EUROPEAN UROLOGY OPEN SCIENCE 62 (2024) 47–53

available at www.sciencedirect.com
journal homepage: www.eu-openscience.europeanurology.com

European Association of Urology

Prostate Cancer

Deep Learning Model for Real time Semantic Segmentation During Intraoperative Robotic Prostatectomy

Sung Gon Park [a], *Jeonghyun Park* [b], *Hong Rock Choi* [b], *Jun Ho Lee* [b], *Sung Tae Cho* [a], *Young Goo Lee* [a], *Hanjong Ahn* [c], *Sahyun Pak* [a,*]

[a] *Department of Urology, Hallym University College of Medicine, Kangnam Sacred Heart Hospital, Seoul, South Korea;* [b] *STARLABS Corp., Seoul, South Korea;* [c] *Department of Urology, University of Ulsan College of Medicine, Asan Medical Center, Seoul, South Korea*

Article info

Article history:
Accepted February 11, 2024

Associate Editor:
Roderick van den Bergh

Keywords:
Prostatectomy
Artificial intelligence
Deep learning
Segmentation

Abstract

Background and objective: Recently, deep learning algorithms, including convolutional neural networks (CNNs), have shown remarkable progress in medical imaging analysis. Semantic segmentation, which segments an unknown image into different parts and objects, has potential applications in robotic surgery in areas where artificial intelligence (AI) can be applied, such as in AI-assisted surgery, surgeon training, and skill assessment. We aimed to investigate the performance of a CNN-based deep learning model in real-time segmentation in robot-assisted radical prostatectomy (RALP).
Methods: Intraoperative videos of RALP procedures were obtained. The reinforcement U-Net model was used for segmentation. Segmentation of the images of instruments, bladder, prostate, and seminal vesicle–vas deferens was performed. The dataset was preprocessed and split randomly into training, validation, and test data in a 7:2:1 ratio. Dice coefficient, intersection over union (IoU), and accuracy by class, which are commonly used in medical image segmentation, were calculated to evaluate the performance of the model.
Key findings and limitations: From 120 patient videos, 2400 images were selected for RALP procedures. The mean Dice scores for the identification of the instruments, bladder, prostate, and seminal vesicle–vas deferens were 0.96, 0.74, 0.85, and 0.84, respectively. Overall, when applied to the test data, the model had a mean Dice coefficient value of 0.85, IoU of 0.77, and accuracy of 0.85. Limitations included the sample size, lack of diversity in the methods of surgery, incomplete surgical processes, and lack of external validation.
Conclusions and clinical implications: The CNN-based segmentation provides accurate real-time recognition of surgical instruments and anatomy in RALP. Deep learning algorithms can be used to identify anatomy within the surgical field and could potentially be used to provide real-time guidance in robotic surgery.

* Corresponding author. Department of Urology, Hallym University College of Medicine, Kangnam Sacred Heart Hospital, 1, Singil-ro, Yeongdeungpo-gu, Seoul 07441, South Korea. Tel. +82 2 829 5198; Fax: +82 2 846 5198.
E-mail address: qkrtkgus0@naver.com (S. Pak).

https://doi.org/10.1016/j.euros.2024.02.005

Patient summary: We demonstrate the potential effectiveness of deep learning segmentation in robotic prostatectomy procedures. Deep learning algorithms could be used to identify anatomical structures within the surgical field and may provide real-time guidance in robotic surgery.

1. Introduction

Prostate cancer is one of the most commonly diagnosed cancers worldwide [1]. Approximately 80–90% of prostate cancers are diagnosed in a nonmetastatic state, and men with nonmetastatic prostate cancer are generally considered to have favorable survival outcomes [2]. While several treatment approaches for nonmetastatic prostate cancer are available, radical prostatectomy is considered a key method [3,4]. Although prostatectomy procedures have traditionally been performed using an open radical retropubic approach, robot-assisted laparoscopy has recently become the procedure of choice, particularly in developed countries [5].

In recent years, artificial intelligence (AI) and deep learning have seen major advances in various fields, including computer vision and natural language processing [6]. In particular, computer vision using deep neural networks has emerged as an important part of the machine learning field [6–8]. Compared with image classification or object recognition, semantic segmentation is one of the most complex tasks in computer vision, and describes the ability to segment an unknown image into different parts and objects [7].

Image segmentation has extensively been studied in the medical domain, especially in applications of computer-aided diagnosis in radiology [9]. In addition, segmentation has also been investigated in vision-based surgical procedures. However, the utility of segmentation in real-time surgery remains largely unknown. This study investigated the performance of segmentation in robotic prostatectomy and its application in intraoperative real-time surgery.

2. Patients and methods

2.1. Study population

We reviewed the records of 150 men with prostate cancer who underwent robot-assisted radical prostatectomy using a Da Vinci robotic platform at two institutions in Hallym University Medical Center between January 2022 and June 2023. Patient data, including demographic characteristics, clinicopathological features, and intraoperative videos, were analyzed retrospectively. The study protocol was approved by the institutional review board of Hallym University Kangnam Sacred Heart Hospital (no. 2022-11-017).

2.2. Dataset

Robot-assisted radical prostatectomy was performed using a traditional anterior approach, and intraoperative videos were recorded in real time. The videos were converted into MP4 video format with 720-pixel resolution and a rate of 30 frames per second. For each surgical procedure, 20–30 images were extracted at regular intervals from the beginning of bladder mobilization to the end of vas deferens and seminal vesicle dissection. All images extracted from each video were annotated manually at a per-pixel level by two expert urologists (Fig. 1). Annotation was performed for each of the classes of surgical instruments, bladder, prostate, and vas deferens/seminal vesicle. Inter-rater agreement between the two urologists was excellent, with a Dice coefficient of 0.85. The dataset was preprocessed and split randomly into training, validation, and test data in a 7:2:1 ratio. After developing the model, validation was performed by three specialists from different external institutions.

2.3. Network architecture

For this research, we utilized the reinforcement U-Net model provided by STARLABS Co. The traditional U-Net architecture is mainly composed of encoders and decoders. The encoder extracts the context of the image, while the decoder reconstructs the refined features based on the encoder extraction.

The reinforcement U-Net model was designed by adding a six-layer convolution-batch normalization-ReLU operation block to both the encoder and the decoder section of the traditional U-Net structure (Fig. 2). This design increases the depth and width of the segmented feature map, intensifying the feature point extraction in the boundary regions of complex and nonuniform objects (Supplementary Fig. 1). Compared with the traditional U-Net model, this model possesses more layers and deeper feature extraction capability, and is therefore anticipated to respond more sensitively to intricate structures and minor changes.

We believe that these characteristics will help the model reduce the frequency of outliers at the object boundary. This in turn will facilitate more accurate class categorization and tissue dissection surface detection. The accurate detection of minor changes and boundaries is crucial, particularly in laparoscopic surgeries, and the features of the reinforcement U-Net model are anticipated to offer significant advantages in such applications.

2.4. Training

Before proceeding with training, preprocessing steps were applied to the dataset. The purpose of preprocessing is to improve image quality for a more effective analysis and use of computer resources. Descriptions of the methods

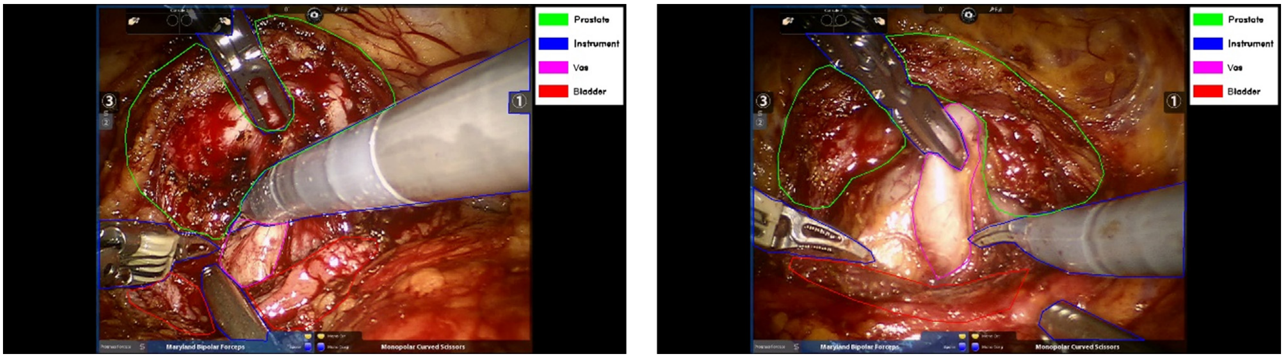

Fig. 1 – Prostatectomy image annotation.

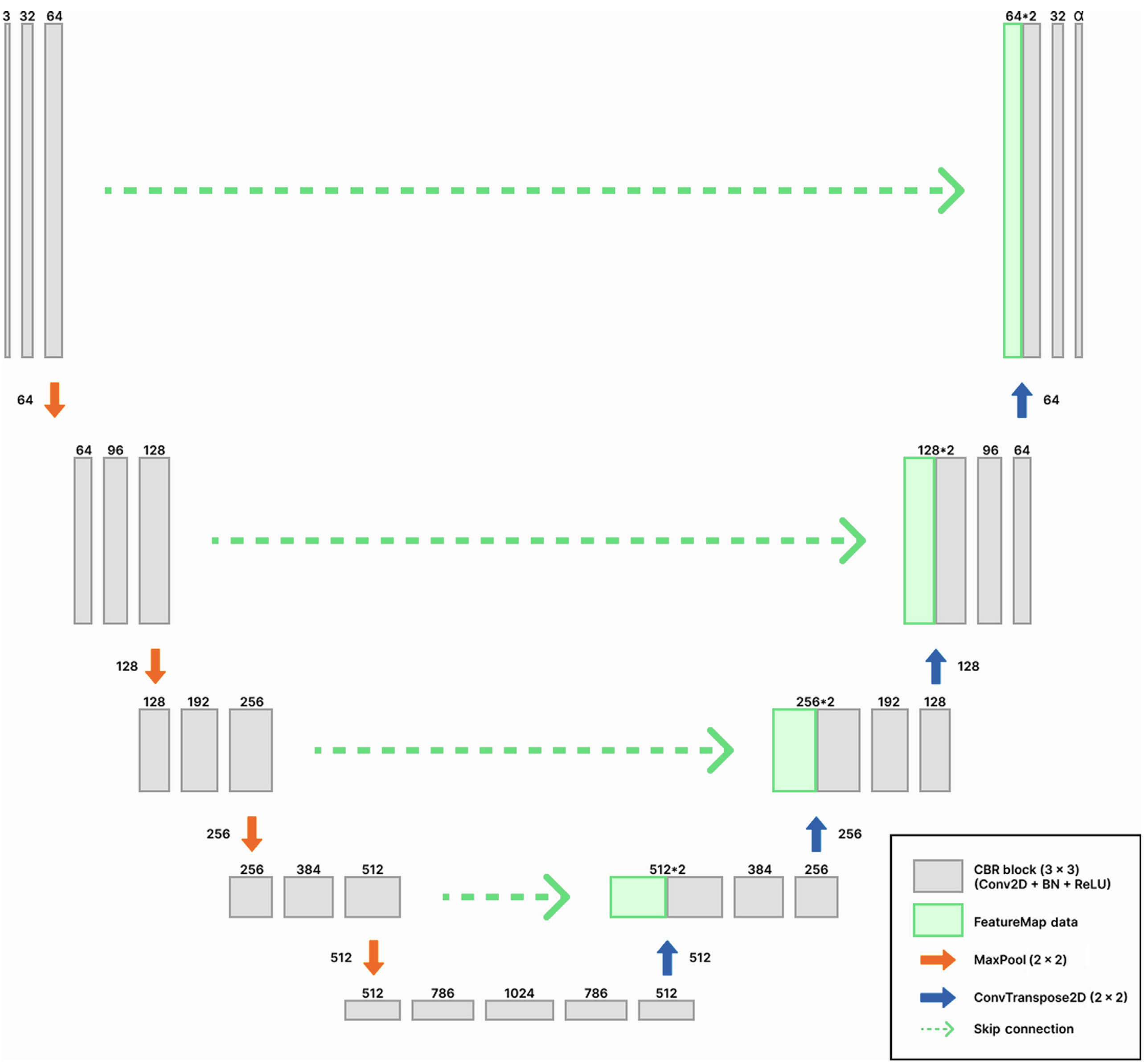

Fig. 2 – Reinforcement U-Net architecture.

used for preprocessing and dynamic transform are provided in the Supplementary material.

Optimizers are algorithms used to change the attributes of the model, such as weights and learning rate, so that the loss function can be minimized. The Adam optimization algorithm was adopted during the training process, with the learning rate set at 0.0001. Focal loss was used as the loss function, and to generalize the model, the batch size was set to 16. The entire training was carried out over 100 epochs. Although an early stop feature was implemented to halt training if there was no improvement in validation loss over a certain period, this feature was not activated in this research. Various data transformation techniques were applied to the training data based on validation loss through dynamic transformation techniques, ensuring diversity in the input data.

Training and validation losses were recorded at every epoch, and the progress of the training was monitored through visualization. The model's weights were saved every time there was an improvement in the validation loss, and intermediate model weights were also saved separately every ten epochs.

2.5. Evaluation metrics

The Dice coefficient, intersection over union (IoU), and accuracy were calculated for each image to evaluate the performance of the developed convolutional neural network (CNN) model in the segmentation task. A confusion matrix method was used to measure the performance of the segmentation model. The Dice coefficient and IoU are commonly used as evaluation metrics in similar medical image segmentation tasks. The Dice coefficient and IoU are similar in that both metrics deal with overlap between predicted segmentation and ground truth, but are different in that the methods used to calculate these are distinct. The Dice coefficient is calculated by 2 $\times$ intersection divided by the total number of pixels in both images, whereas IoU refers to the ratio of the area of the intersection over the area of union of the predicted segmentation and the ground truth. The following formulas were used for these metrics:

$$\text{Dice coefficient} = (2 \times \text{TP})/[(\text{TP} + \text{FN}) + (\text{TP} + \text{FP})]$$

$$\text{IoU} = \text{TP}/(\text{TP} + \text{FN} + \text{FP})$$

$$\text{Accuracy} = (\text{TP} + \text{TN})/(\text{TP} + \text{FN} + \text{FP} + \text{TN})$$

where, TP = true positive, FP = false positive, TN = true negative, and FN = false negative.

3. Results

The clinical and pathological characteristics of the enrolled patients are summarized in Table 1. Of the 120 prostate cancer patients who underwent robot-assisted radical prostatectomy, the mean age was 66.4 yr, the mean prostate-specific antigen level was 11.7 ng/ml, and the mean prostate volume was 40.3 ml. Fifty-one (42.5%) patients had locally advanced prostate cancer (pathological T stage 3a or 3b) and 14 (11.7%) had high-grade (Gleason score $\geq$8) disease.

Our dataset consisted of 2400 ground truth images from 120 videos. After preprocessing, all images were split into training, validation, and test sets in a ratio of 7:2:1. A total of 1680 images were used for training, 480 for validation, and 240 for test data.

Training and validation loss curves for the Dice coefficient, IoU score, and Dice loss are shown in Figure 3. With each training epoch, the training and validation curves for both Dice coefficient and IoU increased, while the loss curves decreased. After 60 epochs of training, the validation loss decreased at a slower rate and plateaued.

Table 1 – Baseline demographic and clinical characteristics of the study patients[a]

	Total no. of patients (n = 120)
Age (yr), mean	66.4
Body mass index (kg/m^2), mean	25.0
Prostate-specific antigen (ng/ml), mean	11.7
Prostate volume (cc)	40.3
Pathological T stage	
T2	69 (57.5)
T3a	31 (25.8)
$\geq$T3b	20 (16.7)
Pathological Gleason score	
6	14 (11.7)
3 + 4	44 (36.7)
4 + 3	48 (40.0)
8	4 (3.3)
9–10	10 (8.3)
Positive surgical margin	31 (25.8)

[a] Data are presented as n (%) unless otherwise indicated.

Overall, our model showed a Dice coefficient of 85% and an IoU score of 77%. Examples of segmentation results obtained from the CNN architecture of our model are shown in Figure 4. The mean Dice scores for the identification of the classes of surgical instruments, bladder, prostate, and seminal vesicle–vas deferens were 0.96, 0.74, 0.85, and 0.84, respectively. These results indicate that our model made fairly accurate predictions of the segmentation mask. In addition, we found that our model worked well on real-time surgical video (Supplementary material, Video 1).

4. Discussion

Recent years have seen a surge of interest in AI and deep learning in various industries. Deep learning, a subset of machine learning, aims to develop algorithms to extract higher-level features from training data and apply them to unknown data [10]. The medical domain is no exception to the application of AI, and deep learning is expected to reshape the landscape of various fields of medicine [11,12]. Currently, the most successful implementations of deep learning in medicine have been for medical image analysis tasks [12].

In deep learning for computer vision, semantic segmentation is an important area of research field [13]. As deep learning algorithms have been developed, numerous neural networks for semantic segmentation have been introduced, and as such, the "state-of-the-art" model is continuously being replaced. In the present study, the reinforcement U-Net architecture was used for surgical image segmentation. U-Net was proposed on the basis of FCN [14] and has seen wide use in medical imaging. U-Net consists of an encoder-decoder structure: the encoder continuously merges layers to extract features to gradually reduce the spatial dimension, and the decoder gradually re-establishes the spatial dimension and target detail accord-

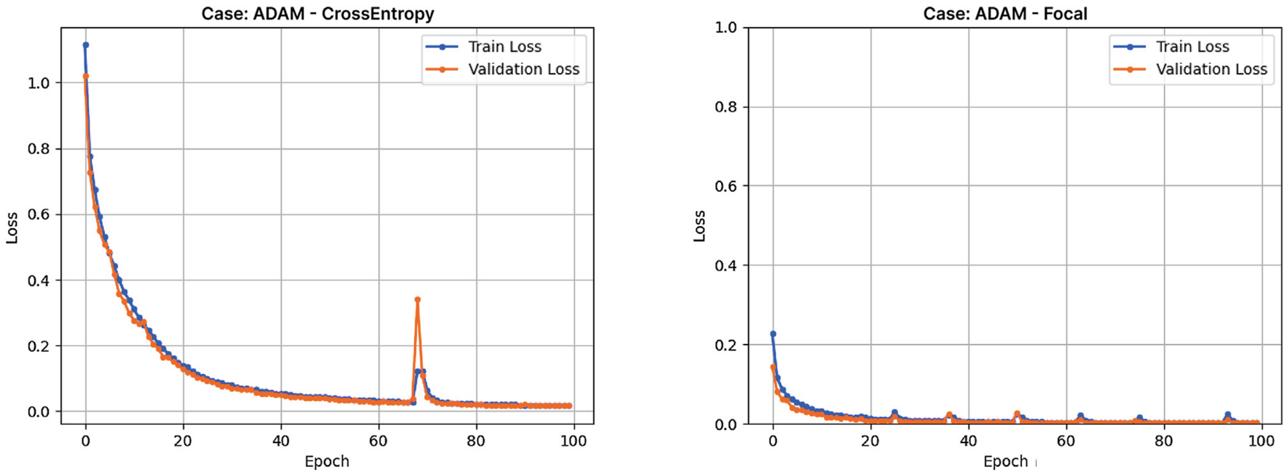

Fig. 3 – Training and validation loss curves.

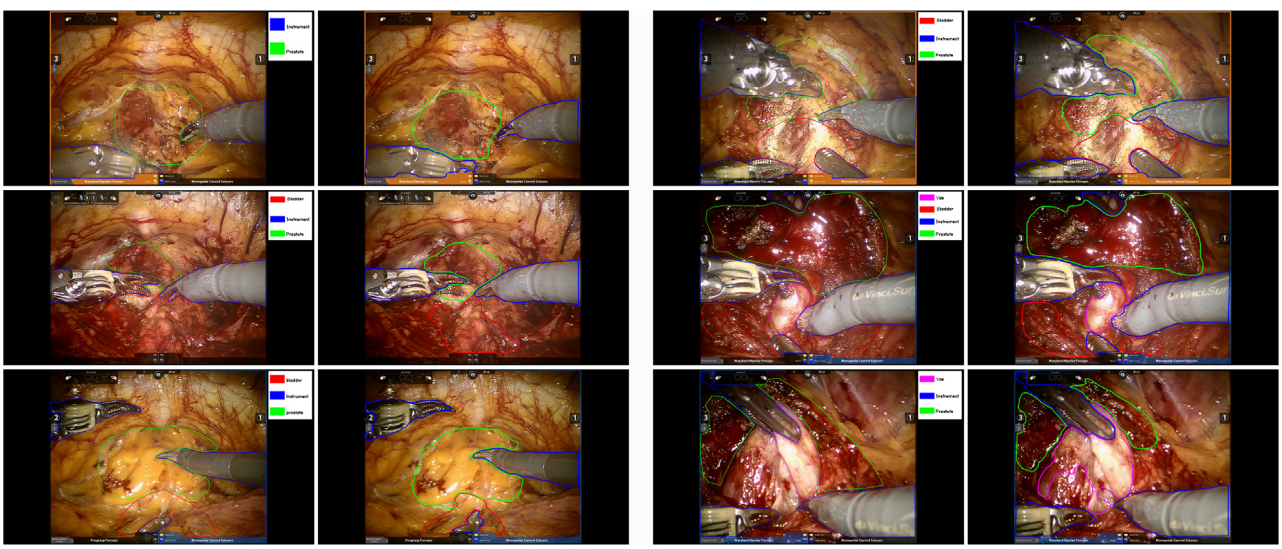

Fig. 4 – Examples of segmentation results.

ing to the extracted features [14]. Differing from conventional CNNs, U-Net uses a skip connection strategy, which directly utilizes shallow features, to use the features of the downsampling part of the encoder to be utilized for upsampling. To achieve a more refined reduction, this strategy is applied to shallow feature information at all scales. We used the modified U-Net architecture in this study for two reasons. First, until recently, U-Net was the most popular model in medical image segmentation. Despite the rapid development of transformer in many vision fields, U-Net has shown high performance and applicability in much of the research [15]. Second, we used some of the data from the present study to test several CNN and transformer models in the pilot phase, including U-Net, DeepLabv3 [16], MANet [17], SegFormer [18], BEiT [19], and DPT [20]. The test results indicated that the models showed similar performance. In addition, the CNN models were appropriate from the perspective of real-time application during surgery.

Computer vision can be used for various purposes in the field of surgical procedures, including endoscopy, laparoscopy, surgical microscopy, fluoroscopy, and multispectral imaging [21,22]. In the present study, we investigated the application of semantic segmentation in robot-assisted laparoscopic surgery. Robot-assisted laparoscopic prostatectomy is the main treatment for nonmetastatic prostate cancer; however, it has a steep learning curve [23], and the boundaries between the bladder and prostate, posterior fascia, and seminal vesicle are often difficult to distinguish intraoperatively because of complex anatomy [24,25]. In robotic surgery, semantic segmentation could possibly play important roles in the assessment of surgical technical skill, detection of anatomy, and navigation [22].

In the present study, we demonstrate the potential effectiveness of a segmentation network in robotic prostatectomy procedures. We used the Dice coefficient and IoU as evaluation metrics. In general, the Dice coefficient and IoU are the most commonly used metrics for medical image segmentation analysis, compared with accuracy, because both metrics penalize false positives, which occur frequently in medical image datasets due to class imbalance. Although there are differences in complexity, most previous surgical image segmentation papers considered a Dice coefficient or IoU score of $\geq$0.7 to indicate good performance. We found that the segmentation performed well, not only when applied to still images, but also when applied to real-time video. There have been several studies on the application of segmentation in vision-based surgery [21,26], but most studies were limited to recognizing surgical instruments and were used in relatively simple surgeries such as cholecystectomy [21,26,27]. Takeshita et al [23] reported that a CNN model provided accurate recognition of the seminal vesicle–vas deferens in a small dataset of 26 prostate cancer patients who underwent robotic prostatectomy. They reported a Dice similarity coefficient of 0.73 for the test data, consistent with this study's finding of 0.79.

There are several limitations to this study. First, the sample size was relatively small. Considering the characteristics of CNN models, training on larger datasets will result in better performance. Second, this study used data from anterior approach prostatectomy performed by four surgeons at two institutions. Therefore, whether this study's findings can be applied to other methods of surgery, such as the Retzius-sparing approach, was not investigated. Third, we focused on real-time segmentation, which is necessary for AI-assisted or AI-enhanced surgery. We observed that our model worked well in all sections and parts of the surgery included in the analysis. However, when the vision was blurred because it was out of focus, or severe bleeding or unusual prostate morphology was present, the performance of the CNN model was found to decrease temporarily. Fourth, vesicourethral anastomosis procedures were excluded from the training data because of complex anatomy. Segmentation of the urethra, which is the most important part of prostatectomy anatomy, was not investigated in this study and needs to be explored in follow-up studies with larger datasets. Lastly, considering that an image segmentation analysis in the surgical domain requires more precision than in other domains, our results suggest that segmentation of organs requires further improvement. Our results indicate that a segmentation analysis of the images of the bladder is poorer than those of other structures, despite good overall performance. Poorer bladder image segmentation may be due to the anatomical characteristics of the bladder. The bladder is positioned adjacent to the prostate and fat-containing soft tissue, which can make it difficult to clearly distinguish the boundary. To overcome this, a larger number of sophisticated annotated training datasets and better AI models may be required.

Despite these limitations, we believe that this study provides a valuable contribution because it is the first to investigate the potential effectiveness of real-time semantic segmentation for instruments and anatomical organs in robotic prostatectomy, including the prostate and bladder. Our results suggest that deep learning algorithms can be used to identify anatomy within the surgical field and may be used to provide real-time guidance in robotic surgery. This technology could be used in AI-assisted robotic surgery, such as on solo surgery platforms with an AI assistant, or in autonomous surgery in the future.

5. Conclusions

The CNN model based on a modified U-Net architecture provided accurate real-time recognition of the surgical instruments and anatomical structures observed during radical prostatectomy. Deep learning algorithms can be used to identify anatomy within the surgical field and may be used to provide real-time guidance in robotic surgery. Further large-scale studies are warranted.

Author contributions: Sahyun Pak had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: Pak.
Acquisition of data: S.G. Park, Cho, Y.G. Lee, Ahn, Pak.
Analysis and interpretation of data: J. Park, Pak.
Drafting of the manuscript: S.G. Park, J. Park, Pak.
Critical revision of the manuscript for important intellectual content: Pak.
Statistical analysis: J. Park, Pak.
Obtaining funding: J.H. Lee, Pak.
Administrative, technical, or material support: J. Park, Choi, J.H. Lee.
Supervision: Pak.
Other: None.

Financial disclosures: Sahyun Pak certifies that all conflicts of interest, including specific financial interests and relationships and affiliations relevant to the subject matter or materials discussed in the manuscript (eg, employment/affiliation, grants or funding, consultancies, honoraria, stock ownership or options, expert testimony, royalties, or patents filed, received, or pending), are the following: None.

Funding/Support and role of the sponsor: This work was supported by the Technology development Program (S3288928) funded by the Ministry of SMEs and Hallym University Research Fund (HURF-2023-24).

Data sharing: The dataset used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.euros.2024.02.005.

References

[1] Siegel RL, Miller KD, Wagle NS, Jemal A. Cancer statistics, 2023. CA Cancer J Clin 2023;73:17–48.
[2] Chierigo F, Wenzel M, Würnschimmel C, et al. Survival after radical prostatectomy versus radiation therapy in high-risk and very high-risk prostate cancer. J Urol 2022;207:375–84.
[3] Würnschimmel C, Wenzel M, Wang N, et al. Radical prostatectomy for localized prostate cancer: 20-year oncological outcomes from a German high-volume center. Urol Oncol 2021;39:830.e17–e26.
[4] Obek C, Doganca T, Argun OB, Kural AR. Management of prostate cancer patients during COVID-19 pandemic. Prostate Cancer Prostatic Dis 2020;23:398–406.

[5] Kim SP, Karnes RJ, Gross CP, et al. Contemporary national trends of prostate cancer screening among privately insured men in the United States. Urology 2016;97:111–7.
[6] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521:436–44.
[7] Guo Y, Liu Y, Georgiou T, Lew MS. A review of semantic segmentation using deep neural networks. Int J Multimed Inf Retr 2018;7:87–93.
[8] Chai J, Zeng H, Li A, Ngai EW. Deep learning in computer vision: A critical review of emerging techniques and application scenarios. Mach Learn Appl 2021;6:100134.
[9] Chen X, Wang X, Zhang K, et al. Recent advances and clinical applications of deep learning in medical image analysis. Med Image Anal 2022;79:102444.
[10] Chassagnon G, Vakalopolou M, Paragios N, Revel M-P. Deep learning: definition and perspectives for thoracic imaging. Eur Radiol 2020;30:2021–30.
[11] Le Q, Miralles-Pechuán L, Kulkarni S, Su J, Boydell O. An overview of deep learning in industry. In: Data analytics and AI. Auerbach Publications; 2020. p. 65–98.
[12] Wang F, Casalino LP, Khullar D. Deep learning in medicine—promise, progress, and challenges. JAMA Intern Med 2019;179:293–4.
[13] Cai L, Gao J, Zhao D. A review of the application of deep learning in medical image classification and segmentation. Ann Transl Med 2020;8:713.
[14] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. Springer; 2015. p. 234–41.
[15] Xiao H, Li L, Liu Q, Zhu X, Zhang Q. Transformers in medical image segmentation: a review. Biomed Signal Process Control 2023;84:104791.
[16] Chen L-C, Papandreou G, Schroff F, Adam H. Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:170605587. 2017.
[17] Chen B, Xia M, Qian M, Huang J. MANet: A multi-level aggregation network for semantic segmentation of high-resolution remote sensing images. Int J Remote Sens 2022;43:5874–94.
[18] Xie E, Wang W, Yu Z, Anandkumar A, Alvarez JM, Luo P. SegFormer: simple and efficient design for semantic segmentation with transformers. Adv Neural Inform Process Syst 2021;34:12077–90.
[19] Bao H, Dong L, Piao S, Wei F. Beit: Bert pre-training of image transformers. arXiv preprint arXiv:210608254. 2021.
[20] Ranftl R, Bochkovskiy A, Koltun V. Vision transformers for dense prediction. Computer Vision Foundation; 2021. p. 12179–88.
[21] Bhandari M, Zeffiro T, Reddiboina M. Artificial intelligence and robotic surgery: current perspective and future directions. Curr Opin Urol 2020;30:48–54.
[22] Chadebecq F, Vasconcelos F, Mazomenos E, Stoyanov D. Computer vision in the surgical operating room. Visceral Med 2020;36:456–62.
[23] Takeshita N, Sakamoto S, Kitaguchi D, et al. Deep learning-based seminal vesicle and vas deferens recognition in the posterior approach of robot-assisted radical prostatectomy. Urology 2023;173:98–103.
[24] Bravi CA, Tin A, Vertosick E, et al. The impact of experience on the risk of surgical margins and biochemical recurrence after robot-assisted radical prostatectomy: a learning curve study. J Urol 2019;202:108–13.
[25] Brunckhorst O, Volpe A, van der Poel H, Mottrie A, Ahmed K. Training, simulation, the learning curve, and how to reduce complications in urology. Eur Urol Focus 2016;2:10–8.
[26] Wang Y, Sun Q, Liu Z, Gu L. Visual detection and tracking algorithms for minimally invasive surgical instruments: a comprehensive review of the state-of-the-art. Robot Auton Syst 2022;149:103945.
[27] Kletz S, Schoeffmann K, Benois-Pineau J, Husslein H. Identifying surgical instruments in laparoscopy using deep learning instance segmentation. In: Proceedings of the 2019 International Conference on Content-Based Multimedia Indexing (CBMI); September 4–6, 2019; Dublin, Ireland. New York, NY: IEEE; 2019. p. 1–6.