Two-Stage Deep Learning Pipeline for Prostate and Lesion Segmentation in MRI Scans

Ryan McGovern, Yoohyuk Chang, Avery Kuo, Sujin Lee

EN.601.475 - Final Project Report

December 6, 2024

Abstract

Prostate cancer segmentation plays a critical role in surgical planning and prostate brachytherapy, where accurate delineation of prostate and cancerous lesions is essential. This study proposes a two-stage deep learning approach for automated segmentation of prostate cancer from MRI data. In the first stage, a Transformer-U-Net model is trained to segment the prostate region by preprocessing the data into NIfTI format, resampling for fixed spacing, and normalizing intensities. Evaluation metrics, including the Dice similarity coefficient (0.85) and the Hausdorff distance, are used to assess performance. In the second stage, a bounding box around the segmented prostate is generated, and another model is trained to localize cancerous lesions within this region. The lesion segmentation model also outputs the results in NifTI format and is evaluated using the same metrics. This pipeline enables precise segmentation of both the prostate and internal lesions, serving as a foundation for grading identified lesions in future work.

1 Introduction

1.1 Background

Prostate cancer is the second leading cause of death among American men, accounting for 1 in 44 deaths [1]. The prostate is a small walnut-shaped gland located just below the bladder. It plays a vital role in male reproductive health by producing seminal fluid that nourishes and transports sperm. Prostate cancer can vary greatly in its behavior; some forms progress slowly and may not cause significant harm. Other forms are more aggressive and spread rapidly to different parts of the body. Early detection is important for successful treatment outcomes. Localized prostate cancer can often be effectively treated with therapies aimed at eradicating the tumor. However, late-stage detection when cancer has metastasized potentially leads to limited treatment options and poor prognoses, which highlights the need to improve diagnostic tools.

1.2 Objective

The objective of this project is to develop a robust and efficient method to identify and localize prostate cancer from magnetic resonance imaging (MRI) scans. Using advanced deep learning techniques, the project creates two transformer-based models capable of performing precise segmentation tasks. These models focus on determining the boundaries of the prostate gland and identifying lesions indicative of cancerous growths.

1.3 Purpose

This work has significant potential to enhance current prostate cancer treatment planning. One leading treatment for prostate cancer is called prostate brachytherapy, where radioactive seeds are surgically placed on the cancerous legion, delivering localized radiation [2]. The biggest risks of this treatment are missing parts of the cancerous region, allowing it to survive and proliferate, or placing radioactive beads on healthy tissue, eventually killing them. These issues can be mitigated by properly segmenting the prostate legion, allowing better planning through improvements in dosage precision and pre-planned seed placement. However, manual segmentation can take over a minute per slice [3]. As standard T2-diffused MRI slice thicknesses are around 1.5mm, this can take well over 30 minutes. If staff is short, this need for manual segmentation may cause patients to wait much longer than 30 minutes or even have to return another day, all while their cancer continues

to attack their bodies. Therefore, there is a need for a reliable prostate legion segmentation model that can be leveraged to make the slow manual segmentation process quick and semi-automatic. To address this need, we created a pipeline of data processing and two U-Net transformer models that can output prostate and legion segmentations. The pipeline consists of data pre-processing, a prostate segmentation (M1) model, a legion segmentation model on a resulting bounding box (M2), and post-processing.

2 Data Processing

2.1 Input Data

First, we downloaded the 'Image' files from The Cancer Imaging Archive [4] using the NBIA data retriever [5]. These consist of 1,152 patients with prostate cancer and their T2-weighted MRI scan(s). Next, we downloaded their corresponding Digital Imaging and Communications in Medicine (DICOM) annotations, curated by Cosmin et al [6].

2.2 Conversion to NIfTI

To more easily integrate these files (both images and annotations) into a PyTorch model training loop, we converted them into the Neuroimaging Informatics Technology Initiative (NIfTI) format using the NiBabel and Slicer Python packages [7, 8]. DICOM and NIfTI are medical image formats, allowing easy access to 3D slice information, affine transformations, headers, and more. Most importantly, they make it possible to automatically align images in space through resampling using the NiBabel package. We decided to use NIfTI over DICOM because creating readable code with NiBabel is easier.

The image files were sorted by patient upon download, some of whom had multiple T2-weighted MRIs. We ignored patients with multiple MRIs on file to ensure we did not have any mismatch between MRI scans and corresponding segmentations. Additionally, not all had segmentations. These patients were also ignored. This left us with 473 images. Of these, 80% were designated for training, 10% for validation, and 10% for testing.

2.3 Cropping with Bounding Boxes

Next, we created cropped images of the prostate and corresponding segmentations. First, we extracted the minimum and maximum indices for each ground truth prostate segmentation. Because the image and both segmentations have different affines (consistent spacing but different origin locations), we had to convert these indices into world coordinates and then into indices for the image and lesion segmentation files using the following equations:

$$v_{\text{img}} = A_{\text{img}}^{-1} A_{\text{pro}} v_{\text{pro}} \quad \text{(Eq 1)}$$

$$v_{\text{leg}} = A_{\text{leg}}^{-1} A_{\text{pro}} v_{\text{pro}} \quad \text{(Eq 2)}$$

where $v_{\text{img}}, v_{\text{pro}}, v_{\text{leg}}$ are the corresponding points relative to the image, prostate, and lesion indices, respectively, and $A_{\text{img}}, A_{\text{pro}}, A_{\text{leg}}$ are the affines of the image, prostate, and lesion, respectively.

Once all minimum and maximum indices were calculated, we cropped the original image and segmentations and added a 10-voxel padding along the axial plane and a 4-voxel padding along other planes.

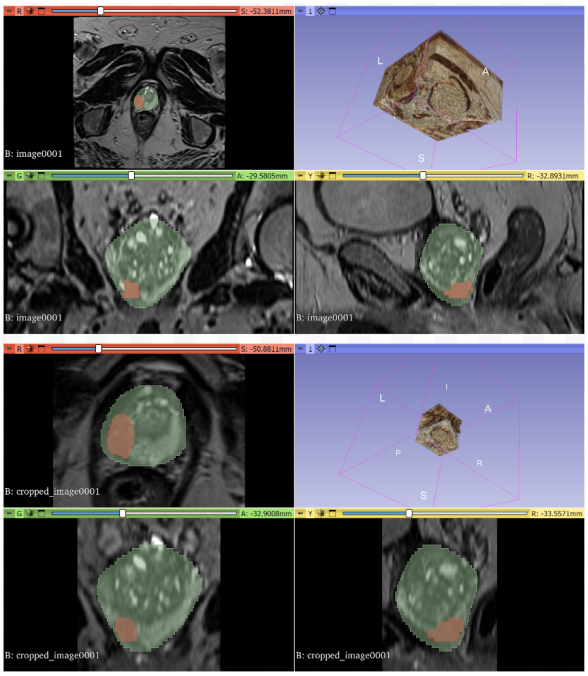

Figure 1: Comparison between original image and prostate/legion segmentations (top) and cropped image and prostate/legion segmentations (bottom)

2.4 Property Extraction

We extracted specific volumetric properties from the dataset. This includes the shape of the images (number of voxels), spacing (height, width,

and depth of voxels), and bounding box information for ground truth prostates. We used the extracted bounding box information to create new files zoomed in on the prostate. This will be used to train the second model in our pipeline. We then extracted the same shape information from the cropped images (spacing will be the same as before).

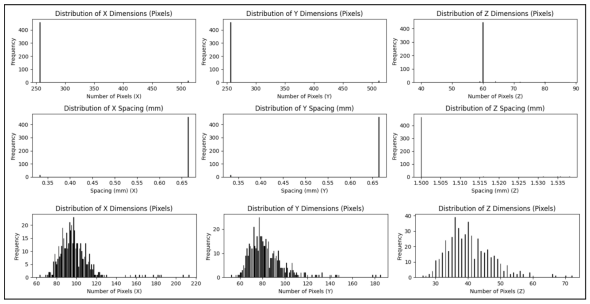

Figure 2: Distribution of image shapes (top), spacing (middle), and prostate shapes (bottom)

2.5 Transforms

As shown in Figure 2, Nearly every MRI image is $256 \times 256 \times 64$ pixels, with spacings $0.66 \times 0.66 \times 1.5$ mm. We chose to use $128 \times 128 \times 64$ pixels and $1.5 \times 1.5 \times 1.5$ mm spacing for M1. Using the raw input was an unnecessary level of precision when looking at the entire MRI. We used $128 \times 128 \times 64$ pixels and $0.66 \times 0.66 \times 1.5$ mm spacing for M2. We used a higher resolution because more precise information was needed to segment small legions. This is accomplished through a MONAI transform that resamples and crops/pads as needed [9]. As the prostate is near the center of every MRI, we do not have to worry about it being cut off for the larger outliers as long as cropping is centered about the image center. The MONAI transform also normalizes based on the 10th and 90th percentile intensities and orients all images to the 'RAS' direction (the x-axis runs from right to left, the y-axis runs from axial to sagittal, and the z-axis runs from superior to inferior). This is important because different MRI scanners may have slightly different image intensity properties and orientations, so setting a standard allows all to be valid inputs.

3 Model Implementation

We use a modified version of the standard U-Net model using convolutional and transformer blocks. A typical U-Net is a convolutional neural network (CNN) with an encoder and decoder path [10]. The encoder path captures context and features, while the decoder path restores lost spatial information using concatenations between corresponding levels on the encoder and decoder paths. This makes it great at segmentation tasks and has become a standard model for medical segmentation.

3.1 Overview of the U-Net Model:

We use a modified version of the standard U-Net model using convolutional and transformer blocks. A typical U-Net is a convolutional neural network (CNN) with an encoder and decoder path [10]. The encoder path captures context and features, while the decoder path restores lost spatial information using concatenations between corresponding levels on the encoder and decoder paths. This makes it great at segmentation tasks and has become a standard model for medical segmentation.

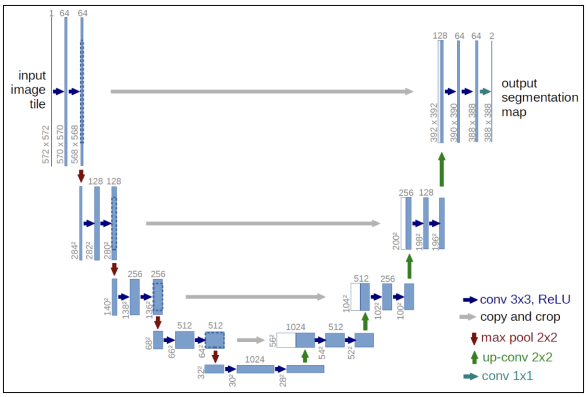

Figure 3: U-Net architecture

3.1.1 Transformers in Medical Segmentation Tasks:

That being said, using attention in a manner like a transformer model is an emerging trend in medical segmentation tasks, leading to the creation of 'Image Transformers' [11]. Image transformers work by separating the image into distinct patches. These patches then become the subject of the typical query (Q), key (K), and value (V) self-attention model in a transformer. Both the patch values and location relative to each other are relevant in generating the key and value. Once the patches are created, you can think of these image transfers as very standard transformers, assuming patches are handled in sequence (patches are akin to words in LLM transformers). However, it is more common

for all patches or ones within a certain distance (Euclidean or Manhattan are used), depending on if the self-attention is global or local, to be looked at regardless of where in the sequence of patch processing they are found.

3.1.2 Model Architecture and Training Details:

Our models (M1 and M2 are the same model with different transforms for their corresponding datasets) use both convolutional and transformer blocks during the encoder path. The convolutional blocks use 3x3x3 convolutions and a reLU activation function. The transformer blocks use multi-head self-attention perceptron followed by layer normalization. The input is split into 4x4x4 patches for this to work. There is one more convolution path than the transformer path. This allows us to use the same method to transition between encoder and decoder paths as U-Net. Like in a normal U-Net decoder, the convolution block outputs are concatenated to up-convolved prior layers. However, this time the output of each transformer block is also concatenated at the corresponding resolution. We used PyTorch and Medical Open Network for AI (MONAI) to create this model [9, 12]. Training was set for 150 epochs with early stopping. Due to the early stopping, M1 finished training in 35 epochs and M2 in 49. The models with the best validation dice were kept.

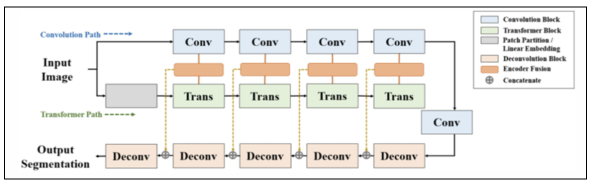

Figure 4: Our model architecture

4 Results

4.1 Model 1: Prostate Segmentation Model

Our model achieved a Dice coefficient of 0.61 with a standard deviation of 0.22 and a Hausdorff distance (95th percentile) of mean 9.44 with a standard deviation of 6.09. The model demonstrated decent segmentation performance, accurately capturing the shape and size of the prostate. This represents a significant improvement, as the manual matching process (including rotation and shift) took less than a minute per prostate, compared to

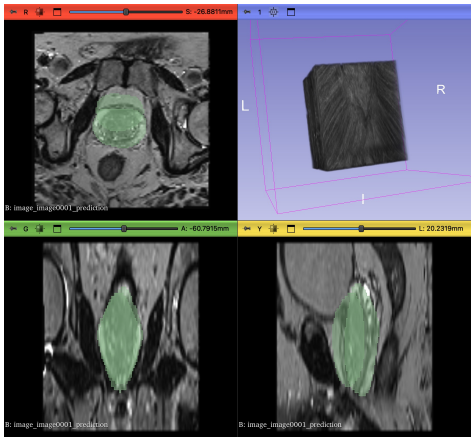

Figure 5: Unaligned Orientations of the Segmented Prostate Region Generated from the Prostate Segmentation Model

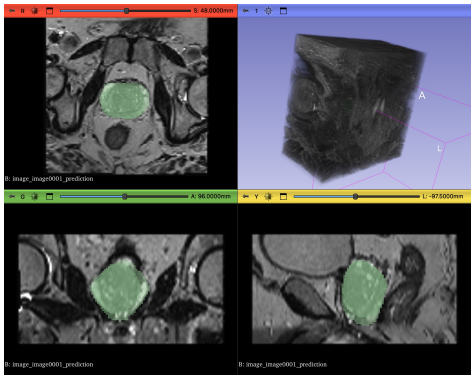

Figure 6: Aligned Orientations of the Segmented Prostate Region Generated from the Prostate Segmentation Model Through Post-Processing

the 30+ minutes typically required for fully manual segmentation by an X-ray technician. We believe this model has the potential to be highly beneficial for prostate brachytherapy planning. However, there remains an issue with the transform that needs to be addressed. For example, although a Dice coefficient of 0.87 was achieved on the validation set during training, the coefficient dropped to __ during deployment. We hypothesized that this problem arised from in the saving or inference process. Despite this limitation, our results demonstrate the potential of the model to segment other organs, provided there is sufficient contrast between the organ and the background. Resolving this issue will be a critical next step in unlocking the broader applicability of our approach.

4.2 Model 2: Lesion Segmentation Model

Our model achieved a Dice coefficient of 0.04 with a standard deviation of 0.08 and a Hausdorff distance (95th percentile) of mean 43.26 with a standard deviation of 20.21. These results highlight the inherent limitations of MRI scans, particularly T2-weighted imaging, in visualizing cancerous lesions. Even when accounting for the resolution of misalignment issues, the model's performance remained suboptimal, underscoring the challenges posed by insufficient contrast in the imaging modality. This limitation is more reflective of the imaging technique itself rather than the model's capabilities, as T2-weighted MRI is not ideally suited for detecting prostate lesions. Its primary objective was to reduce training loss by focusing on segmenting the correct region while minimizing extraneous areas. To achieve this, the model prioritized capturing the majority of the prostate, which led to suboptimal precision. One interesting result we can conclude from this is that legions appear uniformly across all regions of the prostate. If this were not the case, then the model would focus in on a specific region to reduce average dice across tens of thousands of iterations. Achieving higher accuracy for detecting cancerous lesions using this approach will remain a challenge unless different imaging modalities are used or the development of better contrasting agents are realized. These findings emphasize the need for improved imaging modalities to address the limitations of current methods.

5 Conclusion

5.1 Difficulties Encountered

One of the primary challenges we faced was managing variations in image orientations and origins across datasets. Each image and its corresponding segmentation file had unique affine transformations which makes it difficult to directly align them for further processing. These differences often caused misalignment during preprocessing steps such as cropping and padding. Addressing these inconsistencies required careful post-processing adjustments to ensure that all data was properly aligned in the same coordinate space.

A significant limitation of our preprocessing pipeline was the inability to retain metadata after performing cropping and padding operations. Metadata such as image orientation, resolution, and origin plays a critical role in maintaining spatial integrity, especially in medical imaging. Without this information, it was not straightforward to reverse these transformations or align the processed data with the original images. To solve this, we had to manually realign the images and segmentations which is time-consuming and more likely than not lead to errors. Through this, we realized the importance of having robots post-processing pipelines.

Furthermore, we attempted to work directly with DICOM files for both MRI images and segmentations. However, there were multiple difficulties in doing so. Many MONAI functions, which form the backbone of our pipeline, were designed to work seamlessly with NIfTI files and lacked robust support for DICOM. Thus, we had to convert all DICOM files into the NIfTI format which was difficult because of the heterogeneous nature of the DICOM data. Converting DICOM files to NIfTI was challenging because the MRI images and segmentations were stored as different types of DICOM files. MRI images were organized as folders containing multiple slice files, while the segmentation data often came as a single file. Merging these slices into a single volumetric file in the .nii.gz format required extensive preprocessing and careful handling to ensure that spatial relationships between the images and segmentations were preserved. This step was critical for downstream processing and training but introduced significant complexity to the workflow.

5.2 Implications of Results

The objective of our project was to develop models capable of accurately identifying the prostate region and detecting cancerous lesions within that region. By leveraging advanced image segmentation techniques and deep learning algorithms, we aimed to create a framework that supports more efficient and accurate prostate cancer diagnosis.

Our results demonstrate that the model successfully identified cancerous lesions. However, the detected lesion regions were often misaligned within the prostate boundaries. This requires additional post-processing to correct this spatial discrepancy. Despite this limitation, the results highlight the potential of the model as a foundation for further optimization and clinical application.

Our findings have significant implications for the use of machine learning in medical imaging. The ability of the model to roughly identify can-

cerous lesion regions, even with some misalignment, demonstrates its potential to enable faster and more accurate detection of prostate cancer. This capability can reduce the time required for manual analysis by clinicians and enhance diagnostic workflows. Post-processing to correct spatial misalignment can further improve the precision of lesion localization which creates more reliable outcomes. By automating key steps in the diagnostic process, the model has the potential to support clinicians in making informed decisions about prostate cancer diagnosis and treatment.

5.3 Potential Extended Work

A potential extension of our current work is exploring the use of the first model's predicted prostate region as input for training the second model to find the cancerous legion. This approach could provide a comparative analysis against our current second model and offer insights into how we can improve the overall accuracy of our legion segmentation. By leveraging predictions from the initial model, we could potentially refine and enhance the second model's performance, particularly in handling cases where precise localization is critical.

One challenge that currently face is the misalignment that occurs during the cropping and padding process. These operations, while integral to our current model pipeline, inherently disrupt spatial coherence, making it difficult to apply inverse transformations. By extending our work to address this limitation, we could open new possibilities for leveraging more sophisticated preprocessing and postprocessing techniques. Thus, this can allow for improved registration and alignment in the segmentation workflow.

Our model has demonstrated strong performance in segmenting the prostate which is an organs. This shows us that it can be applied and utilized in broader medical imaging applications. Beyond the prostate, this framework could be adapted to segment other body parts, particularly where organ shapes are indicative of disease. By extending this approach, we may uncover vital patterns and correlations in organ morphology. This can better future diagnostic tools and treatment planning in various medical domains.

References

1. https://www.cancer.org/cancer/types/prostate-cancer/about/key-statistics.html

2. Zaorsky NG, Davis BJ, Nguyen PL, et al. The evolution of brachytherapy for prostate cancer. *Nat Rev Urol.* 2017;14(7):415-439. doi:10.1038/nrurol.2017.76

3. Badiei S, Salcudean SE, Varah J, Morris WJ. Prostate segmentation in 2D ultrasound images using image warping and ellipse fitting. *Med Image Comput Comput Assist Interv.* 2006;9(Pt 2):17-24. doi:10.1007/11866763$_3$

4. Natarajan, S., Priester, A., Margolis, D., Huang, J., Marks, L. (2020). Prostate MRI and Ultrasound With Pathology and Coordinates of Tracked Biopsy (Prostate-MRI-US-Biopsy) (version 2) [Data set]. The Cancer Imaging Archive. DOI: 10.7937/TCIA.2020.A61IOC1A

5. Downloading TCIA Images - TCIA Online help - Cancer Imaging Archive Wiki. https://wiki.cancerimagingarchive.net/display/NBIA/Downloading+TCIA+Images

6. Ciausu, C., et al. DICOM Converted Annotations for the Prostate-mri-us-biopsy Collection. *Zenodo,* 3 Nov. 2023. doi:10.5281/zenodo.10069911

7. *Neuroimaging in Python — NiBabel 5.4.0.dev1+g3b1c7b37 documentation.* https://nipy.org/nibabel/

8. *3D Slicer image computing platform. 3D Slicer.* https://www.slicer.org/

9. *MONAI - home.* https://monai.io/

10. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv [csCV].* Published online 2015. http://arxiv.org/abs/1505.04597

11. Parmar N, Vaswani A, Uszkoreit J, et al. Image Transformer. *arXiv [csCV].* Published online 2018. http://arxiv.org/abs/1802.05751

12. *PyTorch. PyTorch.* https://pytorch.org/