Mentors: Dr. Francis Creighton, Dr. Mathias Unberath, Dr. Russell Taylor
Team: Ameen Amanian, Chanha Kim, Yuliang Xiao

Automated Segmentation of the Eustachian Tube – A Deep Learning Platform

Background Reading Critical Review

Project Summary:
To develop a deep learning platform for automated segmentation of the eustachian tube in computer tomography images to improve our understanding of the eustachian tube anatomy, and enhance the diagnostics and treatment of patients with eustachian tube dysfunction.

Briefly, the ground truth will include segmentation of the eustachian tube with near-by structures such as the internal carotid artery, torus tubarius, and Fossa of Rosenmuler. The images will then be co-registered to exhibit the same orientation and will then serve as input into nnUNet, a state-of-the-art deep learning architecture for medical image segmentation. Following training of the model, inference will provide predictions which will undergo validation against the ground using metrics such as the Dice Similarity Coefficient (DSC) and Average Hausdorff Distance (AHD) (Figure 1).

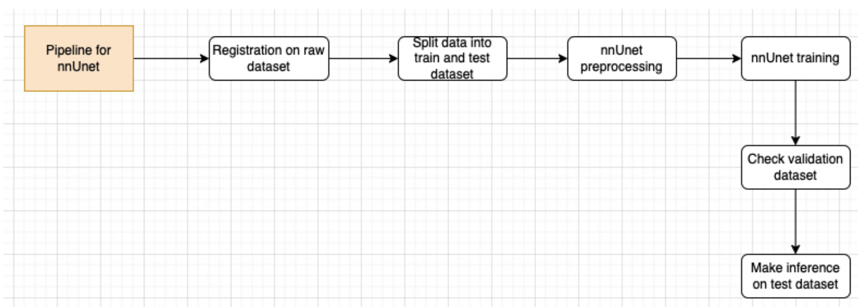

Figure 1 - Project Workflow

Paper Selection:
The two papers described below have been selected as they have utilized nnUNet for medical image segmentation for areas of pituitary adenoma and pancreatic ductal adenocarcinoma. The papers provide results that exhibit their model architecture, hyperparameters, and validation metrics. Key lessons have been deciphered from this review that will be translated into the implementation of our project.

Paper #1: Three-Dimensional Semantic Segmentation of Pituitary Adenomas (PA) Based on the Deep Learning Framework-nnU-Net: A Clinical Perspective[1]
Xujun Shu, Yijie Zhou, Fangye Li, Tao Zhou, Xianghui Meng, Fuyu Wang, Zhizhong Zang, Jian Pu, and Beinan Xu
Journal: Micromachines; **Year:** 2021

Summary of problem:
Pituitary Adenomas, or PAs, arise from the pituitary gland, comprise 10 to 15% of primary brain tumors, and is the third most common type of intracranial tumor. Consequently, segmentations of PAs are routine clinical tasks for treatment decisions, surgical planning, and radiation therapy. However, the manual segmentations are time consuming and laborious. Therefore, this paper

aims to automate segmentation with state-of-the-art results using nnU-Net due to its previous success with 3D biomedical semantic image segmentations.

Key result and Significance:
First, the paper found that the PA volume was one of the most crucial factors that affected their performance of nnUNet models. Specifically, small PAs had dice coefficient similarity score (DSC) of 0.47 while the medium PAs and large PAs had DSC value of around 0.85 which illustrates that the nnUNet had poor segmentation performance for small PAs. Thus, the study concluded that it is appropriate to use nnUnet for medium to large size PAs while it suggested using manual segmentations for small PAs, from a clinical perspective.

Necessary Background:
Deep learning outperforms traditional algorithms in medical image segmentation. nnUNet ("no new network") is a U-Net-based deep learning framework that has enabled successful 3D semantic segmentation of various biomedical image datasets and has been considered the strongest U-Net baseline. In this study, authors developed and evaluated the nnU-Net models to explore a cost-effective way to apply deep learning-based models to PA segmentation in clinical practice. The study had 243 PA MRI images total, where 35 of them were set aside as testing images and the other 208 images were used for training. The first model included all of 208 training images while the second model included only the subset of the training images, 109 cases of non-functional PAs. The study divided datasets into these two models (functional PAs vs. non-functional PAs) because non-functional PAs often look different structurally from functional PAs on imaging.

Technical Approach
Magnetic Resonance Image Dataset: MRIs were collected in DICOM format; T1ce (contrast-enhanced) images of the cases were used as the data samples for deep learning.

Tumor Segmentation: T1ce images in DICOM format were converted into NIFTI images. 243 PAs were segmented by a neurosurgeon specializing in PAs.

nnU-Net Framework: Due to its ease of use and adaptability to diverse biomedical image datasets, nnU-net was used as a deep learning-based segmentation method that automatically configures itself and executes the entire segmentation pipeline. From two nnU-Net models, the results were compared and the sub-group analysis of DSC on testing images were performed.

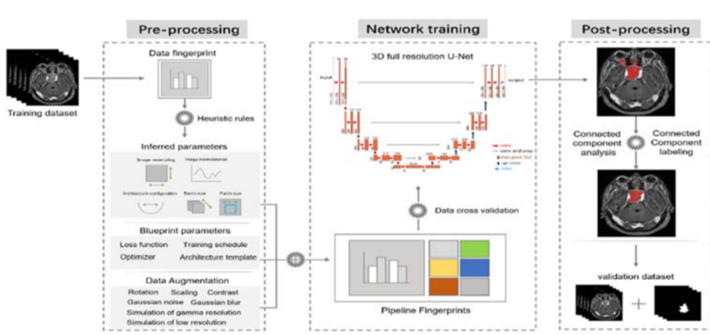

Figure 2 - nnUNet Framework for Pituitary Adenoma Segmentation

Critical Assessment

Both models had overall strong performances during the training (or, validation dataset), where the training loss converged at around 0.85 DSC value. After both models were trained and tested, the study performed a sub-group analysis and comparison of the two models to identify the performance bottleneck for nnUNet: images containing the small PAs. The small PA group had significantly lower DSC value compared to medium and large PA groups.

For the pros of this paper, authors attributed poor performance of nnUNet model for small PAs to the use of dice similarity score as a performance metric. For small structures, DSC is more sensitive because it penalizes errors in small segments to a greater degree than in larger segments. Also, the study points out a need for higher resolution imaging techniques because it was difficult to discern and segment the PA structures when the segmentations are very small. For the cons of this paper, the authors used a biased dataset because there was an uneven distribution of PAs in terms of size, where the medium size PAs constituted the majority for training, validation, and testing datasets. In addition, even though the authors identified the performance bottleneck, they did not attempt to modify nnUnet's DSC loss to resolve the performance bottleneck.

Overall, this study provided our group with several key takeaways. Since we already saw that nnUNet might have difficulties learning very small or thin structures, and the eustachian tube size is comparable to small PAs, we will not just use conventional DSC for our loss function and performance metric. We will look into new evaluation metrics such as the average or weight hausdorff distances, which might emphasize the small segments by assigning a higher weight to pixels in smaller segments. Also, this study showed that we might not always need very large data for state-of-art performance as long as there is a good data distribution, which may save us time from manually segmenting the eustachian tube images. Also, using a contrast-enhanced image during the preprocessing step may improve our nnUNet performance.

<u>Paper #2:</u> Fully Automatic Deep Learning Framework for Pancreatic Ductal Adenocarcinoma Detection on Computed Tomography[2]
Natália Alves, Megan Schuurmans, Geke Litjens, Joeran S. Bosm, John Hermans, and Henkjan Huisman
<u>Journal</u>: Cancers; **<u>Year</u>:** 2022

Summary of problem:
Detection of pancreatic ductal adenocarcinoma (PDAC) is challenging as they are usually small and poorly defined on medical imaging such as computed tomography (CT). Early diagnosis of PDAC may lead to improved survival. Therefore, the authors aimed to use a deep learning architecture to improve detection and localization of PDAC lesions on CT images. Furthermore, they assessed whether inclusion of surrounding anatomical structures as additional labels within the ground truth would improve the performance of the detection model.

Key result and Significance:
Three models were trained:

- nnUNet_T – ROI (Tumor)
- nnUNet_TP – ROI (Tumor and pancreas)
- nnUNet_MS – ROI (Tumor, pancreas, and surrounding anatomical structures)

The authors found that the two latter models had significantly better area under the curve (AUC) for classification of tumors when compared to the first model. Interestingly, with the inclusion of surrounding anatomical structures in addition to the tumor and pancreas, there were cases where the tumor was successfully detected in nnUNet_MS and not detected in nnUNet_TP. The results are significant as they reiterate the advantage of learning-based methods for medical image segmentation given their ability to learn not just from the configuration of the tumor but the anomalies within the tumor surroundings that may not be easily deciphered by the human eye. Therefore, such models have the potential to be integrated into the clinical domain and aid radiologists in early identification and localization of this challenging lesion.

Necessary Background

PDAC is the most common form of pancreatic cancer with a poor 5-year survival at 10.8%. Unfortunately, PDAC is rarely diagnosed in its early stages and most frequently, patients are diagnosed with a late stage or metastatic disease. This is due to several reasons: patients are initially large asymptomatic, and the lesions are difficult to localize on CT scans. This is especially the case if the lesions are smaller than 2 cm. Therefore, a platform that aids radiologists in detecting and localizing these lesions would be of immense clinical benefit.

Technical Approach

For the technical approach, the paper makes a hypothesis that state-of-art deep learning architectures can be used to detect and localize PDAC lesions accurately, especially regarding the subgroup of tumors with a size < 2 cm.

During the experiment, nnU-Net architecture was used to build three models. (1) segmenting only the tumor (nnUnet_T); (2) segmenting the tumor and pancreas (nnUnet_TP); and (3) segmenting the tumor, pancreas, and the multiple surrounding anatomical structures (nnUnet_MS). All of them will use the same CTs but with different segmentations. Then, these three models make inferences on ROI (region of interest) to get tumor likelihood. Additionally, the datasets are downsampled and used to train a low-resolution pancreas segmentation network in order to get regions of interest from full CTs.

Figure 3 demonstrates the pipeline for the PDAC framework. When the test CTs are input, they are firstly downsampled and the low_resolution pancreas segmentation network is used to obtain a coarse segmentation of pancreas. Then, this mask is upsampled back to full image resolution and a spherical kernel is applied to eliminate possible gaps. These images are then cropped based on an extraction margin to get the region of interest which is the input of the PDAC framework. Next, each of the PDAC detection models output a voxel-level tumor likelihood map to decipher where the network will predict PDAC lesions with the corresponding prediction confidence. nnUNet_TP and nnUNet_MS will also generate segmentations for pancreas that could be used in post-processing to reduce false positives outside the pancreas area. After post-processing, candidate PDAC lesions were extracted iteratively from the tumor likelihood map by selecting the voxel with a maximum predicted likelihood and finally we will get the maximum tumor likelihood.

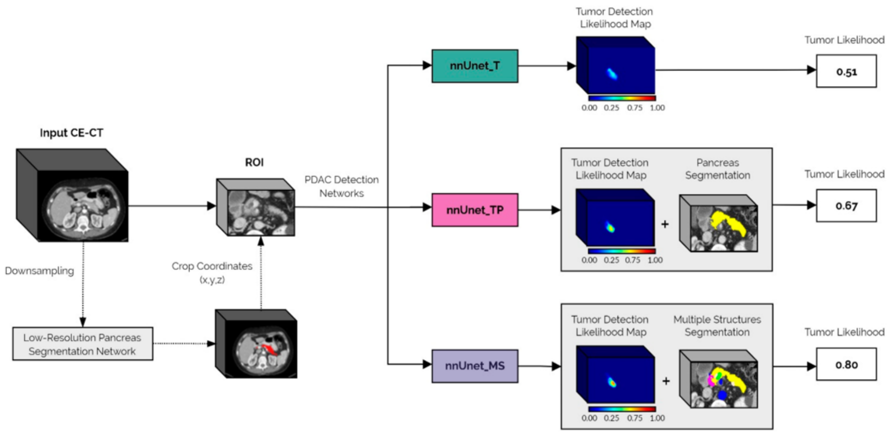

Figure 3 - nnUNet Architecture for Segmentation and Detection of Pancreatic Ductal Adenocarcinoma

Critical Assessment

This is a fully automated deep learning framework for detection and localization of PDAC with the potential to transform the care and survival of patients diagnosed with this lesion. Before this platform can be integrated into the clinical domain, we must address the pros and cons of this study.

First, the authors used an external validation dataset which further proves that their model does not exhibit a bias towards the training dataset. Second, they correctly deciphered that PDAC tumors would exhibit a change to their surroundings. Therefore, by including near-by anatomical structures in their model, they demonstrated increased performance for the detection and localization of PDAC. Finally, their inference methodology was simple and could be used by non-technical personnel for obtaining ROI predictions. However, they only included tumors that were present within the head of the pancreas; as a result, their model may not be generalizable to other datasets that contain PDAC lesions within the body or tail of the pancreas. Additionally, as with any other deep learning model, manual annotation of the ROI is laborious and time-consuming.

Overall, this study provided our group with several key takeaways. First, we may want to train two models: one which includes the eustachian tube only and second, which includes the eustachian tube along with other anatomical structures such as the internal carotid artery and torus tubarius. The default loss function for nnUNet is soft dice and binary cross entropy, however, the authors in this study found the cross-entropy loss function better in delineating the ROI. Therefore, we may look to modify the architecture of our deep learning model by changing parameters such as the loss function. Finally, our project uses the 3D full resolution version of nnUNet; however, using the 3D cascade which trains using both high- and low-resolution CT images may enhance the performance of our model and make it more generalizable to other datasets.

References:

1. Shu X, Zhou Y, Li F, et al. Three-Dimensional Semantic Segmentation of Pituitary Adenomas Based on the Deep Learning Framework-nnU-Net: A Clinical Perspective. *Micromachines*. 2021;12(12):1473. doi:10.3390/mi12121473

2. Alves N, Schuurmans M, Litjens G, Bosma JS, Hermans J, Huisman H. Fully Automatic Deep Learning Framework for Pancreatic Ductal Adenocarcinoma Detection on Computed Tomography. *Cancers*. 2022;14(2):376. doi:10.3390/cancers14020376