Pancreas Segmentation Using Attention U-Net

1st Zengtian Deng
Bioengineering
UCLA
ztdeng@ucla.edu

***Index Terms*—Pancreas Segmentation, U-Net, Attention U-Net**

INTRODUCTION

Pancreatic cancer is one of the common malignancies with increasing incidence worldwide every year. As discussed by Yao et al, it is estimated that there are 43090 pancreatic cancer deaths within US in 2017, which is the fourth highest mortality rate of malignant tumors [1]. In diagnosis, therapy and surgery of pancreatic cancer, the correct segmentation of pancreas is a crucial preliminary process. In general, the segmentation process is done manually by the radiologists, which is time consuming and prone to human error. Therefore, automatic algorithms to correctly segment the pancreas are strongly desired.

In general, the approaches of pancreas segmentation can be split into two categories. The first category is the non-deep learning method, and one common approach is atlas based segmentation. This method first requires a large database of images with pancreas labeled; The algorithm then tries to find the best matched atlas in the database with the sample; Lastly, the algorithm transforms the altas segmentation back to the sample to be the sample segmentation. For exmaple, Karasawa et al. uses vessel structure as the key structure to perform Atlas based segmentation. This method has shown good result comparing to other machine learning methods. However, this method is highly dependent on the atlas database used by the algorithm, which is a manual and time consuming task as well. In contrary, the deep learning method requires less human effort and is able to produce better performance comparing to the atlas based method.

In pancreas segmentation, Oktay et al. have gained success in implementing Attention U-Net for pancreatic on two different CT abdominal datasets [2]. Oktay et al. also showed that Attention U-Net outperforms the generic U-Net model as well [3]. The dataset I found is another CT pancreatic dataset from Medical Segmentation Decathlon, which is a large collection of segmentation datasets on different tasks [4]. Based on literature research, none of the studies has conducted Attention UNet on this dataset. Therefore, the goal of this project is to explore the generalizability of implementing Attention UNet for pancreas segmentation in various domain and whether Attention UNet outperforms generic U-Net in different settings.

PROBLEM DEFINITION

The segmentation problem can be defined as a classification problem. The major task is to correctly classify each pixel into either pancreas or background. Each pixel within the volume represents each data point. However, the features of each data point are implicit characteristics like its own intensity value, its relationship between surrounding pixel intensities, its location within the volume, etc. This type of classification task is very common in the field of medical image analysis because it is crucial in localization of the subject of interest.

METHODS

A. Data Set and Prepossessing

The dataset used for this project contains 284 abdominal phase CT scans with manually labelled pancreas and pancreatic tumor. Figure 1 shows some example slices with the pancreas segmentation map. For most cases, the pancreas only takes a very small portion of the whole CT volume. Figure 2 shows a histogram of background foreground ratio of all volumes for training and validation. As a result, train the model directly on the whole volume would cause a strong class imbalance problem.

Fig. 1. Sample abdominal CT scan with pancreas label(red) and tumor label(green). The transverse view(A), sagittal view(B), coronal view(D), and 3D label volume(C) are shown.

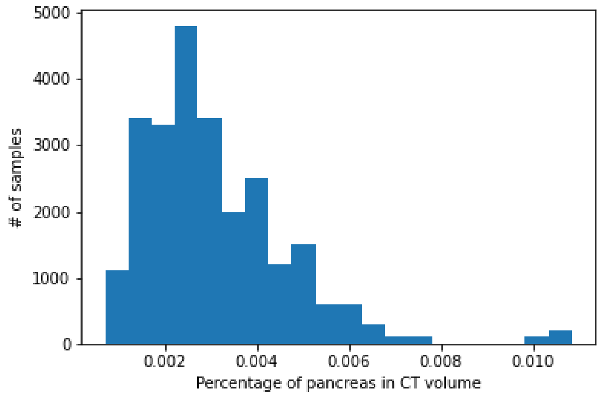

Fig. 2. Histogram of ratios between pancreas volume and whole scan volume.

To address the problem stated above and maximize the performance of the model, several preprocessing steps were taken:

1) The label value is thresholded to blend tumor label into pancreas label.
2) Grey-level mapping is also conducted to constrain the intensity range to be within -100 and 164, which is based on both visual evaluation and online reference [5].
3) The zero-intensity voxels are eliminated by cropping with the smallest cuboid containing all nonzero-intensity voxels.
4) Each image is normalized to have zero mean and 1 standard deviation.
5) Several fixed size volumes are randomly sampled from each sample with constraint on their centers: The centers of all samples will be at either label or background with a fixed ratio. For this project, 4 sub-samples are generated from each sample with the ratio set to 50%, which means fifty percent of the sampled volume will be centered at a pancreatic voxel and the rest centered at non-pancreatic voxel.
6) Finally, all sampled volume undergoes a random affine transform for data augmentation.

B. U-Net

UNet is one of the most common neural networks used for image segmentation due to its simplicity and good performance [3]. Unlike two stage models like Mask R-CNN, UNet uses a set of convolutions followed by a series of up-convolution to extract the feature from the images and then reconstruct the segmentation map based on the features. Besides, the feature map of each feature extraction layer is also fed to the up convolution layers to assist the reconstruction process.

C. Attention U-Net

As shown in Figure 3, Attention U-Net has the same architecture as U-Net. On top of U-Net, Oktay et al modified the skip-connection part of generic U-Net to improve detection performance as shown in Figure 4 [2]. Within Figure 4,

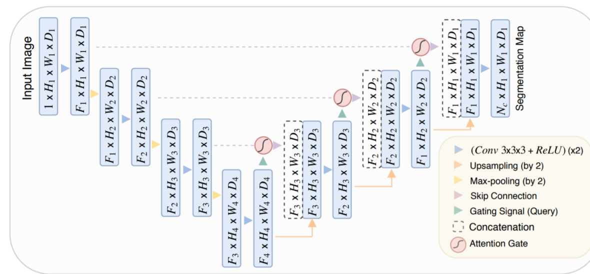

Fig. 3. Block diagram of Attention U-Net shown in the original paper of Oktay et al. [2]

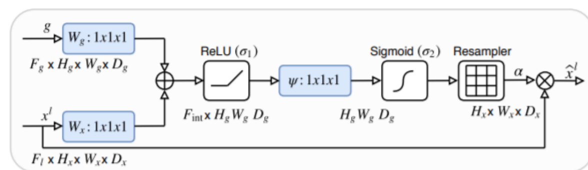

Fig. 4. Flow chart of attention gate shown in the original paper of Oktay et al. [2]

x^l represents the map from feature extraction path before skip connection at layer l, and g represents the deeper level reconstructed feature map before up-convolution. The whole process can be express as follows:

$$\hat{x_i}^l = \alpha_i x_i^l \tag{1}$$

$$\alpha_i = \sigma_2(\psi^T(\sigma_1(W_x^T x_i^l + W_g^T g_i + b_g)) + b_\psi), \tag{2}$$

where σ_1 is the ReLU function and σ_2 is the sigmoid function to introduce non linearity. The W_x^T and W_g^T is the linear transformation over x^l and g by 1x1x1 convolution, with W_g^T having appropriate stride operations to down-sample x^l to the dimension of g. The two feature maps after linear transformations are then added together, also called as additive attention. The combined feature maps are then passed through the ReLU function and then passed through another linear transformation ψ along with the sigmoid function to generate a scale map with range 0 to 1 to be multiplied with the original feature map x_i^l. The resultant $\hat{x}^l$ is then passed through the skip connection to the rest of the model.

The first motivation of the attention gate is to allow the model to focus on certain region of the extracted feature during the reconstruction process instead of treating all regions the same. This process allows the model to ignore features that are not related with pancreas.

The second motivation of the attention gate is to allow layers at different level to share information by utilizing both coarse and fine feature information to make decisions on where to look at during the skip connection process.

D. Implementation

The model was implemented in PyTorch, an open source machine learning library, along with MONAI, which is an open source PyTorch based framework for medical related deep learning studies. To implement the attention gate method, the original U-Net model provided by MONAI was modified and an attention gate layer was constructed to implement the Attention U-Net. The preprocessing steps mentioned before were implemented using MONAI transform method.

Both models are first trained with 80% train 10% validation 10% test split. To explore the performance of both networks based on model complexity, the models were trained using two set of channel sizes: (64, 128, 256, 512,1024) and (16, 32, 64, 128, 256). Due to the limitation of time, cross validation analysis was not finished by the time submitting this report, but the result will be updated once the training finishes.

RESULTS

The performance of the model is evaluated based on the following metrics:

1) Sørensen–Dice coefficient, which is the overlapped region between predicted and ground truth regions over the union of the two regions.
2) Sensitivity and Precision of the result.
3) Hausdorff Distance, which is maximum distance to travel from the boundary of predicted volume to that of the ground truth volume
4) Average Surface Distance, which is the average distance of all boundary points of predicted volume to that of the ground truth volume.

The specificity is excluded in evaluation due to the large class in-balance of the dataset. Since the background volume is too large, the specificity of the model will be inevitably high. The results of all models are shown in the table below.

TABLE I
MODEL PERFORMANCE ON TEST SET

Model Type	DSC(%)	SP(%)	SE(%)	d_H	ASD
$UNet_{small}$	72.55%	72.79%	78.29%	24.28	1.936
$UNet_{large}$	74.95%	77.65%	78.19%	22.44	1.560
$AttnUNet_{small}$	71.10%	74.51%	74.67%	39.20	2.826
$AttnUNet_{large}$	74.56%	75.69%	78.78%	27.48	2.904

[a]Sample of a Table footnote.

As shown in Table 1, it is interesting that the Attention U-Net model is performing worse with test set comparing to U-Net model, which contradicts the validation performance as well as the literature. One possible reason of this unexpected behavior is bugs in model architecture of Attention U-Net model, which causes the model to correctly identify pancreas regions with attention gate but not utilizing the information.

Figure 5 is the inference of the Attention U-Net model on test set with the attention coefficient output from the lower level feature maps. As shown on the figure, the model is able to "give" attention to regions that are highly likely to have pancreas. However, according to the low dice coefficient, the model is obviously not learning from this advantage.

Figure 6 is the training loss and the validation dice score of both U-Net and Attention U-Net model. The validation

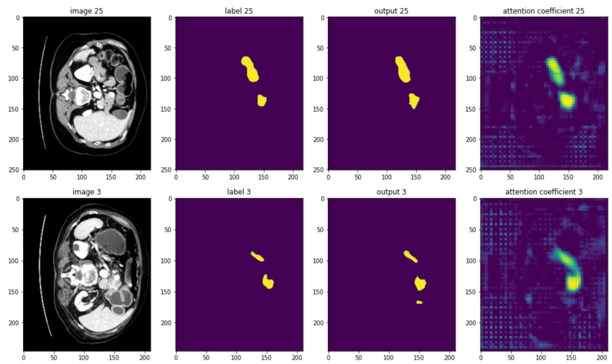

Fig. 5. Two example inference result of Attention U-Net model. From left to right, the columns are original input data, ground truth label, model prediction and Attention coefficients.

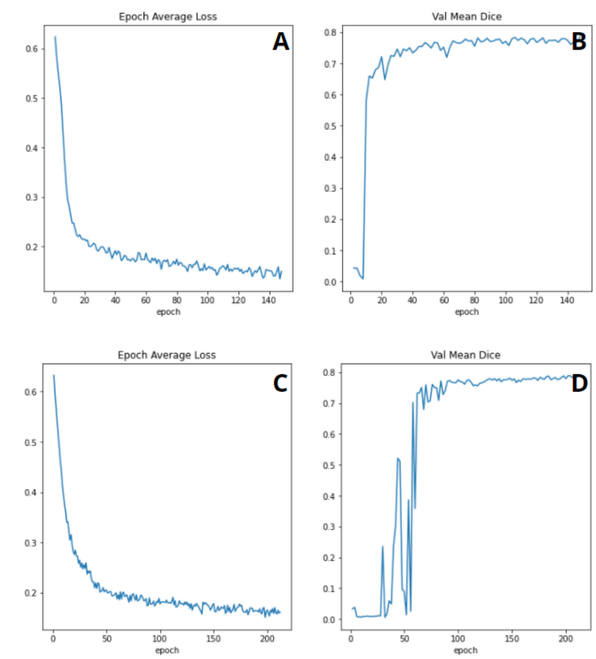

Fig. 6. Training loss(A) and Validation Dice Score(B) of U-Net and Training loss(C) and Validation Dice Score(D) of Attention U-Net. Both with large channel number.

accuracy rises much slower and more unstable comparing to the U-Net model.

In conclusion, the future directly of this project is to first definitely debug the architecture and retrain the Attention U-Net model. In addition to that, in case there are no errors within the model, then it implies that attention mechanism is not able to generalize the improvement on original model's performance in any subjects.

REFERENCES

[1] Xu Yao, Yuqing Song, and Zhe Liu, "Advances on pancreas segmentation: a review," *Multimedia Tools and Applications*, vol. 79, 03 2020.

[2] Ozan Oktay, Jo Schlemper, Loic Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Hammerla, Bernhard Kainz, Ben Glocker, and Daniel Rueckert, "Attention u-net: Learning where to look for the pancreas," 04 2018.

[3] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," 10 2015, vol. 9351, pp. 234–241.

[4] Amber L. Simpson, Michela Antonelli, Spyridon Bakas, Michel Bilello, Keyvan Farahani, Bram van Ginneken, Annette Kopp-Schneider, Bennett A. Landman, Geert Litjens, Bjoern Menze, Olaf Ronneberger, Ronald M. Summers, Patrick Bilic, Patrick F. Christ, Richard K. G. Do, Marc Gollub, Jennifer Golia-Pernicka, Stephan H. Heckers, William R. Jarnagin, Maureen K. McHugo, Sandy Napel, Eugene Vorontsov, Lena Maier-Hein, and M. Jorge Cardoso, "A large annotated medical image dataset for the development and evaluation of segmentation algorithms," 2019.

[5] Masis Isikbay, "Fundamentals of computed tomography studies: Windowing," .