LETTER

First-in-human real-time AI-assisted instrument deocclusion during augmented reality robotic surgery

Jasper Hofman[1,#] | Pieter De Backer[1,2,3,4,#] | Ilaria Manghi[5] | Jente Simoens[1] | Ruben De Groote[1,6] | Hannes Van Den Bossche[7] | Mathieu D'Hondt[8] | Tim Oosterlinck[9] | Julie Lippens[2] | Charles Van Praet[4] | Federica Ferraguti[5] | Charlotte Debbaut[3] | Zhijin Li[10] | Oliver Kutter[10] | Alexandre Mottrie[1,6,##] | Karel Decaestecker[2,11,##]

[1]ORSI Academy, Melle, Belgium

[2]Faculty of Medicine and Health Sciences, Department of Human Structure and Repair, Ghent University, Ghent, Belgium

[3]IBiTech-Biommeda, Faculty of Engineering and Architecture, and CRIG, Ghent University, Ghent, Belgium

[4]Department of Urology, Ghent University Hospital, Ghent, Belgium

[5]Department of Sciences and Methods for Engineering, University of Modena and Reggio Emilia, Modena, Italy

[6]Department of Urology, OLV Hospital, Aalst, Belgium

[7]Department of Urology, AZ West Hospital, Veurne, Belgium

[8]Department of Digestive and Hepatobiliary/Pancreatic Surgery, AZ Groeninge Hospital, Kortrijk, Belgium

[9]Faculty of Medicine, KU Leuven, Leuven, Belgium

[10]NVIDIA, Santa Clara, California, USA

[11]Department of Urology, AZ Maria Middelares Hospital, Ghent, Belgium

Correspondence
Pieter De Backer, ORSI Academy, Proefhoevestraat 12, 9090 Melle, Belgium.
Email: Pieter.De.Backer@orsi.be

[#]Shared first author
[##]Shared senior author

Abstract

The integration of Augmented Reality (AR) into daily surgical practice is withheld by the correct registration of pre-operative data. This includes intelligent 3D model superposition whilst simultaneously handling real and virtual occlusions caused by the AR overlay. Occlusions can negatively impact surgical safety and as such deteriorate rather than improve surgical care. Robotic surgery is particularly suited to tackle these integration challenges in a stepwise approach as the robotic console allows for different inputs to be displayed in parallel to the surgeon. Nevertheless, real-time de-occlusion requires extensive computational resources which further complicates clinical integration. This work tackles the problem of instrument occlusion and presents, to the authors' best knowledge, the first-in-human on edge deployment of a real-time binary segmentation pipeline during three robot-assisted surgeries: partial nephrectomy, migrated endovascular stent removal, and liver metastasectomy. To this end, a state-of-the-art real-time segmentation and 3D model pipeline was implemented and presented to the surgeon during live surgery. The pipeline allows real-time binary segmentation of 37 non-organic surgical items, which are never occluded during AR. The application features real-time manual 3D model manipulation for correct soft tissue alignment. The proposed pipeline can contribute towards surgical safety, ergonomics, and acceptance of AR in minimally invasive surgery.

---

1 | INTRODUCTION

Over the last decade, 3D models have entered oncologic surgery as a means to achieve better outcomes in renal and hepatic surgery [1, 2]. Nevertheless, the integration of 3D models into the operative field has been lacking due to three main reasons. Firstly, proper model alignment with the intraoperative anatomy has proven to be a major challenge due to shifting of organs during surgery and different patient positioning in surgery versus during computed tomography [2, 3]. Secondly, automated organ registra-tion in a continuously moving surgical video has been another major challenge for many years [4]. Thirdly, 3D model overlay obscures the surgical field, including sharp surgical instruments which are manipulated, hence creating a possible hazardous situation rather than facilitating surgery. The latter occlusion problem has been a longstanding study topic [5] which, if solved, would further advance various surgical domains and applications [6]. Already in 2004, Fischer et al. [7] proposed handling instrument occlusion in medical augmented reality (AR) through identification of occlusion zones by creating a virtual map of the existing environment upfront. Four years later, Kutter et al. [8] explored the design and implementation of a high-quality hardware system to enable real-time volume rendering in AR applications. However, to ensure that depth perception was not compromised, the authors needed to apply video colour filtering to handle occlusion, as such limiting its robustness due to the colour prior. Other approaches [9] for occlusion management relied on tracking and 3D positioning of the instrument within the AR environment, which in turn made it prone to certain instruments' directions. As such, previous de-occlusion attempts were unsuccessful in detecting all surgical items with sufficient robustness, whilst at the same time having no prior knowledge of the objects' orientations or positions inside the real-time surgical environment. More recent work [10] showed the potential of deep learning binary instrument segmentation for robust de-occlusion during AR surgery. However, the reported latency reached up to 0.5 s and was considered unfeasible for real-time surgical use.

In this work, a robust real-time binary segmentation pipeline for non-organic items was developed and deployed during three live robot-assisted surgeries: partial nephrectomy (RAPN), migrated endovascular stent removal and liver metastasectomy. Through the use of a state-of-the-art binary segmentation method, as well as software and hardware level acceleration, the pipeline efficiently tackles instrument occlusion caused by the AR 3D model overlay in real-time and reduces delay to a frame-by-frame latency of 13 ms. Qualitative surgical feedback stated the resulting perceived end-to-end latency is acceptable for real-time surgery.

2 | MATERIALS AND METHODS

2.1 | Non-organic binary segmentation

The binary segmentation data set contains 31,812 images on which all non-organic items were manually delineated in the annotation platform SuperAnnotate (Sunnyvale, CA, USA) [11]. The 37 different non-organic items include robotic and laparoscopic instruments, needles, wires, clips, vessel loops, bulldogs, gauzes etc. The images were sampled uniformly across 100 full- length RAPN procedures. The data set was split on a procedural base into 24,087 images for training, 4545 images for validation, and 3180 images for testing. Different encoder–decoder deep learning architectures were evaluated for performance. A Feature Pyramid Network (FPN) architecture [12] with EfficientNetV2 encoder backbone [13] was identified as best performing in a separate optimization study. The model was trained over 50 epochs with batch size 16 and image size 512×512 pixels, using Adam optimizer with a learning rate of 2.25×10^{-4} and a combination of focal and dice loss. The learning rate was reduced on plateau over five epochs with a factor 0.7 and an early stopping criterion was evaluated on the mean Intersection over Union (IoU) with a patience of 15 epochs. It is subsequently compared to other recent work in this field (DeepLabV3+ architecture [10]), in terms of mean IoU and processing time. Inference in the application pipeline requires conversion to ONNX and subsequent TensorRT optimization[1]. This optimization allows for smaller precision, reduced latency and model size, simplified network topology, reduced read and write operations, and dynamic memory allocation to reduce memory footprint. This type of optimization is necessary to meet the real-time needs during surgery. A side-by-side performance comparison of both architectures is performed for both the original Pytorch model and the final implemented TensorRT model. Both are evaluated for Floating Point (FP) 16 and FP32 precision.

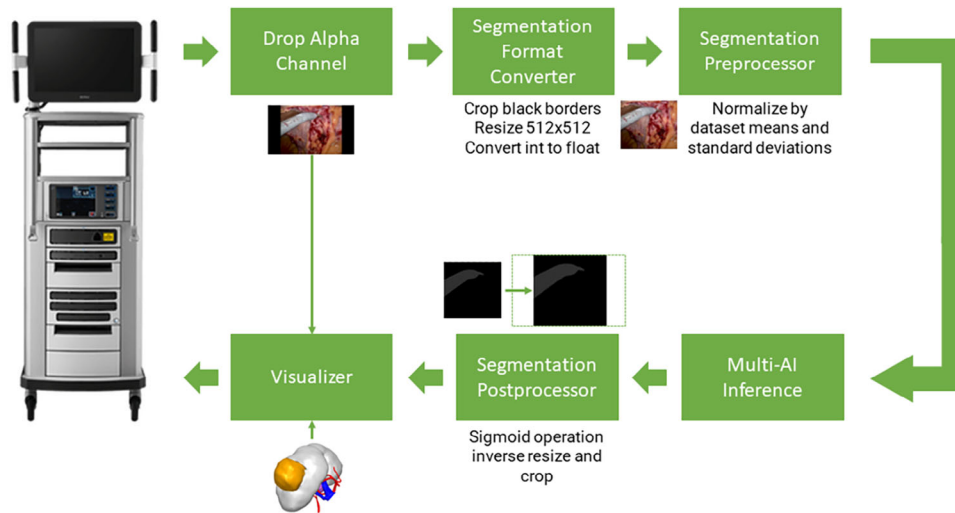# 3 | PIPELINE DEPLOYMENT

3.1 | Hardware framework

This integration addresses delays imposed by different serial components whilst enabling the implementation of more accurate but also heavier deep learning networks for image segmentation. The Nvidia Clara AGX developer kit[2,3] (Nvidia, Santa Clara, California, USA) was identified as embedded computing architecture for highly demanding video processing applications. Live video capture was enabled through a Deltacast DELTA-12G-elp-key capture card[4] (Deltacast, Liege, Belgium). The card provides performance and efficient I/O, as well as a passive bypass, which safeguards original video throughput in case of real-time software malfunctioning.

---

[1] NVIDIA TensorRT: https://developer.nvidia.com/tensorrt.

[2] NVIDIA Developer Kits for medical devices: https://www.nvidia.com/en-gb/clara/intelligent-medical-instruments/.

[3] Clara AGX Product Brief with Technical details: https://resources.nvidia.com/en-us-enabling-smart-hospitals-ai-ep/nvidia-clara-agx-dev?lx=KWlJE5&xs=301547.

[4] DELTA-12G-ELP-KEY 11 details: https://www.deltacast.tv/products/developer-products/sdi-capture-cards/delta-12g-elp-key-11.

FIGURE 1 Schematic overview of the different steps and Graphical eXchange Format (GXF)-extensions in the segmentation application.

3.2 | Software framework

The intra-operative AI and AR application was developed using the NVIDIA Holoscan SDK[5], an extensible open-source framework for implementing real-time, low-latency medical AI applications. The pipeline was implemented through a combination of existing Holoscan operators, extended with use case tailored ones. Figure 1 displays a schematic overview of the pipeline and corresponding operators. The pipeline can be divided into four main blocks: pre-processing, inference, post-processing, and visualization. The 1920 × 1080 p captured frames are reformatted to serve as input to the segmentation model. For every frame, the alpha-channel information is dropped, black borders are removed, and the frame is resized to 512 × 512 pixels. The colour channels are normalized with means and standard deviations derived from the binary segmentation model training set.

After inference, a sigmoid activation is applied to the 512 × 512 × 2 output, yielding a 512 × 512 pixels binary mask indicating whether the corresponding pixels make up non-organic items or soft tissue. The 512 × 512 pixels segmentation mask is subsequently resized to match the original input resolution. The 3D model is rendered through Visualization Toolkit (VTK)[6] and composited with the live full-quality surgical video and segmentation mask to create the final image. The 3D models are manually segmented pre-operatively using Mimics (Materialise, Leuven, Belgium) from a 4-phasic CT scan sequence [14]. The models consist of separate Standard Triangle Language (STL) files for different structures like parenchyma, tumours, stents, arteries, veins, and other anatomical entities relevant to the procedure (Figure 4f). All STL files can be toggled with hot-keys during overlay and the transparency can be edited in real-time. The 3D model requires manual alignment with the surgical scene as can be seen in Figure 2a.

4 | LIVE SURGERIES

The study was performed with patient consent under institutional review board approval (B6702020000442). Figure 2 displays the operating room setup during the live RAPN procedure. The DELTA-12G-elp-key capture card, which is integrated into the Holoscan box, receives the live video feed through serial digital interface (SDI). The patient-specific 3D model was preloaded onto the Clara AGX developer kit for rendering inside the Holoscan application. The processed frames were sent over DisplayPort-out into an active HDMI-splitter. One of both HDMI-out signals was sent to a monitor through which the users could continuously interact with the application. The other signal was fed back with an HDMI-DVI cable into the Intuitive Xi robotic system (Intuitive, California, USA) by means of the DVI-TilePro-input (Figure 2c). As such, whenever the surgeon enables the TilePro feature, perfectly aligned. Furthermore, a keyboard and mouse were connected to the Clara AGX developer kit to be able to manipulate and interact with the 3D model as discussed above. Figure 2a shows the physical experiment setup during the in-human procedure in the operating theatre. The Clara AGX developer kit and other hardware were installed next to the surgical tower, allowing continuous model alignment during the procedure and direct communication between the user and the surgeon whenever needed.
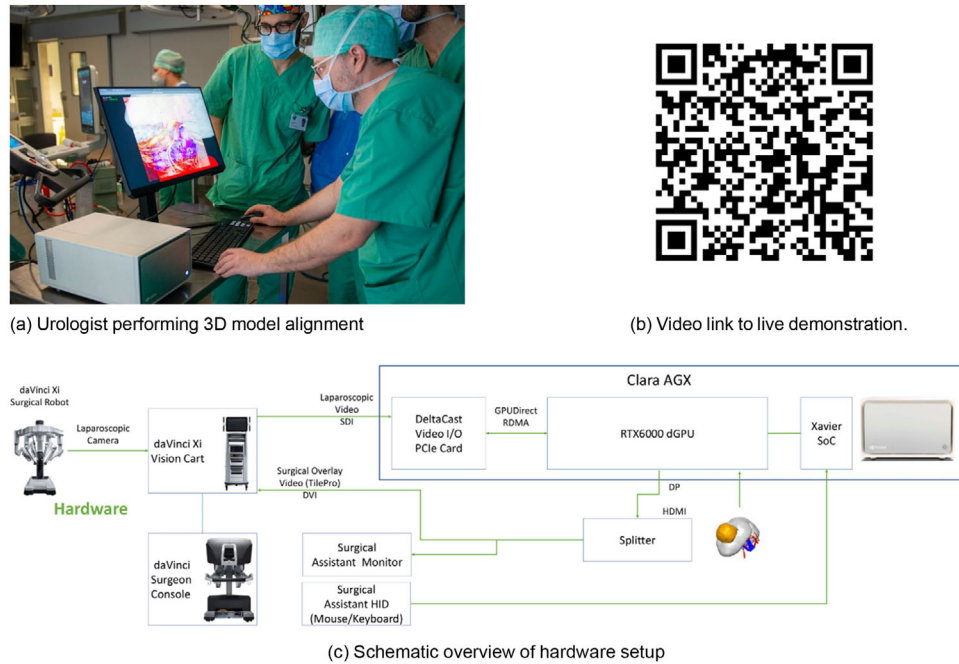
5 | RESULTS

5.1 | Non-organic binary segmentation

Table 1 summarizes the comparison of evaluation metrics of the FPN model versus the DeepLabV3+ model [10] on an identical test set of 1345 images derived from four distinct RAPN procedures. The baseline model for the reported side-by-side improvements is the DeepLabV3+ with FP32 precision, implemented in Pytorch. The Δ mean IoU and Δ inference

---

[5] Holoscan SDK: https://github.com/nvidia-holoscan/holoscan-sdk.
[6] Visualization Toolkit (VTK): https://vtk.org/

(a) Urologist performing 3D model alignment

(b) Video link to live demonstration.

(c) Schematic overview of hardware setup

FIGURE 2 Experiment setup overview. (a) Urologist performing 3D model alignment on the Clara AGX developer kit positioned next to the surgical tower. (b) Video link to live demonstration. (c) Schematic overview of hardware setup and connection types.

TABLE 1 Performance comparison of (TensorRT optimized) DeepLabV3+ and FPN segmentation model architectures evaluated on Pytorch container.

Model	Precision	mean IoU		Inference time (ms)		Δ IoU		Δ inference time (ms)	
		Pytorch	TensorRT	Pytorch	TensorRT	Pytorch	TensorRT	Pytorch	TensorRT
DEEPLAB	FP32	0.90318	0.90318	48.6	20.4	N/A	0	N/A	−28.2
DEEPLAB	FP16	0.90317	0.90316	52.6	8.5	−0.00001	−0.00002	4	−40.1
FPN	FP32	0.94621	0.94621	36	14.2	0.04303	0.04303	−12.6	−34.4
FPN	FP16	0.94621	0.94623	40.5	5.1	0.04303	0.04305	−8.1	−43.5

FPN, feature pyramid network; IoU, intersection over union.

(a) Original image (b) DeepLabV3+ FP16 (c) DeepLabV3+ FP32 (d) FPN FP16 (e) FPN FP32
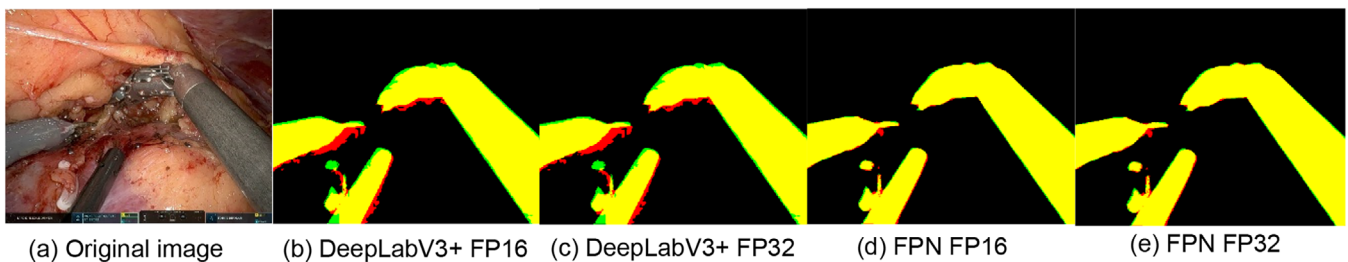
FIGURE 3 Segmentation performance of DeepLabV3+ and FPN models (TensorRT optimized). The yellow regions indicate true positive pixels, the red regions indicate false-positive pixels, and the green regions indicate false-negative pixels. FPN, feature pyramid network.

time reflect stepwise improvements in segmentation quality and inference time, with respect to this baseline. Changing the model architecture to FPN resulted in increased mean IoU while TensorRT optimization significantly reduced inference times, with the largest time reduction for FP16 precision. Figure 3 provides an example of the improvement in segmenta-

tion quality in TensorRT. When compared to DeepLabV3+, the FPN model reduced both false-positive and false-negative pixel regions (respectively represented by red and green regions). Switching to FP16 has no effect on segmentation performance for both FPN and DeepLabV3+ architectures. The TensorRT FPN model with FP16 precision was identified as the most
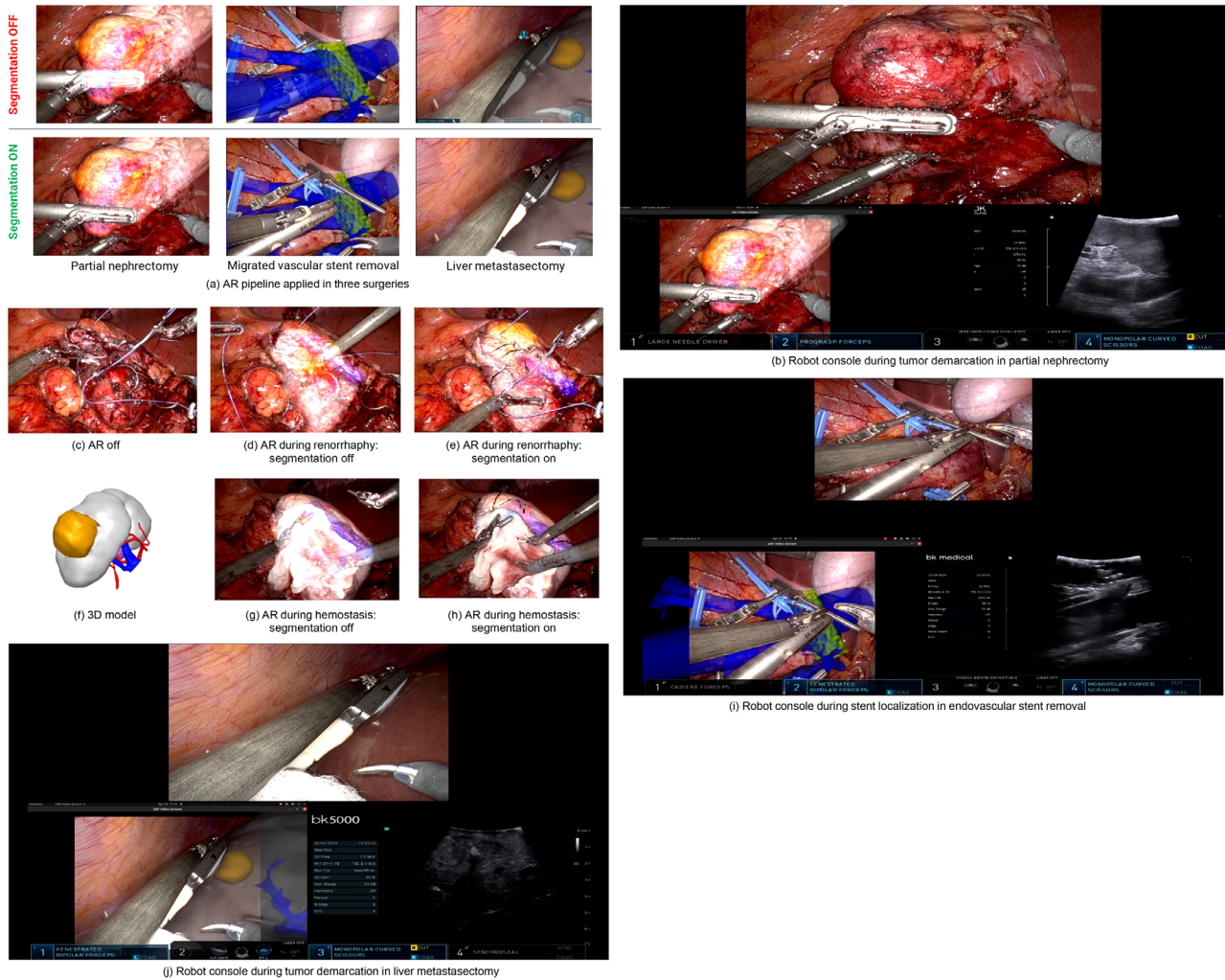
FIGURE 4 Illustration of different segmentation scenarios during three live surgeries. (a) Shows the AR pipeline applied in liver metastasectomy, endovascular stent removal, and partial nephrectomy, with both enabled and disabled segmentation to illustrate de-occlusion of non-organic items. (c) Displays renorrhaphy after tumour resection with AR inactive. (j, i, and b) Show the respective console views during surgery with two Tile Pro inputs. The tumour or stent localization and 3D model alignment (left) is being validated by the ultrasound probe (right). (f) Shows an example of a patient-specific 3D model. (d, e) Show the 3D model overlaid with the surgical scene with segmentation off and on respectively to illustrate segmentation performance on smaller non-organic items. (g, h) Display respectively segmentation off and on during the application of surgical hemostatics. AR, augmented reality.

promising network for the live experiment. To profile the application pipeline, the Nvidia Nsight Systems profiling tool[7] was used. The experiments were run over 20 s at an input frame rate of 80 frames per second. The resulting median processing time for the application pipeline was less than 13 ms with an average GPU utilization of 42%. As such, the device still has GPU bandwidth for additional workloads [15].

5.2 | Live demonstration and user feedback

Figure 2b shows a QR code with a link to a video containing highlighted segments during surgery. During the RAPN, the application was first enabled after the identification and isolation of the renal artery. The 3D model overlay confirmed the orientation of the kidney and tumour with respect to the artery, providing initial support for navigation, as well as confirmation of clamping level with respect to possible earlier bifurcations of the vessel. During this phase both our previous solution [10] and current solution were simultaneously visually compared through separate Tile Pro inputs inside the console. The surgeon (R.D.G) stated a significant improvement in perceived latency, where the latency of the current solution is small enough for surgical adoption. Thereafter, the AR application was enabled during tumour demarcation. Figure 4b displays the surgeon's console view, where the application is used in parallel with the ultrasound probe. The resection margins and tumour depth estimation are augmented by the 3D model overlay with the segmented endoscopic ultrasound probe on top and as confirmed by the ultrasound imagery. Figure 4b also illustrates that the application solely served as support next to the endoscopic vision as not to impair or alter the surgeon's original vision or

---

decision. The application was not used during the time-critical surgical phases for renal artery clamping and tumour resection. After tumour resection and arterial unclamping, the application was once more enabled during renorrhaphy. Figures 4d and 4e illustrate the model's segmentation performance on suturing needles, wires, and hem-o-lok clips as other non-organic materials. Finally, Figures 4g and 4h illustrate the performance for gauze segmentation during hemostasis. The 3D setup was experienced as easy to manipulate without prior knowledge by the clinician performing the alignment (H.V.D.B.). The ability to align the 3D overlay and toggle the tumour visibility added insights regarding localization of the tumour bed, while the segmentation effectively provided a sense of depth while suturing the renal capsule. Automatic model alignment was reported to be the major next clinical improvement. The second case entails a robotic removal of a migrated endovascular stent, placed for a nutcracker syndrome. This stent had migrated into the vena cava inferior causing relapsing symptoms and danger of further migration towards the right atrium. The stent was removed and vena cava reconstruction with left renal vein transposition was performed to treat the primary nut cracker syndrome. Figure 4i displays the surgeon's console view, with the endoscopic ultrasound depicted in the right lower TilePro window. We note the validation of the stent location, represented by the oval hyperreflective structure in the top of the ultrasound image. The surgeon (K.D.) confirmed that the delay was negligible and acceptable for surgery, and that the AR Tile Pro input was sufficiently informative and responsive to even be used as the main screen during this phase of the surgery. Finally, the AR pipeline was applied during a robotic liver metastasectomy. Figure 4j shows the surgeon's console view where tumour demarcation and 3D model alignment are again validated using ultrasound. The surgeon (M.D.) stated that, although this setup solves the delay and de-occlusion problem, the application is not yet applicable in liver surgery due to the organ's deformative nature which complicates 3D model alignment. As for the first case, automatic model alignment with extension to deformable registration was reported to be the next major clinical improvement. All patients experienced a normal postoperative course and recovery.

6 | CONCLUSION AND FUTURE WORK

This work presents the implementation of a robust novel real-time approach for occlusion handling in surgical AR scenarios. It shows that AR-induced instrument occlusion is a resolvable issue when integrating software directly in a dedicated hardware pipeline. Our segmentation algorithm is shown to transfer smoothly across three different robot-assisted renal surgeries and the setup is applicable across Intuitive Xi systems, as shown in three different testing hospitals. Despite being trained only on robot- assisted partial nephrectomy instrument segmentation, the algorithm seems to generalize well across other surgeries using similar instruments. This could facilitate towards a broader AR adoption in robotic surgery. The subjective surgical feedback indicated that the application can bring clinical value to several parts of the procedure and that the delay is acceptable for real-time surgery. Specific perceived surgical benefits include better insights into tumour localization below the renal surface and corresponding arterial tree, together with improved tumour delineation due to de-occlusion. The pipeline is built on top of extensible open-source technologies, allowing to replicate and translate the work to other challenges in computer-assisted intervention (CAI) and surgical data science (SDS) for real-time adoption in surgery. By solving the long-standing problem of real-time instrument occlusion, the work is a demonstrator for translational research from lab to operating room. Furthermore, by optimizing compute resources, a frame latency of less than 13 ms and average GPU utilization of 42% was achieved. These results reflect room for the integration of additional workloads such as parallel deep learning inference pipelines. Future work resides in the implementation of a parallel soft tissue segmentation pipeline for automatic 3D model alignment, and by extension non-rigid body registration. The system should be further evaluated on user experience in a more formalized manner, by, for example, constructing a questionnaire and applying the pipeline in a greater number of surgeries. Other future minor hardware improvements include the use of the capture card SDI-out to further reduce delays and enable continuous passive bypass for surgical safety.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The segmentation algorithm, and the real-time holoscan application are publicly available for non-commercial use under the Creative Commons Attribution CC-BY-NC-SA 4.0 at https://github.com/nvidia-holoscan/holohub. Due to privacy restrictions, the datasets used in the present work cannot be publicly shared. If readers want to use this algorithm, they must cite this article or acknowledge Orsi Academy and the primary authors, Jasper Hofman and Pieter De Backer. For inquiries on commercial use, the corresponding author Pieter De Backer should be contacted.

ORCID

Jasper Hofman 🄳 https://orcid.org/0009-0001-1867-4762

REFERENCES

1. Piramide, F., Kowalewski, K.F., Cacciamani, G., et al.: Three-dimensional model–assisted minimally invasive partial nephrectomy: A systematic review with meta-analysis of comparative studies. Eur. Urol. Oncol. 5(6), 640–650 (2022)
2. Acidi, B., Ghallab, M., Cotin, S., Vibert, E., Golse, N.: Augmented reality in liver surgery, where we stand in 2023. J. Visc. Surg. 160(2), 118–126 (2023)
3. Khaddad, A., Bernhard, J.C., Margue, G., et al.: A survey of augmented reality methods to guide minimally invasive partial nephrectomy. World J. Urol. 41(2), 335–343 (2023)
4. Madad Zadeh, S., Francois, T., Calvet, L., et al.: SurgAI: Deep learning for computerized laparoscopic image understanding in gynaecology. Surg. Endosc. 34, 5377–5383 (2020)
5. Qian, L., Wu, J.Y., DiMaio, S.P., Navab, N., Kazanzides, P.: A review of augmented reality in robotic-assisted surgery. IEEE Trans. Med. Robot Bionics 2(1), 1–16 (2019)
6. Suzuki, R., Karim, A., Xia, T., Hedayati, H., Marquardt, N.: Augmented reality and robotics: A survey and taxonomy for ar-enhanced human-robot interaction and robotic interfaces, In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, pp. 1–33 (2022)
7. Fischer, J., Bartz, D., Straßer, W.: Occlusion handling for medical augmented reality using a volumetric phantom model. In: Proceedings of the ACM Symposium on Virtual Reality Software and Technology, pp. 174–177 (2004)
8. Kutter, O., Aichert, A., Bichlmeier, C., et al.: Real-time volume rendering for high quality visualization in augmented reality. International Workshop on Augmented environments for Medical Imaging including Augmented Realiy in Computer-aided Surgery (AIM-ARCS), pp. 104–113 (2008)
9. Frikha, R., Ejbali, R., Zaied, M.: Handling occlusion in augmented reality surgical training based instrument tracking. In: IEEE, pp. 1–5 (2016)
10. De Backer, P., Van Praet, C., Simoens, J., et al.: Improving augmented reality through deep learning: Real-time instrument delineation in robotic renal surgery. Eur. Urol. 84(1), 86–91 (2023)
11. De Backer, P., Eckhoff, J.A., Simoens, J., et al.: Multicentric exploration of tool annotation in robotic surgery: Lessons learned when starting a surgical artificial intelligence project. Surg. Endosc. 36(11), 8533–8548 (2022)
12. Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S.: Feature pyramid networks for object detection. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2117–2125 (2017)
13. Tan, M., Le, Q.: Efficientnetv2: Smaller models and faster training. In: PMLR, pp. 10096–10106 (2021)
14. De Backer, P., Vermijs, S., Van Praet, C., et al.: A novel three-dimensional planning tool for selective clamping during partial nephrectomy: Validation of a perfusion zone algorithm. Eur. Urol. 83(5), 413–421 (2023). https://doi.org/10.1016/j.eururo.2023.01.003
15. De Backer, P., Simoens, J., Mestdagh, K., Hofman, J.: Automated robotic surgical video anonymization: Enabling privacy proof video sharing through deep learning. Session Useful Techniques for your Future OR - European Association of Endoscopic Surgery Congress - Rome (2023)