Ureteroscopy and
Percutaneous Procedures

Open camera or QR reader and
scan code to access this article
and other resources online.

Automated Upper Tract Urothelial Carcinoma Tumor Segmentation During Ureteroscopy Using Computer Vision Techniques

Daiwei Lu, BS, MS,[1] Amy Reed, MD,[2] Natalie Pace, MD,[2] Amy N. Luckenbaugh, MD,[2]
Maximilian Pallauf, MD,[3,4] Nirmish Singla, MD,[3] Ipek Oguz, PhD,[1] and Nicholas Kavoussi, MD[2]

Abstract

Introduction: Endoscopic tumor ablation of upper tract urothelial carcinoma (UTUC) allows for tumor control with the benefit of renal preservation but is impacted by intraoperative visibility. We sought to develop a computer vision model for real-time, automated segmentation of UTUC tumors to augment visualization during treatment.

Materials and Methods: We collected 20 videos of endoscopic treatment of UTUC from two institutions. Frames from each video ($N = 3387$) were extracted and manually annotated to identify tumors and areas of ablated tumor. Three established computer vision models (U-Net, U-Net++, and UNext) were trained using these annotated frames and compared. Eighty percent of the data was used to train the models while 10% was used for both validation and testing. We evaluated the highest performing model for tumor and ablated tissue segmentation using a pixel-based analysis. The model and a video overlay depicting tumor segmentation were further evaluated intraoperatively.

Results: All 20 videos (mean 36 ± 58 seconds) demonstrated tumor identification and 12 depicted areas of ablated tumor. The U-Net model demonstrated the best performance for segmentation of both tumors (area under the receiver operating curve [AUC-ROC] of 0.96) and areas of ablated tumor (AUC-ROC of 0.90). In addition, we implemented a working system to process real-time video feeds and overlay model predictions intraoperatively. The model was able to annotate new videos at 15 frames per second.

Conclusions: Computer vision models demonstrate excellent real-time performance for automated upper tract urothelial tumor segmentation during ureteroscopy.

Keywords: upper tract urothelial carcinoma, computer vision, artificial intelligence, endoscopy, ureteroscopy

Introduction

THE GOLD STANDARD for treating upper tract urothelial carcinoma (UTUC) has historically been nephroureterectomy.[1] However, modern endoscope and laser technologies enable a minimally invasive means of diagnosis and treatment of UTUC.[2] Specifically, advancements in endoscopic visualization and flexible scopes facilitate retrograde intrarenal surgery and laser ablation of tumors. Completely

treating tumors, however, can be challenging due to the limited field and depth of view (10 and 6 mm on average, respectively) of current flexible endoscopes.[3] Furthermore, bleeding and debris during treatment can obscure tumors and prevent complete ablation.[4,5] Further improvements in tumor localization and visualization during ablation could improve operative efficiency and tumor treatment.

Computer vision, a subfield of artificial intelligence that focuses on analysis of visual data, may improve endoscopic

[1]Department of Computer Science, Vanderbilt University School of Engineering, Nashville, Tennessee, USA.
[2]Department of Urology, Vanderbilt University Medical Center, Nashville, Tennessee, USA.
[3]Department of Urology, Johns Hopkins University, Baltimore, Maryland, USA.
[4]Department of Urology, University Hospital Salzburg, Paracelsus Medical University Salzburg, Salzburg, Austria.

visualization of upper tract urothelial tumors during endoscopic surgery. Using a labeled set of endoscopic video frames, deep learning-based computer vision models can be trained to automatically identify and track UTUC tumors during endoscopic surgery. Similar tools have shown feasibility in identifying bladder tumors during surveillance cystoscopy.[6,7] Moreover, we have previously demonstrated the feasibility of such techniques during localization and treatment of kidney stones.[8] In a similar manner, computer vision methods could enhance tumor identification and localization during endoscopic treatment.

In this study, we developed and compared computer vision-based models for automated segmentation of UTUC tumors during endoscopic treatment. We additionally evaluated the highest performing model's ability to segment areas of ablated tumor. The highest performing model was also implemented in the operating room, and we assessed real-time performance speed of an automated video overlay of UTUC tumor.

Materials and Methods

Cohort

After approval from the Institutional Review Board, we recorded endoscopic surgical video from 20 patients during UTUC treatment from August 2022 to May 2023 from two institutions. All patients were 18 years of age or older, and either had suspicion for UTUC tumor based on preoperative CT imaging or had been referred for endoscopic management after initial diagnosis. A flexible ureteroscope (Karl Storz FLEX-XC or FLEX-X2S) was used for each case. Tumor position and size information based on preoperative imaging was manually recorded.

Dataset and model training

We recorded raw ureteroscopy videos of UTUC tumor treatment at standard resolution (1920×1080 pixels) in 22 patients. All videos were edited to specifically highlight both initial tumor identification and inspection of areas of ablated tumor. The portion of the video that depicted tumor identification included complete endoscopic evaluation of the localized tumor. The portion of the video that illustrated areas of ablated tumor included evaluation of the area immediately after completion of laser treatment. All videos were reviewed for quality, extracted at 5 frames per second [FPS], cropped, and then resized to 512×512 pixels. Two videos were excluded for video quality issues. The frames ($n = 3387$) were manually annotated to identify both UTUC tumors and areas of ablated tumor by two endourologists (Cohen's Kappa agreement of 0.87 based on pixel annotation).

The annotated frames comprised a "ground truth" data set for model training and evaluation. We used MakeSense.ai (San Francisco, CA) for analysis of the annotated frames.[9] We trained two independent models for tumor and ablation segmentation, respectively. We randomly selected 80% of the frames ($n = 2969$) for supervised model training and reserved 10% for validation and 10% for testing ($n = 209$ frames each). The validation set was randomly selected to monitor model training performance. Training was a standard iterative pipeline with forward propagation and loss computation, followed by back-propagation with frequent validation checks. Models were independently evaluated for segmentation of both UTUC tumor and areas of ablated tumor.

To evaluate and optimize a computer vision model for segmentation, we compared three state-of-the-art deep neural network architectures: U-Net, U-Net++, and UNeXt.[10–12] UNet and UNet++ were implemented using pretrained segmentation models in Pytorch using ResNet backbones. The UNext model was constructed based on standard implementations.[13] U-Net is a fast and precise network architecture for high-resolution imaging segmentation.[10] U-Net++ expanded upon this architecture by implementing "skip connections," for potentially enhanced feature extraction.[11] UNeXt aimed to improve upon U-Net with the introduction of binary classifiers (i.e., perceptron blocks) instead of convolutional blocks to improve performance and drastically increase computational speed.[12] Specifically, we tuned hyperparameters as follows: 0.001 learning rate, 0.0001 weight decay (with AdamW optimizer), 0.00001 minute learning rate (with CosingAnnealingLR), Binary Cross Entropy Loss Batch size of 16, and 100 epochs for training. Figure 1 shows the workflow for model training and testing.

Evaluation metrics

We evaluated the model performance for segmentation of UTUC tumors and areas of ablated tumor. Each model was trained using binary cross entropy (BCE, evaluation of the dissimilarity between the automatic and manual segmentations) loss and compared via mean dice similarity coefficient (DSC, a spatial overlap index) score.[14,15] We evaluated the average area under the receiver operating curve (AUC-ROC) of the best performing model for both UTUC and ablated tumor segmentation across all annotated frames. We further computed the overall pixel accuracy, sensitivity, and specificity for the best models. All analyses were done using comet.ml (New York, NY).

We further evaluated the model's performance in real-time in the operative room. Predicted annotations were displayed using a heatmap showing pixelwise probabilities of segmented tumors and areas of ablation. We assessed the real-time performance speed of an automated video overlay of UTUC tumor in the operating room. We report our findings following the Standardized Reporting of Machine Learning Applications in Urology framework to enhance reproducibility, comparability, and interpretability of our results.[16] In addition, we have made our models available for public use and validation.[17]
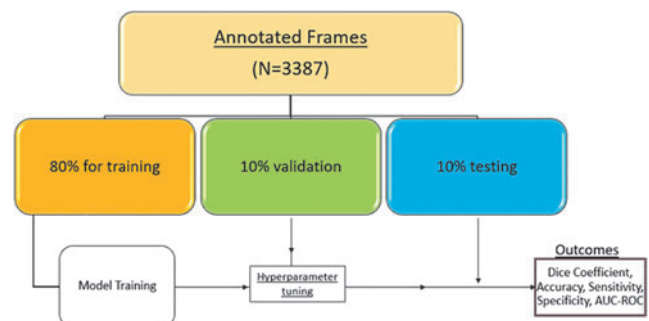

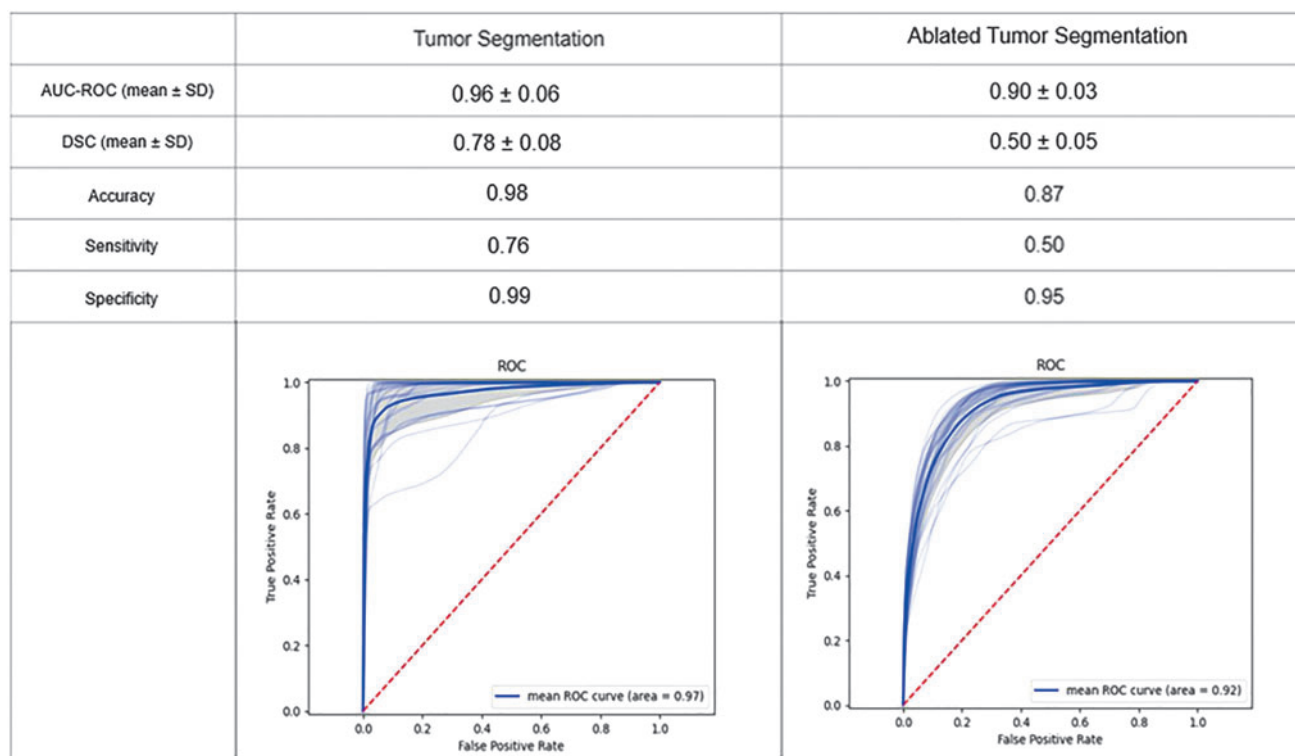**FIG. 1.** Workflow for model training and testing.

TABLE 1. TUMOR CHARACTERISTICS FROM EACH VIDEO

| Video	Location	No. of tumors identified	Histologic grade Low grade = 1 High grade = 2
1	Renal pelvis	1	1
2	Middle pole	2	1
3	Upper pole	3	2
4	Upper pole	2	2
5	Lower pole	2	2
6	Upper pole and renal pelvis	6	1
7	Upper pole	2	2
8	Renal pelvis, ureter	2	2
9	Upper pole	1	1
10	Middle pole	1	1
11	Upper pole, renal pelvis	4	1
12	Interpolar calyx and renal pelvis	4	1
13	Upper, middle, lower pole	5	2
14	Upper pole	1	2
15	Renal pelvis	2	2
16	Middle pole	1	2
17	Lower pole	1	1
18	Renal pelvis	1	1
19	Ureter	1	2
20	Upper pole	1	2

Results

Postoperative biopsy results revealed a mix of low-grade (47%) and high-grade (53%) tumors used during training and analysis. Tumor characteristics are reported in Table 1. Digital and fiber optic ureteroscopes (Karl Storz FLEX-XC or FLEX-X2S) were used in 17 and 3 cases, respectively. Twelve out of the 20 cases underwent treatment with laser ablation. The average surgical video time of tumor identification was 36 seconds (standard deviation ±58 seconds). The average video time depicting the area of tumor ablation was 24 seconds (standard deviation ±16). The U-Net model achieved the highest DSC and lowest BCE loss (mean ± standard deviation of 0.78 ± 0.08 and 0.23 ± 0.07, respectively) compared to U-Net++ (0.73 ± 0.10 and 0.15 ± 0.07) and U-Next models (0.55 ± 0.09 and 0.40 ± 0.26). Thus, the U-Net model was selected for further testing.

The U-Net model demonstrated excellent performance for UTUC tumor segmentation with a mean AUC-ROC of 0.96 ± 0.06 and an overall pixel accuracy, sensitivity, and specificity of 0.98, 0.76, and 0.99, respectively. The model also showed good performance for segmentation of the area of ablated tumor with AUC-ROC of 0.90 ± 0.03 and DSC of 0.50 ± 0.05, and an overall pixel accuracy, sensitivity, and specificity of 0.87, 0.50, and 0.95, respectively (Fig. 2). Example frames from videos containing the raw input, prediction, and corresponding heatmaps are shown in Figure 3. Subanalysis showed decreased model performance for fiberoptic ureteroscope video (AUC-ROC of 0.60 ± 0.09, and DSC of 0.55 ± 0.11).

	Tumor Segmentation	Ablated Tumor Segmentation
AUC-ROC (mean ± SD)	0.96 ± 0.06	0.90 ± 0.03
DSC (mean ± SD)	0.78 ± 0.08	0.50 ± 0.05
Accuracy	0.98	0.87
Sensitivity	0.76	0.50
Specificity	0.99	0.95

FIG. 2. Summary statistics of the U-Net model during tumor and ablated tissue segmentation. Values include pixel-based evaluation of the model compared with the manually annotated frames. Area under the receiver operating curve and dice similarity coefficient values reported from the average of all annotated frames with standard deviation. Each light *blue curve* represents model performance for a single frame. Overall model accuracy, sensitivity, and specificity values reported across annotated frames.
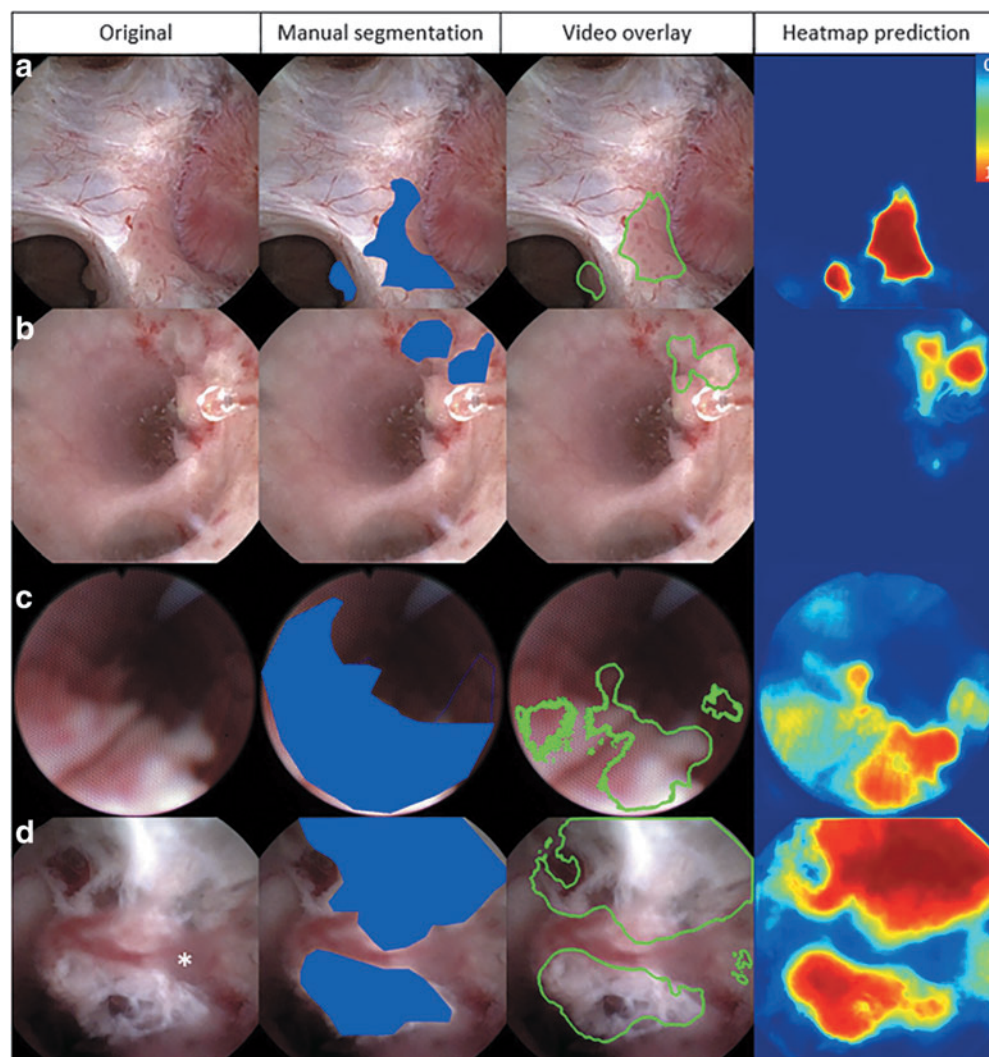
FIG. 3. Qualitative results of automated segmentation during endoscopic UTUC surgery demonstrating (*left* to *right*) the original endoscopic video image, manual segmentation performed by the surgeons (*blue* overlay), the contour of the automated segmentation (*green* overlay), and a heatmap prediction demonstrating the raw probability output per pixel. *Blue* and *red* pixels represent low and high probability of correct segmentation, respectively. **(a)** Digital scope tumor identification, **(b)** area of ablated tumor, **(c)** fiberoptic scope tumor identification, **(d)** example of model performing in instance with decreased visibility during tumor ablation. *Area of hemorrhage obstructing target tissue. UTUC = upper tract urothelial carcinoma.

During intraoperative testing, the model annotated new videos at 15 FPS and maintained segmentation performance while displaying a video overlay outlining the tumor (Fig. 4).

Discussion

We show that computer vision models can automatically segment UTUC tumors during endoscopic surgery. Specifically, our model demonstrated good segmentation performance for tumors and areas of ablated tumor. We further implemented these models intraoperatively and observed maintained performance while displaying an automated overlay at 15 FPS. Note that we visualize the contour of the automatically segmented tumor to avoid obstructing the video feed during active treatment. Taken together, these findings demonstrate the potential for improved endoscopic visualization during UTUC treatment.

Despite advances in endoscopic technology, UTUC recurrence rates are high.[1,18] Current guidelines recommend endoscopic surgery as the initial management for low-risk disease and select patients with high-risk disease who cannot undergo radical nephroureterectomy.[1] However, even large and more complex tumors can be effectively managed endoscopically.[19–21] For example, Scotland et al. evaluated the effectiveness of ureteroscopy with laser ablation for UTUC lesions that were larger than 2 cm (~20% of which were high grade) and found a 65% progression-free survival and 75% overall survival at 5 years.[19] Despite this, endoscopic management of UTUC remains underutilized compared to nephroureterectomy (index case rates of 11% *vs* 56%), even with a decrease in risk of complications, renal insufficiency, and dialysis dependence.[20–24]

Several challenges complicate endoscopic tumor ablation. Effective tumor ablation is impacted by endoscopic visibility
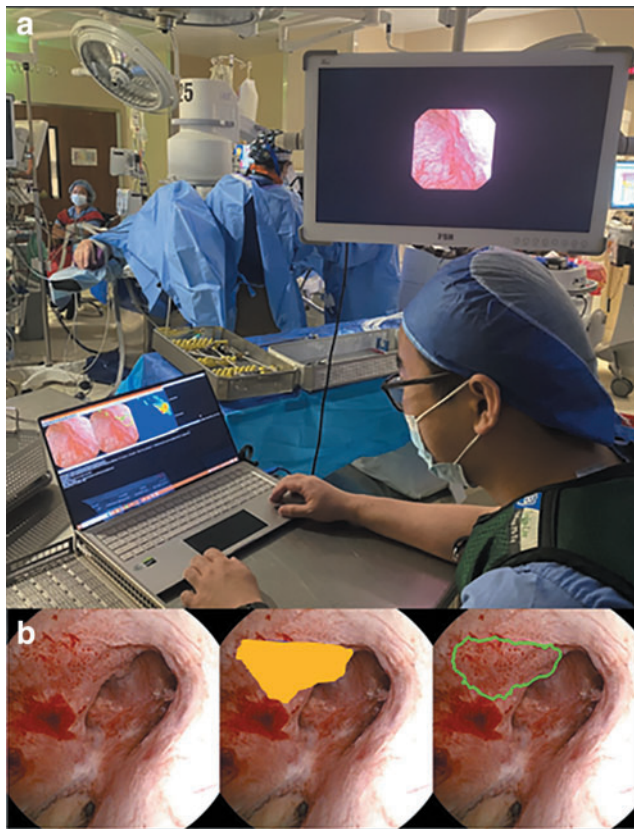
FIG. 4. **(a)** Intraoperative implementation of the segmentation model during endoscopic upper tract urothelial carcinoma treatment. **(b)** Example of qualitative segmentation results from the intraoperative experiment. (*Left panel*), the raw endoscopic video feed. (*Middle panel*), the automated segmented UTUC tumor (*orange* overlay). (*Right panel*), an outline of the tumor (*green* overlay) that could be used to visualize the model prediction without obstructing the surgeon's view during tumor treatment.

and localization of tumors, manually performed by the surgeon in the current paradigm. Factors such as debris and blood clots limit visibility and impact tumor treatment. In addition, most upper tract tumors are multifocal, and tracking multiple tumors in different parts of the branched renal collecting system is difficult and affected by clinician experience. Furthermore, effective tumor ablation is limited by endoscopic visibility due to the limited field and depth of view of current ureteroscopes. Often, to achieve complete tumor ablation in complex cases, multiple surgeries are required to adequately locate and visualize multiple tumors (or the full extent of larger tumors).[24] Tools that could enhance visualization and localization of tumors and areas of prior ablation, therefore, could improve outcomes after endoscopic UTUC surgery. Moreover, they could be used to improve endoscopic surgical education and assist less experienced surgeons.

We have previously evaluated computer vision models for kidney stone disease and demonstrated good performance both during stone localization and fragmentation.[8,25] In addition, we have previously identified automated performance metrics from computer vision models and associated them with surgeon experience during endoscopic stone surgery.[26] Moreover, similar computer vision models have been applied to endoscopic evaluation and treatment of bladder cancer.[27,28] Shkolyar et al. demonstrated a computer vision-based deep learning algorithm that could accurately detect bladder cancer during white light cystoscopy.[28] The algorithm has the potential to aid in training and clinical decision-making during cystoscopy and transurethral resection of bladder tumor by leveraging widely available software, rather than specialized equipment. Similarly, our models show the potential for improving decision-making during endoscopic UTUC treatment. Furthermore, in the future, this technology could be integrated with endoscopic robotic platforms and inform automated processes for tumor detection and treatment.

There are several limitations in this study. First, our model is trained on a limited dataset. The training set may not robustly represent the variety of endoscopic findings that could be seen during UTUC surgery. However, the model performed well despite the limited number of videos. More videos could be integrated in the workflow for model training to improve performance in the future. In addition, most videos used for training were from digital ureteroscopes. This likely explains the decrease in model performance during segmentation of fiberoptic videos. Newer ureteroscope technologies have been developed with an emphasis on digital video (i.e., disposable scopes) due to improved video quality. Likewise, we often utilize digital scopes for endoscopic UTUC treatment to improve visualization. Still, future models could incorporate more fiberoptic videos for generalizability.

In this study, we used tumors identified by expert surgeons as the ground truth to train our model. We thus do not know whether our model can detect tumors that are difficult or ambiguous to detect with the human eye and further testing would be required to evaluate the model's ability to identify these lesions. Moreover, the model currently segments surgical video at 15 FPS in the operating room using conventional graphics cards, which can lead to a slight lag time in overlaying the outline intraoperatively. However, the model can run on more advanced graphics cards (e.g., NVIDIA RTX A6000 Ada GPU) at 30 FPS or greater, which would improve the speed and make real-time accurate segmentation feasible. Finally, the current model was trained from single segmented frames without temporal context.

In future work, we will further improve the models by incorporating time through leveraging consecutive frames, which can improve performance in frames with partial or full occlusions. Despite these limitations, our study demonstrates the potential of computer vision models for segmentation of UTUC tumors during endoscopic surgery. Larger and more varied datasets will allow for more clinical applications of the models.

Conclusion

We have developed and evaluated a computer vision model that can segment tumors and areas of ablated tissue during endoscopic surgery for UTUC. The model was implemented intraoperatively and showed good performance. Moreover, our demonstration of a video overlay intraoperatively shows the potential for enhancing endoscopic visualization of UTUC, which could improve outcomes and tumor recurrence.

Authors' Contributions

D.L.: methodology, software, formal analysis, and writing-original draft, A.R.: data curation and writing-original draft, N.P.: data curation and writing-original draft, A.N.L.: data curation, conceptualization, and investigation, M.P.: data curation, writing- review and editing, and project administration, N.S.: conceptualization, data curation, and visualization, I.O.: conceptualization, methodology, writing-review and editing, supervision, and funding acquisition, N.K.: conceptualization, data curation, methodology, writing-review and editing, supervision, and funding acquisition.

Author Disclosure Statement

D.L., N.K., and I.O. have funding through the National Institute of Health. N.S. has funding through the Department of Defense.

References

1. Coleman JA, Clark PE, Bixler BR, et al. Diagnosis and management of non-metastatic upper tract urothelial carcinoma: AUA/SUO guideline. J Urol 2023;209(6):1071–1081; doi: 10.1097/JU.0000000000003480

2. Verges DP, Lallas, CD, Hubosky S, et al. Endoscopic treatment of upper tract urothelial carcinoma. Curr Urol Rep 2017;18(4):31; doi: 10.1007/s11934-017-0675-x

3. Abdelshehid C, Ahlering MT, Chou D, et al. Comparison of flexible ureteroscopes: Deflection, irrigant flow and optical characteristics. J Urol 2005;173(6):2017–2021; doi: 10.1097/01.ju.0000158139.65771.0a

4. Krogh J, Kvist E, Rye B. Transitional cell carcinoma of the upper urinary tract: Prognostic variables and post-operative recurrences. Br J Urol 1991;67(1):32–36; doi: 10.1111/j.1464-410x.1991.tb15064.x

5. Shen CY, Jou YC, Kan WC, et al. Outcome of non-muscle invasive upper tract urothelial carcinoma receiving endoscopic ablation: An inverse probability of treatment weighting analysis. J Clin Med 2022;11(5):1307; doi: 10.3390/jcm11051307

6. Wu S, Chen X, Pan J, et al. An artificial intelligence system for the detection of bladder cancer via cystoscopy: A multicenter diagnostic study. J Natl Cancer Inst 2022;114(2):220–227; doi: 10.1093/jnci/djab179

7. Ikeda A, Nosato H, Kochi Y, et al. Cystoscopic imaging for bladder cancer detection based on stepwise organic transfer learning with a pretrained convolutional neural network. J Endourol 2021;35(7):1030–1035; doi: 10.1089/end.2020.0919

8. Setia SA, Stoebner ZA, Floyd C et al. Computer vision enabled segmentation of kidney stones during ureteroscopy and laser lithotripsy. J Endourol 2023;37(4):495–501.

9. Skalski P. Make sense. Git Code. 2019. Available from: https://github.com/SkalskiP/make-sense [Last accessed: April 29, 2024].

10. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference. Springer: Cham; 2015; pp. 234–241.

11. Zhou Z, Rahman Siddiquee MM, Tajbakhsh N, et al. Unet++: A nested u-net architecture for medical image segmentation. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS. Springer: Cham; 2018; pp. 3–11.

12. Jégou S, Drozdzal M, Vazquez D, et al. The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2017; pp. 11–19.

13. Yakubovskiy, P. Segmentation models pytorch. GitHub Repos. 2020: 57. Available from: https://github.com/qubvel/segmentation_models.pytorch [Last accessed: April 29, 2024].

14. Ruby U, Yendapalli V. Binary cross entropy with deep learning technique for image classification. Int J Adv Trends Comput Sci Eng 2020;9(10):5393–5397.

15. Zhao R, Qian B, Zhang X, et al. Rethinking dice loss for medical image segmentation. In: IEEE International Conference on Data Mining (ICDM). IEEE: Sorrento, Italy; 2020; 851–860.

16. Kwong JCC, McLoughlin LC, Haider M, et al. Standardized reporting of machine learning applications in urology: The STREAM-URO framework. Eur Urol Focus 2021;7(4):672–682; doi: 10.1016/j.euf.2021.07.004

17. Lu D, Kavoussi N, Oguz I. Automated UTUC carcinoma tumor segmentation during ureteroscopy using computer vision techniques. Git Code. 2024. Available from: https://lu-d.github.io/utuc/ [Last accessed: April 29, 2024].

18. Cornu JN, Rouprêt M, Carpentier X, et al. Oncologic control obtained after exclusive flexible ureteroscopic management of upper urinary tract urothelial cell carcinoma. World J Urol 2010;28(2):151–156; doi: 10.1007/s00345-009-0494-x

19. Scotland KB, Kleinmann N, Cason D, et al. Ureteroscopic management of large ≥2 cm upper tract urothelial carcinoma: A comprehensive 23-year experience. Urology 2018;121:66–73; doi: 10.1016/j.urology.2018.05.042

20. Villa L, Haddad M, Capitanio U, et al. Which patients with upper tract urothelial carcinoma can be safely treated with flexible ureteroscopy with holmium:YAG laser photoablation? Long-term results from a high volume institution. J Urol 2018;199(1):66–73; doi: 10.1016/j.juro.2017.07.088

21. Shvero A, Abu-Ghanem Y, Laufer M, et al. Endoscopic treatment for large multifocal upper tract urothelial carcinoma. J Urol 2021;205(4):1039–1046; doi: 10.1097/JU.0000000000001505

22. Ng Chieng Hin J, Hettiarachchilage D, Gravestock P, et al. Role of ureteroscopy in treatment of upper tract urothelial carcinoma. Curr Urol Rep 2021;22(10):49; doi: 10.1007/s11934-021-01065-7

23. Cutress ML, Stewart GD, Tudor EC, et al. Endoscopic versus laparoscopic management of noninvasive upper tract urothelial carcinoma: 20-year single center experience. J Urol 2013;189(6):2054–2060; doi: 10.1016/j.juro.2012.12.006

24. Gadzinski AJ, Roberts WW, Faerber GJ, et al. Long-term outcomes of nephroureterectomy versus endoscopic management for upper tract urothelial carcinoma. J Urol 2010; 183(6):2148–2153; doi: 10.1016/j.juro.2010.02.005

25. Stoebner ZA, Lu D, Hong SH, et al. Segmentation of kidney stones in endoscopic video feeds. In: Medical Imaging 2022: Image Processing, Vol. 12032. SPIE Medical Imaging, San Diego, CA, USA; 2022; 900–908.

26. Cabo J, Lu D, Stoebner Z, et al. Using a computer vision-mediated analysis to distinguish surgeon experience during endoscopic stone surgery. J Urol 2023;209(Supplement 4): e953.

27. Lurie KL, Angst R, Zlatev DV, et al. 3D reconstruction of cystoscopy videos for comprehensive bladder records. Biomed Opt Express 2017;8(4):2106–2123.

28. Shkolyar E, Jia X, Chang TC, et al. Augmented bladder tumor detection using deep learning. Eur Urol 2019;76(6): 714–718.

Address correspondence to:
Nicholas Kavoussi, MD
Department of Urology
Vanderbilt University Medical Center
1211 Medical Center Drive
Nashville, TN 37323
USA

E-mail: nicholas.l.kavoussi@vumc.org

Abbreviations Used

AUC-ROC = area under the receiver operating curve
BCE = binary cross entropy
CT = computed tomography
DSC = dice similarity coefficient
FPS = frames per second
UTUC = upper tract urothelial carcinoma