Comparative Study between Single and Two-Step Approaches on Automatic Segmentation of Prostate Abnormalities Using Deep Learning

Cheh Kit Hong[1][0009-0000-1206-6793], Syaheerah Lebai Lutfi[1,3][0000-0001-7349-0061], Gayathri Delanerolle[2][0000-0002-9628-9245], and Yassine Bouchareb[3][0000-0001-8819-9736]

[1] School of Computer Sciences, Universiti Sains Malaysia, Penang 11800, Malaysia
[2] Research & Innovation Department, Hampshire & Isle of Wight Healthcare NHS Foundation Trust, Southampton SO30 3JB, UK
[3] College of Medicine and Health Sciences, Sultan Qaboos University, Muscat 123, Oman

Abstract. Accurate diagnosis through MRI is the main non-invasive method in the diagnosis of PCas through manual segmentation of the prostate lesion. This process is time-consuming and prone to interreader variability. Deep learning-based image processing has emerged as a solution to the problem. However, most deep-learning-based approaches in prostate lesion segmentation use a one-step direct lesion segmentation approach, which may cause the model to learn from unnecessary inputs that could contribute to noise. To address this problem, some researchers suggested introducing prior prostate anatomical knowledge using a two-step approach that involves segmenting the prostate gland and using the extracted prostate gland as the input for the lesion segmentation model. Despite demonstrating high performance in lesion segmentation, there remains a lack of systematic comparative studies that quantify its practical advantages using standardised metrics, consistent architectures, and the same dataset. In this study, a comparative evaluation of one-step and two-step approaches in prostate lesion segmentation was conducted using the benchmark dataset Prostate158 with DeepLabV3, U-net, TransUNet, and TransAttUnet. The findings reveal that all four models with the two-step approach achieve significantly higher results relative to the one-step approach, with a percentage increase of 160.0%, 105.4%, 100.6%, and 68.1% for DeepLabV3, U-net, TransUNet, and TransAttUnet, respectively. However, the sequential dependencies of the two-step approach may also introduce error propagation and inconsistent results.

Keywords: Deep Learning, Computer Vision, Prostate Cancer, Semantic Segmentation

1 Introduction

As one of the most common and lethal cancers affecting men worldwide, early detection and effective diagnosis of prostate cancer (PCa) are critical for deciding the most appropriate medical treatment. Magnetic Resonance Imaging (MRI) plays a vital role

in helping medical experts diagnose PCas by providing a non-invasive, clear internal body structure. This allows the experts to visually distinguish and segment the region of the prostate affected by the disease, commonly known as prostate lesions. However, the manual segmentation of MRI images is time-consuming and prone to interreader variability, leading to inconsistencies in diagnosis [1].

To address this issue, Computer-aided Diagnosis (CAD) that utilises computational or traditional machine learning algorithms to handle segmentation of prostate lesions is proposed [2]. In this context, various deep-learning approaches are at the forefront of innovation. Deep learning (DL) provides a non-biased yet robust CAD method in the clinical workflow. Various deep learning architectures, such as Convolutional Neural Networks (CNN), Transformer, and CNN-Transformer hybrids, are promising solutions that allow them to capture local or global image feature representation [3].

Despite their advantages, DL approaches mostly use a single-step, direct lesion/tumour segmentation method, which may require the models to learn additional parameters to ensure a good performance for lesion localisation and segmentation. Recently, researchers have suggested incorporating prior anatomical knowledge of the prostate zones/gland to benefit and aid in model learning. [4]. However, there is a lack of research that directly studies the performance difference between two different models, incorporating prior anatomical knowledge and models that learn from scratch. This study aims to address the gap and to provide a quantitative analysis and comparison between the single and two-step approaches in lesion segmentation using the same model architecture, dataset, and performance metrics.

2 Related Works

2.1 Background and DL models on Prostate Image Segmentation

The traditional image-based method of diagnosis and determination of PCa lesions through MRI is performed by manual segmentation. However, such a method requires human expert knowledge and labour, making it time-consuming while suffering from interreader/interobserver variability even among highly knowledgeable experts [1]. Various computational algorithms for image segmentation have been employed. This includes classical computer vision techniques such as active contour models [5] and also machine learning algorithms [2]. However, due to the multi-sliced nature of MRI images, these techniques faced limitations in maintaining the consistency of the segmentation, resulting in relatively lower segmentation accuracy [2, 5]. The use of DL models demonstrates the possibility of getting the computer to perform the segmentation tasks, which effectively eliminates the subjectivity of humans.

U-net is a DL model designed for the image segmentation task [6]. U-net incorporates an encoder-decoder architecture where the encoder is responsible for feature learning while the decoder is responsible for restoring spatial information lost during encoding. Skip connections from the encoder to the decoder are built by concatenating feature maps from the early stage to the later stage to preserve both low and high-level detail information. [3]. Another CNN segmentation model includes DeepLabV3, which introduced a more advanced method known as Atrous Spatial Pyramid Pooling (ASPP)

to apply Atrous Convolution (AC) in a parallel manner with different dilation rates, enabling the model to understand the context at different scales [7].

Recent architectures that use the transformer encoder have also been used in this task [8]. For instance, the Swin-UNETR model uses a Swin Transformer as the encoder to learn features from non-overlapping 3D patches of MRI data using shifted windows and passes it to a CNN decoder [9]. Another architecture is the TransUNet proposed in [10]. TransUNet incorporates CNN as the encoder backbones to capture the local feature representation and pass the learned feature maps to the transformer encoder for learning global feature representation. Another similar U-based CNN-Transformer hybrid model is TransAttUnet by [11]. It also uses a CNN backbone and passes the features captured to two different transformer modules, which are Transformer Self-Attention (TSA) and Global Spatial Attention (GSA), for multiscale feature extraction.

Another model, known as the Segment Anything Model (SAM), was developed by the team MetaAI [12] that incorporates user prompts in the form of points or boxes to specify the exact region for the segmentation target to enhance the model's flexibility and zero-shot generalisation. Later research [13] reported that the SAM can be fine-tuned for medical image segmentation. However, the model was optimised by taking the user prompt, which may pose challenges in implementing the model in the automated workflow. Table 1 below presents a summary of the DL architectures reviewed.

Table 1. Summary of the DL models on lesion segmentation

Experimented Model	Dataset	Segmentation Scope	Score
U-net [14]	Private dataset of 40 patients' MRI images	Prostate lesion	91.76 DSC
DeepLabV3 [15]	Private dataset of 50 patients' pelvic MRI	Prostate lesion	73.97 IOU
Swin-UNETR [16]	PICAI	Prostate lesion	0.762 PICAI score
TransUNet [10]	ADC MRI on pelvic*	Pelvic organ	88.27 DSC
TransAttUnet [3]	Clean-CC-CCII*	Pelvic organ	86.57 DSC
MedSAM [13]	Internal validation of the prostate MRI dataset	Pelvic lesion	97.2 DSC

2.2 Basics and Gap Analysis on the Two-Step Approach

As mentioned in the section above, most of these deep learning approaches in prostate lesion segmentation only involve the direct lesion segmentation step. The sequential, two-stage strategy is specifically designed to overcome inherent challenges in direct, end-to-end lesion segmentation, particularly due to the difficulty posed by the small size of prostate lesions relative to the entire MRI image by introducing prior anatomical knowledge to the model. By accurately localising the prostate first, the search space for lesions is drastically reduced, leading to improved computational efficiency and enhanced accuracy in lesion detection [17]. The rationale for this approach is inspired by

the conventional radiologist's clinical workflow in manual segmentation of the lesion. The approach attempts to emulate how radiologists manually delineate the prostate gland in their minds before searching for the lesions within it [4]. By introducing the prostate gland region into the model, the model can focus its learning and inference on a much smaller, more relevant anatomical area. This approach is similar to the action of central cropping that crops out the important region of interest (ROI) of the prostate for the machine/deep learning model to focus on segmenting the region within it, effectively reducing false positives outside the prostate and improving the signal-to-noise ratio for lesion detection [18]. But instead of cropping a central region, this approach performs a "smart cropping" mechanism that only perfectly fits the whole prostate gland while eliminating other pixels that are outside the prostate gland segment boundary. Table 2 summarises the attempts at the experiments in a two-step approach.

Table 2. Summary of past works on the two-step lesion segmentation approach

Model	Descriptions	Results/Performance metrics
3D Attention Guided Unet (AGUnet) [4]	One AGUnet was trained for gland segmentation to extract the gland region and pass it to another AGUnet for lesion segmentation.	DSC 0.82 and AUROC 0.85 on PICAI
Mask R-CNN + Weakly Supervised Deep Neural Network (DNN) [17]	A Mask R-CNN was trained to segment prostate structure and pass it to the DNN for the lesion detection and classification.	AUROC 0.912 and 0.882 on ProstateX and the authors' local cohort
Unet [19]	One Unet was trained for gland segmentation to extract the gland region and pass it to another Unet for lesion segmentation.	DSC 0.87 on the author's internal dataset
Squeeze-Excitation CNN (SECNN) and Residual Attention Unet (RAU-net) [20]	Using SECNN to identify the presence of prostate for each slice in the 3D volume and passing the result to RAU-net for lesion segmentation.	0.860 DSC on authors' internal dataset
Cascaded scoring CNN [21]	Locating the ROI (prostate lesion) to score the dominant intraprostatic lesions and followed by segments further the lesion via a cascaded CNN.	84.3% DSC on private dataset of 70 patients

While researchers have proposed multiple DL architectures for the two-step framework that showed promising performance in lesion segmentation, there is a lack of systematic empirical comparison between the two-step segmentation and direct segmentation methods under the same training configuration. Additionally, the datasets used in these studies were extracted from different sources and may have used different data configurations, like 2D or 3D inputs, resolution, and stacking of different MRI

modalities that lack a unified standard for evaluation. This gap highlights the need for a comprehensive comparative evaluation of two-step segmentation and direct segmentation methods to determine their relative performance and practical usability using a standardised model, metrics, and other configurations.

3 Methodology

3.1 Dataset

The dataset used in this study is Prostate158 [22]. Prostate158 is a well-maintained dataset of parametric 3-Tesla prostate MRI images. The dataset contains 158 independent cases, of which 139 were released as the training and validation set in [23] while the remaining 19 cases were released as the testing set in [24]. All Prostate158 cases include T2-weighted (T2W) and diffusion-weighted (DWI) images with apparent diffusion coefficient (ADC) maps for the pelvic MRI sequence. Each MRI also comes with a pixel-wise gland mask (central zone and transitional zone) and PCa lesion mask, which are annotated by two board-certified radiologists with 6 and 8 years of experience in uro-oncologic imaging, that are resampled into the same orientation, direction, and spacing. However, in this study, the central and transitional zones are combined into one single class to reduce the task complexity. This approach also accounts for the fact that other datasets may label only the prostate gland as a whole, without distinguishing between the central and transitional zones, ensuring a lower effort is required to reproduce the same result of this study with different datasets. Figure 1 shows an example of an MRI slice for the Prostate158 after combining the gland classes.

Fig. 1. Example of Prostate158 MRI slice

3.2 Segmentation Workflow (ROI Lesion Segmentation)

Initial Data Preparation - This workflow starts with the data preprocessing and augmentation step. The augmentations include random rotation, random horizontal flip, random vertical flip, and elastic transformation. After that, each MRI and gland mask is resized into a shape of (256, 256) and normalised. Each 2D MRI slice will be treated as one sample input for the model.

Stage 1 Gland Segmentation – The processed MRIs and gland masks will be used as the input for the gland segmentation model. In this step, a gland segmentation model will be trained until it can accurately segment the prostate gland to ensure the model can define a high-quality ROI for the later lesion segmentation task. The U-net model

is selected for this stage as it is lightweight and proven to show accurate and robust results in organ segmentation.

Stage 2.1 Data Preparation - With a good gland segmentation model established, this model will be used to generate the gland mask for the dataset. The gland masks generated will be used in the ROI extraction step, where all the gland masks will be applied to their corresponding MRI slices to extract the ROI prostate gland. After the ROI extraction, only the remaining slices with the presence of the prostate gland will be passed as the input for the later lesion segmentation stage. After aligning all the extracted ROI MRI slices with their corresponding lesion masks, they will be passed to another preprocessing and augmentation pipeline similar to the gland segmentation stage.

Stage 2.2 Lesion Segmentation - This segmentation task is also a binary segmentation in segmenting the region affected by cancer. Models used in this step include U-net, DeepLabV3, TransAttUnet, and TransUNet. These models are chosen as they are flexible for accepting either 2D slices or 3D volumetric MRI for flexibility. Second, the model chosen has relatively low computation and requirements due to having a smaller number of parameters, while showing an acceptable result in a task in a similar field. Besides, due to the automated requirements of the model, models that require human prompts, like SAM, are not selected for the initial experiment purposes.

Stage 3 Evaluation and Analysis - After setting up all the segmentation experiments, all models will be assessed in Stage 3 for monitoring the training result and analysis. Figure 2 shows the diagram of the full workflow.

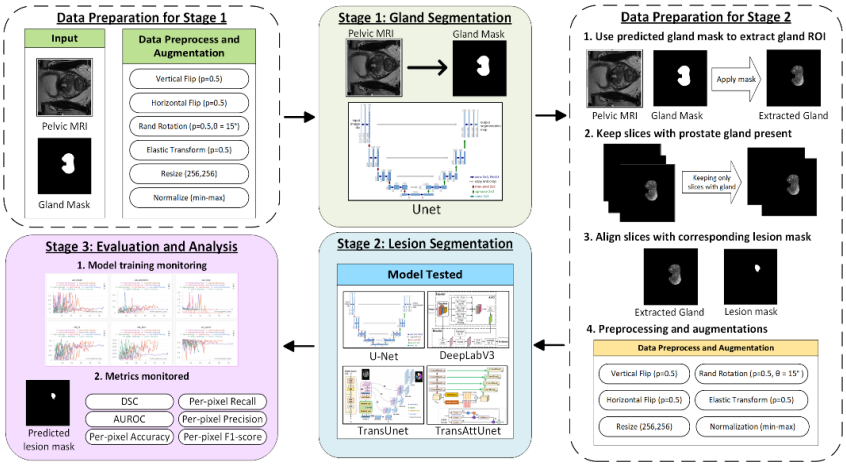

Fig. 2. ROI Segmentation workflow

Besides the experiment for the two-stage ROI segmentation method, another experiment that performed a conventional direct lesion segmentation by using the full MRI, a direct lesion segmentation that does not involve the step of extracting the ROI, was also conducted as the control group in this study.

3.3 Experiment Settings

Three different loss functions are combined into a single loss function for training with their respective weights. The first is the Binary Cross Entropy Loss to monitor the per-pixel accuracy of the model. The second loss function used is the Dice Loss function, which is used to measure the degree of overlap between the prediction and the ground truth mask. The third loss function used is the Focal Loss to address the issue of class imbalance by assigning a larger weight to the hard-to-classify instance/pixel in the data, since the lesion is relatively small when compared to the whole image. This combined loss function, equation (4), will be used for the final model training.

$$L_{bce} = -\sum_{i=1}^{N}[y_i \log \hat{y}_i + (1 - y_i) \log(1 - \hat{y}_i)] \quad (1)$$

$$L_{dice} = 1 - \frac{2\sum_{i=1}^{N} y_i \hat{y}_i + \varepsilon}{\sum_{i=1}^{N} y_i + \sum_{i=1}^{N} \hat{y}_i + \varepsilon} \quad (2)$$

$$L_{focal} = -\frac{1}{N}\sum_{i=1}^{N}[\alpha y_i (1 - \hat{y}_i)^{\gamma} \log \hat{y}_i + \hat{y}_i^{\gamma}(1 - \alpha)(1 - y_i) \log(1 - \hat{y}_i)] \quad (3)$$

$$L = w_1 L_{bce} + w_2 L_{dice} + w_3 L_{focal} \quad (4)$$

Where,

N = number of observations/pixels
y_i = ground truth label of the observation, $y_i \in \{0,1\}$
$\hat{y}_i$ = predicted label of the observation, $\hat{y}_i \in \{0,1\}$
ε = smoothing constant, 1e-6
α = weight for class balance, $\alpha \in [0,1]$
γ = focusing parameter, $\gamma \in [3,5]$
w_1, w_2, w_3 = Corresponding weights for each loss function, $w_{i=1,2,3} \in [0,1]$

Considering the major aim of the project is a segmentation task, the main performance metric to be monitored is the DSC to measure the similarity between two sets. It can be obtained numerically by taking twice the intersection area divided by the sum of each individual area. The DSC values range from 0 (no similarity) to 1 (completely identical). Other metrics like accuracy, recall, precision, F1-score, and AUROC are also monitored to record the per-pixel prediction performance of the model.

In both experiments, all the lesion segmentation models are trained using the NVIDIA T4 GPU 16 GB GDDR6. The optimiser used is the Adam optimiser [25], initialized with a learning rate of 1×10^{-4} and weight decay of 1×10^{-5} to prevent the model from overshooting convergence while ensuring regularization. A ReduceLROnPlateau learning rate scheduler was implemented such that it will reduce the learning rate whenever the validation DSC stops showing improvement for 5 epochs until it reaches the minimum learning rate of 1×10^{-6}. The factor of learning rate reduction is set to 0.8, so that the learning rate is reduced in a less aggressive manner. All models are trained with a max epoch of 200, with a batch size of 64 for each step to leverage the full GPU memory of 16 GB. An EarlyStopping callback was implemented to monitor the validation DSC, which will stop the training progress whenever the model stops improving for over 20 epochs to preventoverfitting.

4 Results and Discussion

For ROI segmentation, a basic U-net of five encoding layers was trained as the gland segmentation model. This gland segmentation model achieved a DSC of 0.9101 on the test set, proving its capabilities to act as the ROI extractor. The control experiment set was also conducted with the same training configuration as stated in the previous section. Table 3 summarises the detailed results of the experiment.

Table 3. Performance Metrics for the experiment

Model	Mode	DSC	Precision	Recall	F1	AUROC	Epoch
U-net	ROI	0.1208	0.1166	0.3733	0.1777	0.9322	79
DeepLabV3	ROI	0.1630	0.1862	0.2357	0.2081	0.8487	87
TransUnet	**ROI**	**0.1826**	**0.2534**	**0.3952**	**0.3003**	**0.9861**	**93**
TransAttUnet	ROI	0.1076	0.1034	0.3191	0.1562	0.9832	73
U-net	Direct	0.0588	0.1017	0.2735	0.1778	0.8539	103
DeepLabV3	Direct	0.0627	0.108	0.427	0.1085	0.927	104
TransUnet	**Direct**	**0.0910**	**0.1264**	**0.2629**	**0.1701**	**0.8964**	**94**
TransAttUnet	Direct	0.0640	0.0441	0.3901	0.0787	0.8394	88

Table 3 indicates that, in the ROI lesion segmentation, TransUNet achieved the highest overall performance with a DSC of 0.1826, followed by DeepLabV3, U-net, and finally TransAttUnet of DSC 0.1630, 0.1208, and 0.1076, respectively. For the direct lesion segmentation, TransUnet achieved the highest overall performance with a DSC of 0.091, followed by TransAttUnet, DeepLabV3, and finally U-net of DSC 0.064, 0.0627, and 0.0059, respectively.

TransUNet produced the best performance in both experiment sets, acquiring a DSC of 0.0910 and 0.1826 in direct lesion segmentation and ROI lesion segmentation, respectively. It consistently outperformed other models, highlighting the benefit of its hybrid encoder, which combines CNN's local feature extraction with the transformer's capacity for global context modelling [10]. Figure 3 shows the predictions of the best-performing model (TransUNet) for each experiment set.

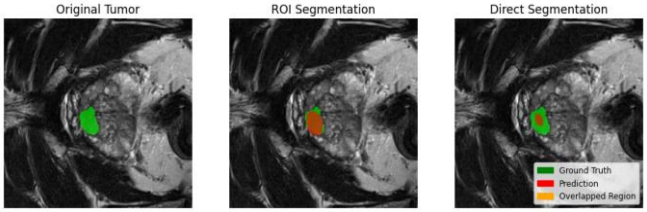

Fig. 3. Ground truth (left), predictions ROI lesion segmentation (middle) and direct lesion segmentation (right)

However, the best-performing DSC of 0.1826 is still much lower than the original Prostate158's baseline of 0.453 [26], although the baseline was trained using the 3D volumetrics as the input data, which might leverage the interrelated features between

slices. Besides that, due to the focus of the study is the comparative analysis between two segmentation methods, hyperparameter tuning is not performed. Each model was initialized with the same hyperparameters, which may not be their optimal hyperparameter settings. It is also important to note that the DSC in this study was computed on a per-batch average basis, which inherently excludes slices without any lesions. This setting would omit the cases that would otherwise achieve a perfect DSC of 1.0, potentially lowering the overall DSC.

From both experiment sets, it can be observed that the ROI lesion segmentation would improve the performance of all models under the same experiment configuration and hyperparameter settings. Figure 4 shows the graphical comparison of performance metrics between the two approaches.

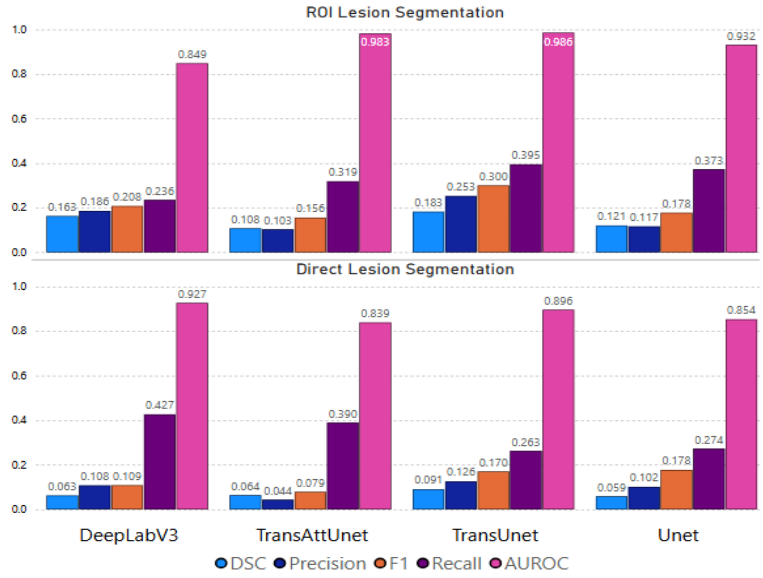

Fig. 4. Result for ROI lesion segmentation (top) and direct lesion segmentation (bottom)

All models show a positive improvement in percentage change in DSC of the ROI segmentation relative to direct lesion segmentation, with DeepLabV3 showing the best improvement of 160%. These results suggest that including anatomical context through prior gland segmentation allows lesion segmentation models to perform better in segmenting prostate lesions. The percentage change of each model is shown in Table 4.

Table 4. Percentage change of DSC in ROI lesion segmentation relative to direct lesion segmentation

Model	Percentage Increased
DeepLabV3	**160.0%**
Unet	105.4%
TransUnet	100.6%
TransAttUnet	68.1%

From the result, it can be inferred that the prior gland segmentation with the extraction of ROI may provide anatomical prior knowledge to the later lesion segmentation model by limiting its focus on relevant regions [17]. By limiting the input image to the prostate gland only, this approach ensures that the lesion segmentation model only needs to focus on adjusting its parameters to learn the feature in the relevant prostate gland region. It could also reduce the complexity of training and eliminate the possibility of the model learning irrelevant features (noise) outside the prostate gland [18].

Despite the higher observed performance, there is a potential error propagation that might exist in this technique. In this ROI segmentation, there are two models used, where the lesion segmentation model is highly dependent on the gland segmentation model. The errors or noise learned from the gland segmentation stage might propagate to the later lesion segmentation model, which may cause the subsequent degradation of performance in lesion segmentation. This error propagation is similar to the issue faced in [20], where the slices that are misclassified as not containing the prostate will yield no lesion mask output in the segmentation model, severely affecting the quality of the diagnosis. Moreover, this workflow requires the training of two DL models, which may get too computationally intensive if dealing with a large dataset or a more complex model with more parameters, such as the Vision Language Model (VLM). However, if the clinical workflow itself requires the automated DL-based gland segmentation in the pipeline or there is already a robust and production-ready gland segmentation model, then this two-step approach may not introduce any redundant burdens to the workflow.

5 Conclusion

This study aimed to compare a two-step deep learning ROI lesion segmentation approach with a single-step direct lesion segmentation approach under the same configurations. The result shows that the two-step ROI lesion segmentation demonstrated higher performance relative to the direct lesion segmentation approach with a percentage increase of 160.0%, 105.4%, 100.6%, and 68.1% in the models DeepLabV3, Unet, TransUNet, and TransAttUnet, respectively. This two-step approach produced an overall best-performing model of TransUNet with a DSC of 0.1826. Despite this, there are still potential issues, including error propagation and dependencies in performance between two models, while requiring a higher computational requirement. While the experiment has proven to favour the two-step approach, the experiment was conducted using Prostate158, which still possesses an issue with generalizability. While valuable, Prostate158 consists of a relatively homogeneous set of prostate MRI images. More experiments are needed to examine the approach in generalising data from different scanners, modalities, or populations with greater anatomical and pathological variability. For future research, this study can be extended by incorporating other well-curated prostate MRI datasets like the PICAI Challenge to investigate the approach's generalizability or using 3D volumetric MRI images for training to investigate its performance with 3D data and its abilities to learn features between adjacent slices.

References

1. Liechti, M.R., Muehlematter, U.J., Schneider, A.F., Eberli, D., Rupp, N.J., Hötker, A.M., Donati, O.F., Becker, A.S.: Manual prostate cancer segmentation in MRI: interreader agreement and volumetric correlation with transperineal template core needle biopsy. Eur Radiol. 30, 4806–4815 (2020). https://doi.org/10.1007/s00330-020-06786-w.
2. Niaf, E., Rouvière, O., Mège-Lechevallier, F., Bratan, F., Lartizien, C.: Computer-aided diagnosis of prostate cancer in the peripheral zone using multiparametric MRI. Phys. Med. Biol. 57, 3833–3851 (2012). https://doi.org/10.1088/0031-9155/57/12/3833.
3. Xiao, H., Li, L., Liu, Q., Zhu, X., Zhang, Q.: Transformers in medical image segmentation: A review. Biomedical Signal Processing and Control. 84, 104791 (2023). https://doi.org/10.1016/j.bspc.2023.104791.
4. Wei, C., Liu, Z., Zhang, Y., Fan, L.: Enhancing prostate cancer segmentation in bpMRI: Integrating zonal awareness into attention-guided U-Net. DIGITAL HEALTH. 11, 20552076251314546 (2025). https://doi.org/10.1177/20552076251314546.
5. Chen, Y., Ge, P., Wang, G., Weng, G., Chen, H.: An overview of intelligent image segmentation using active contour models. Intell Robot. 3, 23–55 (2023). https://doi.org/10.20517/ir.2023.02.
6. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W.M., and Frangi, A.F. (eds.) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. pp. 234–241. Springer International Publishing, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28.
7. Chen, L.-C., Papandreou, G., Schroff, F., Adam, H.: Rethinking Atrous Convolution for Semantic Image Segmentation, https://arxiv.org/abs/1706.05587, (2017). https://doi.org/10.48550/ARXIV.1706.05587.
8. Khan, A., Rauf, Z., Khan, A.R., Rathore, S., Khan, S.H., Shah, N.S., Farooq, U., Asif, H., Asif, A., Zahoora, U., Khalil, R.U., Qamar, S., Asif, U.H., Khan, F.B., Majid, A., Gwak, J.: A Recent Survey of Vision Transformers for Medical Image Segmentation, https://arxiv.org/abs/2312.00634, (2023). https://doi.org/10.48550/ARXIV.2312.00634.
9. Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H.R., Xu, D.: Swin UNETR: Swin Transformers for Semantic Segmentation of Brain Tumors in MRI Images. In: Crimi, A. and Bakas, S. (eds.) Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. pp. 272–284. Springer International Publishing, Cham (2022). https://doi.org/10.1007/978-3-031-08999-2_22.
10. Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A.L., Zhou, Y.: TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation, http://arxiv.org/abs/2102.04306, (2021). https://doi.org/10.48550/arXiv.2102.04306.
11. Chen, B., Liu, Y., Zhang, Z., Lu, G., Kong, A.W.K.: TransAttUnet: Multi-level Attention-guided U-Net with Transformer for Medical Image Segmentation, https://arxiv.org/abs/2107.05274, (2021). https://doi.org/10.48550/ARXIV.2107.05274.
12. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.-Y., Dollár, P., Girshick, R.: Segment Anything, https://arxiv.org/abs/2304.02643, (2023). https://doi.org/10.48550/ARXIV.2304.02643.

13. Ma, J., He, Y., Li, F., Han, L., You, C., Wang, B.: Segment Anything in Medical Images. (2023). https://doi.org/10.48550/ARXIV.2304.12306.
14. Hambarde, P., Talbar, S., Mahajan, A., Chavan, S., Thakur, M., Sable, N.: Prostate lesion segmentation in MR images using radiomics based deeply supervised U-Net. Biocybernetics and Biomedical Engineering. 40, 1421–1435 (2020). https://doi.org/10.1016/j.bbe.2020.07.011.
15. Nobnop, N., Yamcharoen, N., Sukjamsri, C., Piboonthummasak, T., Charoenpong, T., Kiatisevi, P.: Pelvic Tumor Segmentation in MRI Images Using Deep Learning with DeepLabV3+ and U-Net: A Performance Comparison. In: 2024 16th Biomedical Engineering International Conference (BMEiCON). pp. 1–5. IEEE, Chon Buri, Thailand (2024). https://doi.org/10.1109/BMEiCON64021.2024.10896343.
16. Molstad, N., Myhre, E.: Segmentation of Clinically Significant Prostate Cancer using Swin UNETR, (2024).
17. Liu, Z., Jiang, W., Lee, K.-H., Lo, Y.-L., Ng, Y.-L., Dou, Q., Vardhanabhuti, V., Kwok, K.-W.: A Two-Stage Approach for Automated Prostate Lesion Detection and Classification with Mask R-CNN and Weakly Supervised Deep Neural Network. In: Nguyen, D., Xing, L., and Jiang, S. (eds.) Artificial Intelligence in Radiation Therapy. pp. 43–51. Springer International Publishing, Cham (2019). https://doi.org/10.1007/978-3-030-32486-5_6.
18. Zhang, G., Jia, M., Gao, L., Luo, J., Zhang, A., Chen, Y., Shan, P., Zhao, B.: Saliency Transfer Learning and Central-Cropping Network for Prostate Cancer Classification. Neural Process Lett. 55, 2391–2403 (2023). https://doi.org/10.1007/s11063-022-10999-z.
19. Venâncio, L.M.T.F.: Prostate Lesion Segmentation with Convolutional Neural Networks, (2020).
20. Wang, Z., Wu, R., Xu, Y., Liu, Y., Chai, R., Ma, H.: A two-stage CNN method for MRI image segmentation of prostate with lesion. Biomedical Signal Processing and Control. 82, 104610 (2023). https://doi.org/10.1016/j.bspc.2023.104610.
21. Eidex, Z.A., Wang, T., Lei, Y., Axente, M., Akin-Akintayo, O.O., Ojo, O.A.A., Akintayo, A.A., Roper, J., Bradley, J.D., Liu, T., Schuster, D.M., Yang, X.: MRI-based prostate and dominant lesion segmentation using cascaded scoring convolutional neural network. Medical Physics. 49, 5216–5224 (2022). https://doi.org/10.1002/mp.15687.
22. Adams, L.C., Makowski, M.R., Engel, G., Rattunde, M., Busch, F., Asbach, P., Niehues, S.M., Vinayahalingam, S., Van Ginneken, B., Litjens, G., Bressem, K.K.: Prostate158 - An expert-annotated 3T MRI dataset and algorithm for prostate cancer detection. Computers in Biology and Medicine. 148, 105817 (2022). https://doi.org/10.1016/j.compbiomed.2022.105817.
23. Keno Bressem, Lisa Adams, Günther Engel: Prostate158 - Training data, https://zenodo.org/doi/10.5281/zenodo.6481141, (2022). https://doi.org/10.5281/ZENODO.6481141.
24. Bressem, K.: Prostate158 - test data, https://zenodo.org/doi/10.5281/zenodo.6592345, (2022). https://doi.org/10.5281/ZENODO.6592345.
25. Kingma, D.P., Ba, J.: Adam: A Method for Stochastic Optimization, https://arxiv.org/abs/1412.6980, (2014). https://doi.org/10.48550/ARXIV.1412.6980.
26. Bressem, K.: Models for Prostate158, https://zenodo.org/record/6397057, (2022). https://doi.org/10.5281/ZENODO.6397057.