Deep Learning for Automatic Segmentation of Human Inner Ear

BMEN4460 Deep Learning in Biomedical Imaging Project Final Report

Author: Yunlin Zhou, Christopher Awosogba

ABSTRACT

This thesis aims to develop a deep learning approach for automatic segmentation of the human inner ear in micro-CT scan images, with the ultimate goal of optimizing microneedle design and drug delivery strategy for intracochlear access. Specifically, we implemented a U-Net neural network model and trained it using three datasets. We discussed potential improvements, such as using cross-validation, sliding window algorithm, and alternative loss functions and activation functions for non-binary mask data. Our study demonstrates the promise of deep learning in medical image segmentation and its potential for improving treatment options for inner ear diseases.

INTRODUCTION

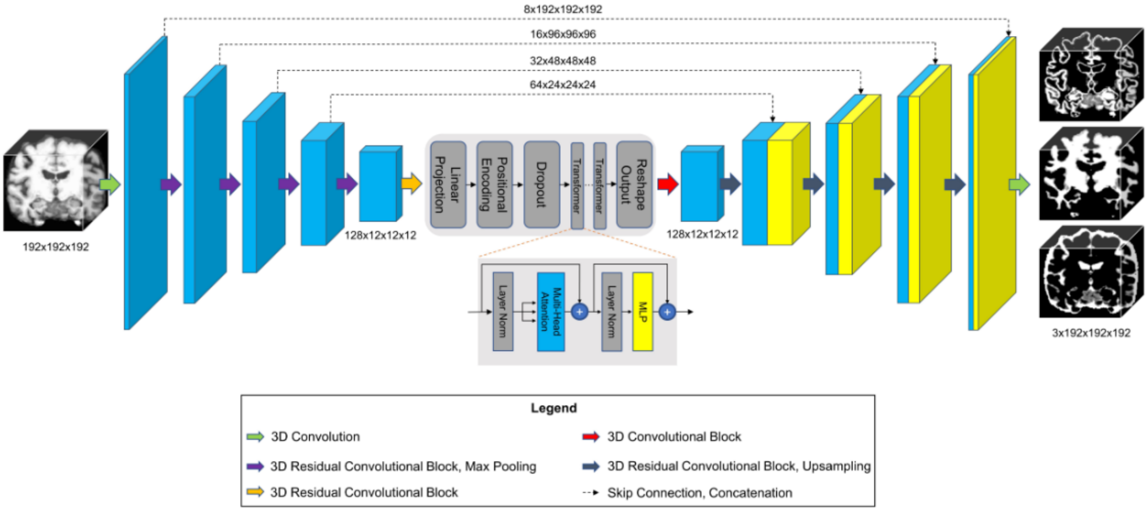
Disorders of the inner ear comprise the largest and most serious class of diseases responsible for hearing loss, with 250 million people worldwide suffering from disabling hearing loss [1]. The major challenge in treatment of inner ear diseases is the inaccessibility of targets for therapy, due largely to the presence of the blood-cochlear barrier, which oral medications are typically blocked by [2]. The round window membrane (RWM) is the only soft tissue portal from the middle ear into the cochlea and, therefore, is an ideal candidate for intracochlear access. Microneedle-mediated perforation of the RWM is a novel means of achieving intracochlear access and can facilitate reliable and predictable perforation of the RWM, with minimal anatomic and functional damage among the intracochlear delivery methods [3][4]. Real cochlea model simulation helps optimize the microneedle design and drug delivery strategy. Segmentation of micro-CT scan is the foundation of such model. However, manual segmentation requires experienced readers, is time-consuming and prone to intra-and inter-observer variability [5-7]. Recent studies have demonstrated the successful application of deep learning techniques for detection, segmentation, and classification tasks in the medical field [8]. Therefore, our study's objective was to develop a deep-learning approach for the automatic segmentation of the human inner ear in micro-CT scan images.

METHOD

We implemented TABS (Transformer-based Automated Brain Tissue Segmentation) for this experiment as it proved to be more efficient and adept at generalizing a variety dataset compared to CNN alone. This is because it relies on a Res-Unet backbone, with a Vision Transformer embedded between the 5-layered 3D CNN encoder and decoder layers, and the transformer consists of 4 layers and 8 heads.

Model Architecture

TABS takes an input dimension of 192 × 192 × 192, and the five encoder layers downsample the original image to f x28 × 28 × 28, where f represents the number of encoded features. We convert the encoded feature tensor into 512 tokenized vectors that are sequentially fed into the Transformer module in the order determined by the learned positional embeddings. Our Transformer encoder consists of 4 layers and 8 heads following the implementation. The output of the Transformer is 512 × 21,952, which we then reshape to 512 × 28 × 28 × 28 and reduce the feature dimensionality to f via convolution. The decoder portion of the network reconstructs the image to the original input dimension, and a final convolution operation is applied to generate a 3-channel output with each channel corresponding to an individual tissue type.



Experiment Set - Up

Original Data Set

For our research, we utilized three micro-CT scans of the human cochlea as our data sets. Each data set consisted of a raw micro-CT scan and its corresponding mask, which served as the ground truth. We separate the data sets into the training dataset (1R), the validation dataset (2R), and the test dataset (5R).

Data Preparation

To optimize the training outcome, we first utilized the "as_closest_canonical" function to reorient the CT scan image data and its corresponding mask data so that they matched. We normalized the intensity of the images to a range of -1 to 1 to reduce the influence of noise, artifacts, and other factors that can affect the model's accuracy. This rescaling of the intensity values to a common range enabled the model to learn more robust features and become less sensitive to variations in the input data.

As there were limited data sources for this project, we used random selection during data cropping. Our data sets had dimensions of (751, 469, 735), (579, 691, 774), and (695, 573, 782). We created a random selection data loader to crop the data into smaller pieces, selecting sample cubes of size (80, 80, 80) from the original data cube.

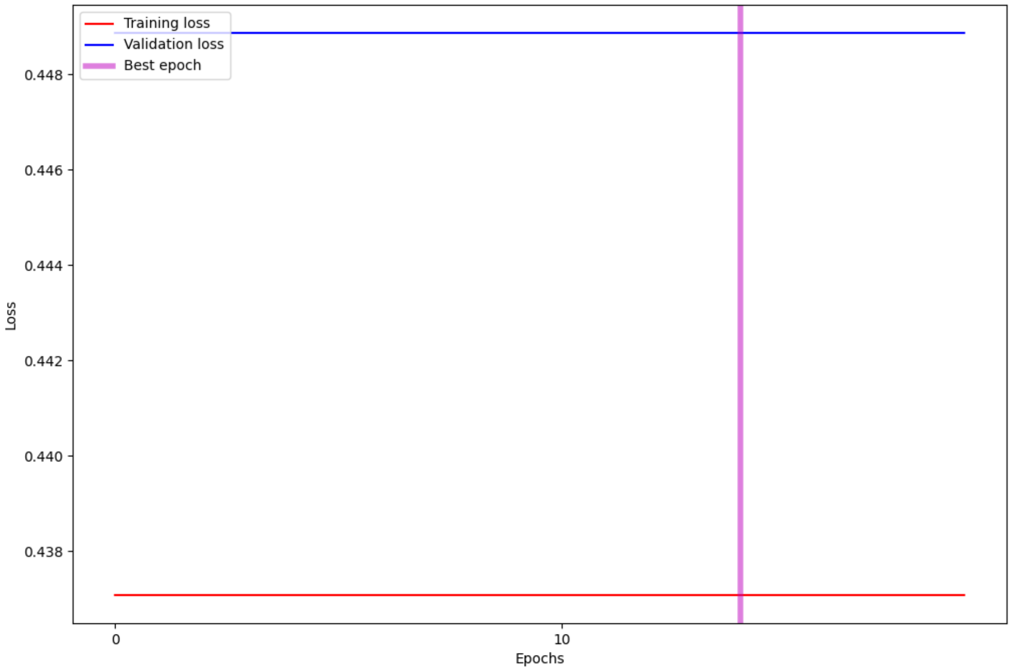
Training details

Due to the limitation of the Kaggle GPU, we trained our models for 20 epochs on 60 samples. The batch size was 3, the input channel size was 1, and the output channel was 5, with a learning rate of 1e-5.

RESULTS

Training Model

After running 20 epochs of training on the provided dataset, the best epoch was found to be epoch 14. However, the performance of the model was poor, with both the training and validation losses hovering around 0.44. Further analysis revealed that there was no significant advantage of epoch 14 compared to the other epochs. Overall, the training model did not perform well and may require additional adjustments or improvements in order to achieve better results.



Test Model

Unfortunately, we were not able to obtain the test results for our model due to some technical limitations. One issue that we encountered was that our mask data was not binary, and as a result, we could not use the original activation function (argmax) for the model predictions. Since we did not have enough information to set the threshold for the mask data, we were unable to figure out a new activation function that would be suitable for our needs. Another challenge was that due to the limitations of the Kaggle GPU, we could not run the training and validation data in the same notebook. As a result, we could not directly test the model after training, and we were not able to obtain the final performance results.

DISCUSSION

Firstly, while we used three datasets (train, validation, and test) in this project, a better way would be to use cross-validation to ensure that our model is more robust and can generalize well to new data. This would involve partitioning the data into several folds and training the model on each fold while testing it on the remaining folds. By doing so, we can better estimate the performance of our model on unseen data and avoid overfitting to our current datasets.

Secondly, for the validation and test data loaders, we could use a sliding window algorithm to improve the performance of our model. This would involve dividing the input image into smaller patches or tiles, running the model on each patch, and then stitching the results together to obtain the final output. By doing so, we can better capture the details and context of the inner ear structure and potentially improve the accuracy of the segmentation.

Thirdly, while we used the mean squared error (MSE) as the loss function in this project, we could try other loss functions such as the Dice loss or cross-entropy loss to see if they perform better. The Dice loss is commonly used for image segmentation tasks, especially when dealing with imbalanced classes, while the cross-entropy loss is a popular choice for multi-class classification problems. By experimenting with different loss functions, we can potentially improve the accuracy and robustness of our model.

Fourthly, as we encountered some issues with the activation function due to the non-binary nature of the mask data, we could explore different activation functions such as the sigmoid or softmax functions. Alternatively, if we have more information, we could set a threshold for the mask data and continue using the argmax activation function. By doing so, we can better handle the non-binary nature of the data and obtain more accurate segmentation results.

Finally, while the performance of our model on the inner-ear dataset was poor, we plan to improve it in the future by experimenting with different model architectures and hyperparameters, using more advanced techniques such as data augmentation and transfer learning, and incorporating domain-specific knowledge into the model design. Additionally, we plan to compare the performance of our model with other state-of-the-art models on this dataset to better understand its strengths and weaknesses.

REFERENCES

[1] Holley MC. Keynote review: The auditory system, hearing loss and potential targets for drug development. Drug Discov Today. 2005;10(19):1269-82. doi: 10.1016/s1359-6446(05)03595-6. PubMed PMID: 16214671.

[2] McCall AA, Swan EE, Borenstein JT, Sewell WF, Kujawa SG, McKenna MJ. Drug delivery for treatment of inner ear disease: current state of knowledge. Ear Hear. 2010;31(2):156-65. doi: 10.1097/ AUD.0b013e3181c351f2. PubMed PMID: 19952751; PMCID: PMC2836414.

[3] Szeto B, Chiang H, Valentini C, Yu M, Kysar JW, Lalwani AK. Inner ear delivery: Challenges and opportunities. Laryngoscope Investig Otolaryngol. 2020;5(1):122-31. Epub 20191211. doi: 10.1002/lio2.336. PubMed PMID: 32128438; PMCID: PMC7042639.

[4] Valentini C, Szeto B, Kysar JW, Lalwani AK. Inner ear gene delivery: vectors and routes. Hearing, balance, and communication. 2020;18(4):278-85.

[5] Nogovitsyn N, Souza R, Muller M, Srajer A, Hassel S, Arnott SR, Davis AD, Hall GB, Harris JK, Zamyadi M. Testing a deep convolutional neural network for automated hippocampus segmentation in a longitudinal sample of healthy participants. Neuroimage. 2019; 197:589-97.

[6] Men K, Zhang T, Chen X, Chen B, Tang Y, Wang S, Li Y, Dai J. Fully automatic and robust segmentation of the clinical target volume for radiotherapy of breast cancer using big data and deep learning. Physical Medica. 2018; 50:13-9.

[7] Akkus Z, Galimzianova A, Hoogi A, Rubin DL, Erickson BJ. Deep learning for brain MRI segmentation: state of the art and future directions. Journal of digital imaging. 2017; 30:449-59.

[8] Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, Van Der Laak JA, Van Ginneken B, Sánchez CI. A survey on deep learning in medical image analysis. Medical image analysis. 2017; 42:60-88.

[9] Yamashita R, Nishio M, Do RKG, Togashi K. Convolutional neural networks: an overview and application in radiology. Insights Imaging. 2018;9(4):611-29. Epub 20180622. doi: 10.1007/s13244-018-0639-9. PubMed PMID:29934920; PMCID: PMC6108980.

[10] Wang YM, Li Y, Cheng YS, He ZY, Yang JM, Xu JH, Chi ZC, Chi FL, Ren DD. Deep Learning in Automated Region Proposal and Diagnosis of Chronic Otitis Media Based on Computed Tomography. Ear Hear. 2020;41(3):669-77. doi:10.1097/aud.0000000000000794. PubMed PMID: 31567561.

[11] Zhang D, Wang J, Noble JH, Dawant BM. HeadLocNet: Deep convolutional neural networks for accurate classification and multi-landmark localization of head CTs. Med Image Anal. 2020; 61:101659. Epub 20200128. doi: 10.1016/j.media.2020.101659. PubMed PMID: 32062157; PMCID: PMC7959656.

[12] Cho YS, Cho K, Park CJ, Chung MJ, Kim JH, Kim K, Kim YK, Kim HJ, Ko JW, Cho BH, Chung WH. Automated measurement of hydrops ratio from MRI in patients with Ménière's disease using CNN-based segmentation. Sci Rep.2020;10(1):7003. Epub 20200424. doi: 10.1038/s41598-020-63887-8. PubMed PMID: 32332804; PMCID: PMC7181627.

[13] Rao VM, Wan Z, Arabshahi S, Ma DJ, Lee P-Y, Tian Y, Zhang X, Laine AF, Guo J. Improving across-dataset brain tissue segmentation for MRI imaging using transformer. Frontiers in Neuroimaging. 2022;1. doi:10.3389/fnimg.2022.1023481.