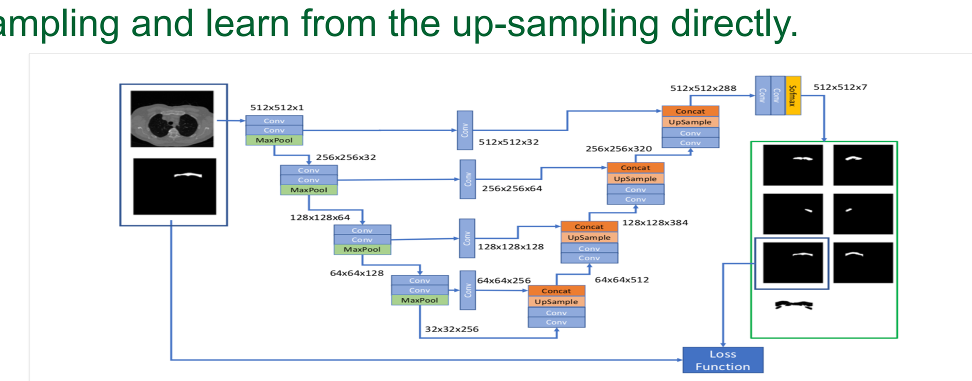

Figure 4: Architecture of CUNet

Experiment

Data: The experiment is performed on Japanese Society of Radiological, (JSRT), lung CT dataset, for which there are masks annotated manually for different organs in the lung CT image.

Image Preprocessing: Raw images in JSRT are converted to .png format and cropped into size of 512x512x1. Since JSRT has only limited number of images, data augmentation is performed in training data, which can greatly increase the number of available training data.

Training: There are 247 images in JSRT dataset. Before each run of training, 27 images are randomly selected as the validation data, and the others become training data. Training is performed in 30 epochs and both training loss and validation metrics, dice score and accuracy, are monitored.

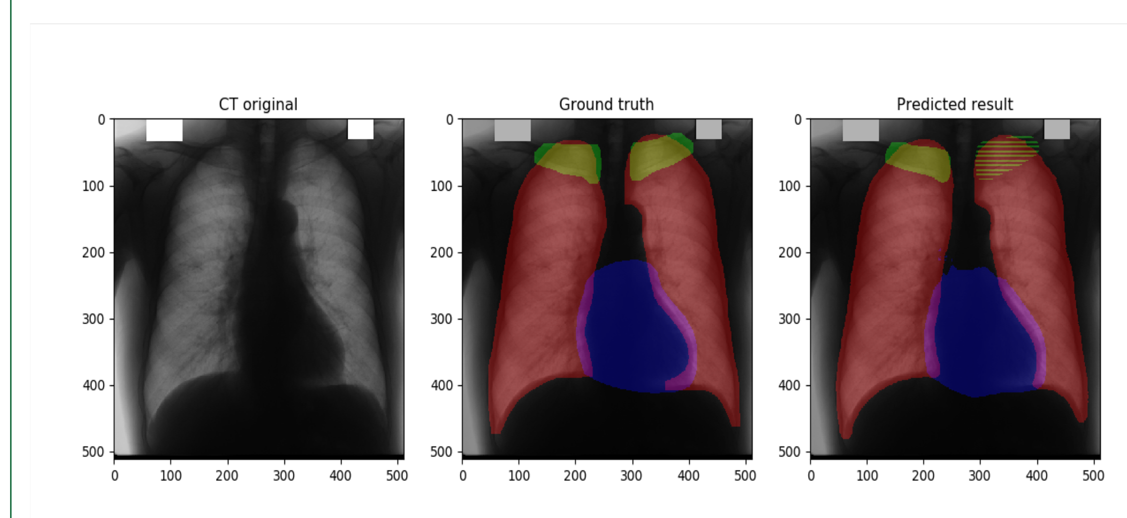

Figure 5: CUNet prediction results and ground truth mask in lung CT

Statistical Analysis: The performance of CUNet is compared with original U-Net architecture. Both CUNet and U-Net are trained under 30 epochs following the same training trajectory. And the experiment results are compared in binomial test and student t-test.

Table 1: Average Validation Error in Dice Score and Accuracy with Standard Errors

	Dice Score	Accuracy
CUNet	0.883±0.0145	0.895±0.021
U-Net	0.885± 0.0149	0.895±0.0188

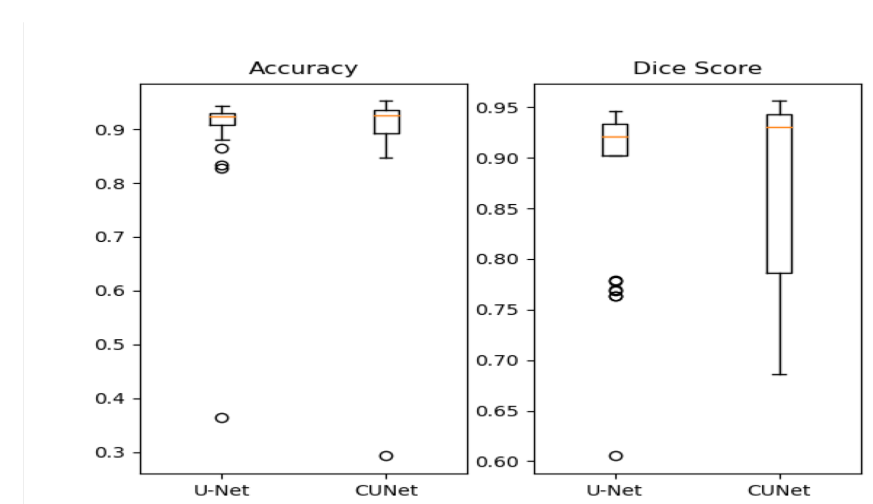

Figure 6: Boxplot for Validation Error in U-Net and CUNet

Discussion

At the beginning of discussion, a Null hypothesis is made that two deep learning frameworks have same performance obtained from training. Binomial test is an one-tailed test and Student t-test is a two-tailed test. And p-value used for rejecting the null hypothesis in the experiment is set as 0.1.

Table 2: p-value

	Dice Score	Accuracy
Binomial test	1	1
Student t-test	0.929	0.975

The average validation metrics are similar between CUNet and U-Net. Therefore, according to the validation dice score and accuracy of experiment, CUNet and U-Net can achieve same prediction performance in partial supervised learning in JSRT dataset.

CUNet can achieve excellent performance when extended to multi-label image segmentation. The intersections between ground truth and predictions are in pink color in Figure 7. And the gray color is ground truth of the annotated mask. CUNet do not achieve high accuracy in left clavicle prediction. Since the symmetric of clavicles and data augmentation performed in preprocessing, the overall prediction is still excellent.

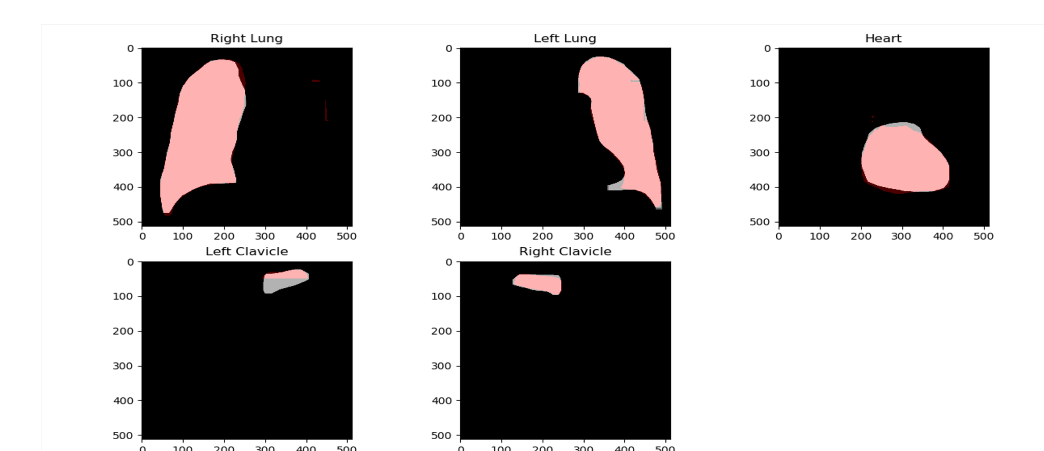

Figure 7: Difference between ground truth and prediction in CUNet

Reference

1. González, Germán, et al. "Multi-Structure Segmentation from Partially Labeled Datasets. Application to Body Composition Measurements on CT Scans." *Image Analysis for Moving Organ, Breast, and Thoracic Images Lecture Notes in Computer Science*, 2018, pp. 215–224., doi:10.1007/978-3-030-00946-5_22.

2. Ronneberger, et al. "U-Net: Convolutional Networks for Biomedical Image Segmentation." *ArXiv.org*, 18 May 2015, arxiv.org/abs/1505.04597.