NephCNN: A deep-learning framework for vessel segmentation in nephrectomy laparoscopic videos

Alessandro Casella*‡, Sara Moccia†*, Chiara Carlini‡, Emanuele Frontoni†, Elena De Momi‡, Leonardo S. Mattos*
*Department of Advanced Robotics, Italian Institute of Technology, Genoa, Italy
†Department of Information Engineering, Università Politecnica delle Marche, Ancona, Italy
‡Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milan, Italy

***Abstract*—Objective: In the last years, Robot-assisted partial nephrectomy (RAPN) is establishing as elected treatment for renal cell carcinoma (RCC). Reduced field of view, field occlusions by surgical tools, and reduced maneuverability may potentially cause accidents, such as unwanted vessel resection with consequent bleeding. Surgical Data Science (SDS) can provide effective context-aware tools for supporting surgeons. However, currently no tools have been exploited for automatic vessels segmentation from nephrectomy laparoscopic videos. Herein, we propose a new approach based on adversarial Fully Convolutional Neural Networks (FCNNs) to kidney vessel segmentation from nephrectomy laparoscopic vision. Methods: The proposed approach enhances existing segmentation framework by (i) encoding 3D kernels for spatio-temporal features extraction to enforce pixel connectivity in time, and (ii) perform training in adversarial fashion, which constrains vessels shape. Results: We performed a preliminary study using 8 different RAPN videos (1871 frames), the first in the field, achieving a median Dice Similarity Coefficient of 71.76%. Conclusions: Results showed that the proposed approach could be a valuable solution with a view to assist surgeon during RAPN.**

***Index Terms*—kidney segmentation, fully convolutional neural networks, adversarial training, blood vessel segmentation**

I. Introduction

In 2018, the estimated prevalence of renal cancer in Europe was around 136.5 thousands cases (3.5% of total cancer cases) and this number has been growing since 2012 [1]. The mortality rate of people suffering from renal cancer is around 40% [2]. Renal cell carcinoma (RCC), which is the most common kidney cancer in adults, accounts for approximately $2-3\%$ of all adult malignancies [3].

Nowadays, with a 5-year after treatment survival rate equal to 95%, nephrectomy is considered the main treatment option for RCC [4]. In the last few decades, robot-assisted partial nephrectomy (RAPN) has been increasingly adopted in the treatment of RCC, offering significant advantages in terms of patient safety, health outcomes and hospitalization costs over conventional open and laparoscopic nephrectomy [5].

Despite the advantages brought by RAPN, surgeons still have to face many obstacles, such as small field of view, field occluded by surgical tools, and reduced maneuverability. These obstacles highlight the need of introducing, in the operating room, computer-assisted tools to support surgeons and enhance his/her view. This could potentially reduce the occurrence of intraoperative accidents (*e.g.*, unwanted vessel resection, bleeding due to surgical tool misplacement), while minimizing surgeon mental workload [6]. Technical challenges include the detection and localization of anatomical structures and surgical instrumentation [7], [8], intraoperative registration [9], and workflow modeling and recognition [10].

The specific aim of this work is to provide vessel segmentation in nephrectomy laparoscopic videos. The segmentation could represent the baseline for enhanced surgeon visualization (e.g., through highlighting vessels during surgery) and also serve as baseline for vessel avoidance with virtual-fixture control [11]. Nonetheless, despite its importance, few efforts have been put in vessel segmentation from nephrectomy laparoscopic videos. This may be mainly due to the shortage of public available annotated datasets and the challenges of the task, such as presence of surgical tools and several anatomical structures in the image, vessels covering a small portion of the image and with varying position in the image, as well as low image quality due to noise and blur.

To cope with these issues, researchers in closed fields highlighted the importance of integrating temporal information [8], naturally encoded in the video, and shape-constraining training strategy, to improve segmentation performance [12].

Following this considerations, we propose an adversarial strategy to train a 3D Fully-Convolutional Neural Network (FCNN) (the *segmentor*) for automatic renal artery segmentation in RAPN videos. Here, the third dimension refers to time, allowing for spatio-temporal features extraction naturally encoded in intraoperative videos. In this preliminary study based on 8 RAPN annotated videos (1871 frames), we investigated the following research hypotheses: (i) the use of spatio-temporal information improves features extraction and (ii) shape-constrained training can boost segmentation performance by enforcing segmentation mask contiguity across frame and preserving shape.

A. Related work on segmentation in laparoscopic images

Blood vessel segmentation is a topic of high interest in medical research [14]. However, focusing on intraoperative laparoscopic images, the work in [15] represents one of the few approaches to blood vessels segmentation.

Anyway, an established literature on intraoperative tissue segmentation exists. First attempts used handcrafted features (e.g., [7], [16]–[18]). In the last years, FCNNs have become the standard for segmentation in laparoscopy. The typical use

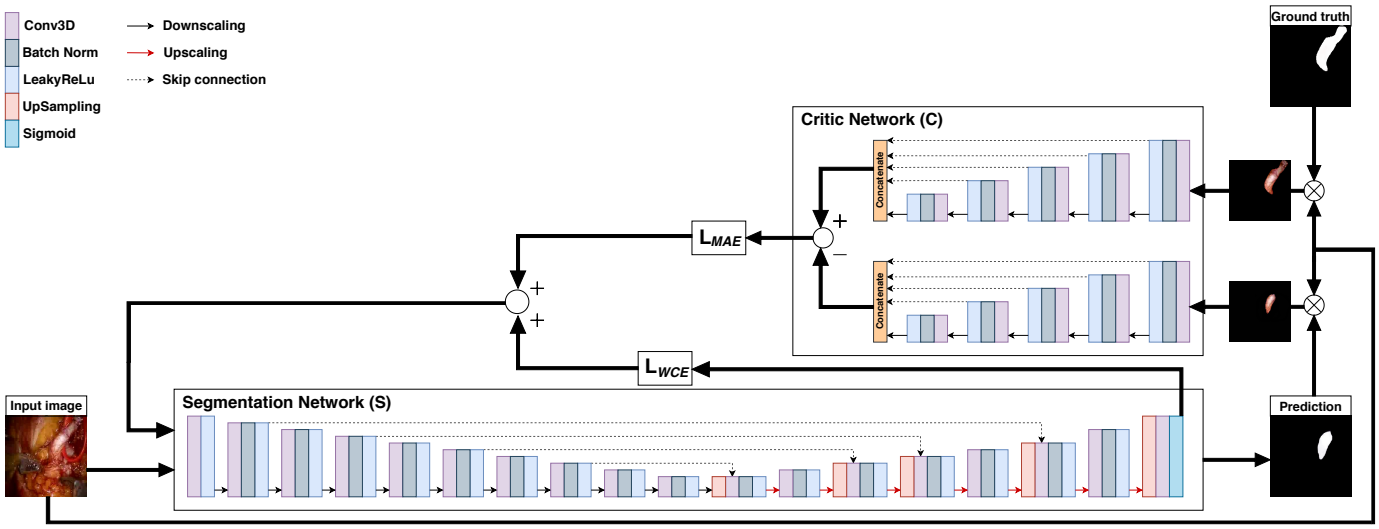


Fig. 1: Schematic representation of *NephCNN* architecture for vessel segmentation in intraoperative nephrectomy. The *segmentor* is a U-shaped network with long-skip connections (dashed arrows) inspired on 3D U-Net [13] architecture. Each block of descending path (black thin arrows) is composed of a 3D convolution followed by batch normalization (BN) and *Leaky ReLU* activation function. Only the first downscaling and the last upscaling blocks do not include a BN layer. In the ascending path (red thin arrows), Conv3D-BN-LeakyReLU modules alternate with UpSampling-Conv3D-BN-LeakyReLU ones. The last upscaling block has sigmoid activation function instead of *Leaky ReLU* one. The *critic*, inspired by [12], consists of the encoder branch of 3D U-Net. During the training process, illustrated in Sec. II-C, the *critic* extracts the feature vectors from the input image masked with the *segmentor* output and with the ground-truth. The Mean Absolute Error (MAE) computed between the two feature vectors contributes to the loss that is minimized during training. The other contribution to this loss is given by the *segmentor* loss, *i.e.* the Weighted Cross-Entropy Loss (L_{WCE}).

of a CNN for vessel segmentation is based on pixel–wise classification: each pixel in the image is assigned to the class vessel or background. [19]–[21] Recently, the work in [8] introduced the idea of exploiting temporal information for processing laparoscopic videos. Specifically, a 3D FCNN was implemented to extract spatio-temporal features for instrument joint detection in order to exploit the temporal information intrinsically embedded in laparoscopic videos.

A second recent innovation in the field has been brought by the introduction of adversarial networks for shape-constrained segmentation. Some interesting results were presented in [22], [23] for blood vessel segmentation and, more recenlty, in [12] for fetal surgery.

Following considerations in the literature, the purpose of this work is to exploit the potentiality of 3D FCNNs, in combination with adversarial training, to provide automatic and robust vessel segmentation in nephrectomy laparoscopic images.

II. Methods

The proposed method consists of a 3D FCNN *segmentor* (Sec. II-A) and a discriminator network, called *critic* (Sec. II-B), both trained in an adversarial fashion (Sec. II-C). The overall architecture of *NephCNN* is shown in Fig. 1.

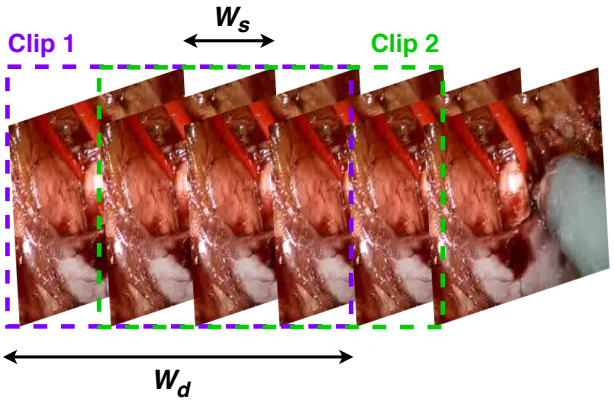


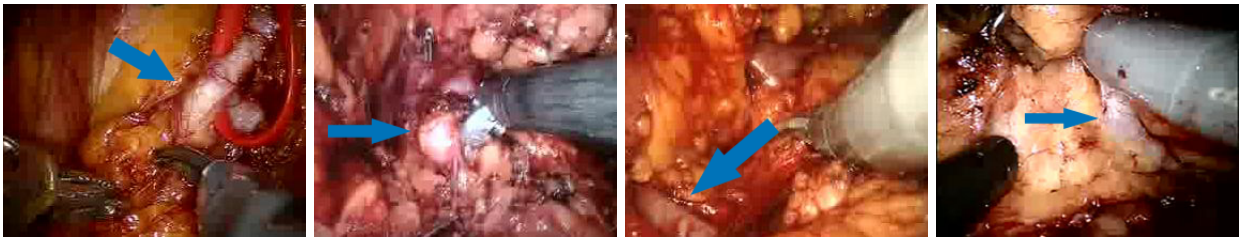
Fig. 2: Sliding window algorithm: starting from the first video frame, an initial clip with W_d frames (dotted violet line) is selected and combined to generate a 4D datum of dimensions image width $\times$ image height $\times$ W_d $\times$ 3. Then the window moves of W_s frames along the temporal direction and a new clip (dotted green line) is selected.

A. Segmentor

The *segmentor* has a U-Net-like architecture [24] consisting of downsampling and upsampling paths, made of 10 pro-

TABLE I: Specification of the initial dataset, the number of frames for each patient is shown.

Dataset: 1871 frames							
Patient 1	Patient 2	Patient 3	Patient 4	Patient 5	Patient 6	Patient 7	Patient 8
240	240	240	240	210	240	240	221



(a) (b) (c) (d)

Fig. 3: Challenges in the acquired dataset include varying vessel position and size, vessel occlusion by surgical tools, presence of noise, blur, and varying illumination level.

cessing blocks each, linked via long-skip connections. Each processing block of descending path is composed of a convolution followed by batch normalization (BN) and *Leaky ReLU* activation function. Only the first downscaling and the last upscaling blocks do not include a BN layer. In the ascending path, processing block alternate with upsampling layers. The output module, at the end of the *segmentor*, consists of a 1x1 convolution layer activated with the sigmoid function.

To take the temporal information into account, we use 3D convolutional layers to build the processing block. Hence, the input of our *segmentor* is a temporal clip (*i.e.*, set of W_d temporally consecutive video frames) obtained with a sliding window algorithm, as in [8]: starting from the first video frame, the first W_d frames contribute to the temporal clip of dimensions *width* $\times$ *height* $\times$ W_d $\times$ $N_{channels}$, where *width* and *height* are the frame width and height, respectively, and $N_{channels}$ is the number of image channels. The window then slides of W_s frames along the temporal direction, generating a collection of overlapping temporal clips. To feed the 3D architectures, temporal clips were obtained with the sliding-window algorithm proposed in [8]. The algorithm is controlled by the window temporal length (W_d) and step (W_s). Starting from the first video frame, the first W_d images are collected and used to generate a 4D tensor of dimension *width* $\times$ *height* $\times$ W_d $\times$ $N_{channels}$, where $N_{channels} = 3$ for RGB images. The window then moves of W_s frames along the temporal direction and a new temporal clip is generated. This results in a collection of 4D clips. The *segmentor* produces as output 4D tensors, each one of dimension *width* $\times$ *height* $\times$ W_d $\times$ $N_{channels}$, where $N_{channels} = 1$. These outputs are 4D clips of binary segmentation masks, each referring to the renal artery in the current frame.

B. Critic

The *critic* is fully inspired by that proposed in [12], and consists of a standard U-Net encoding path for spatial feature extraction to preserve segmentation mask macro-appearance, The *critic* implements 3D convolution to process the temporal information while performing feature extraction. It receives a two-fold input, which consists of the temporal clip masked (*i.e.*, pixel-wise multiplication) by (i) the ground-truth ($x \cdot y$, where x is the input clip and y is the ground-truth) and (ii) the output of the *segmentor* ($x \cdot S(x)$, where $S(x)$ refers to the output of the *segmentor*). The two masked inputs are processed by the network to obtain two feature vectors. These are compared via Mean Absolute Error (*MAE*), defined as:

$$MAE(C(x \cdot y), C(x \cdot S(x))) = \frac{\sum_{i=1}^{W_d} \sum_{j=1}^{M} |C_j(x_i \cdot y_i) - C_j(x_i \cdot S(x_i))|}{W_d \cdot M} \quad (1)$$

where C is the output of the *critic* (*i.e.*, the feature vector), x_i is the i-th frame of the input clip, y_i is the i-th frame of the ground-truth mask and M is the length of the feature vector. The *MAE* is minimized in an adversarial fashion during training, as explained in Sec. II-C.

C. Training strategy

The training of *NephCNN* is performed with stochastic gradient descend (SGD) as optimizer, to minimize the adversarial loss proposed in [12]. The L sums up the contribution of two loss functions, *i.e.*, the per-pixel Weighted Cross-Entropy Loss (L_{WCE}) from the *segmentor* and the MAE from the *critic*:

$$L = -\frac{1}{\Omega} \sum_{\mathbf{k} \in \Omega} (\beta y_{\mathbf{k}} \cdot log(S(x)_{\mathbf{k}}) + (1 - y_{\mathbf{k}}) \cdot log(1 - S(x)_{\mathbf{k}})) + MAE[C(x \cdot y), C(x \cdot S(x))] \quad (2)$$

where $y_{\mathbf{k}}$ and $S(x)_{\mathbf{k}}$ denote the ground-truth value and the corresponding prediction of the *segmentor* at pixel location $\mathbf{k}$ in the frame domain Ω and β denotes the weight coefficient (set equal to 2).

The (L_{WCE}) proved to provide better performances in learning from class unbalanced datasets [25] (e.g., images that contains structures of interest occupy small part of the overall image).

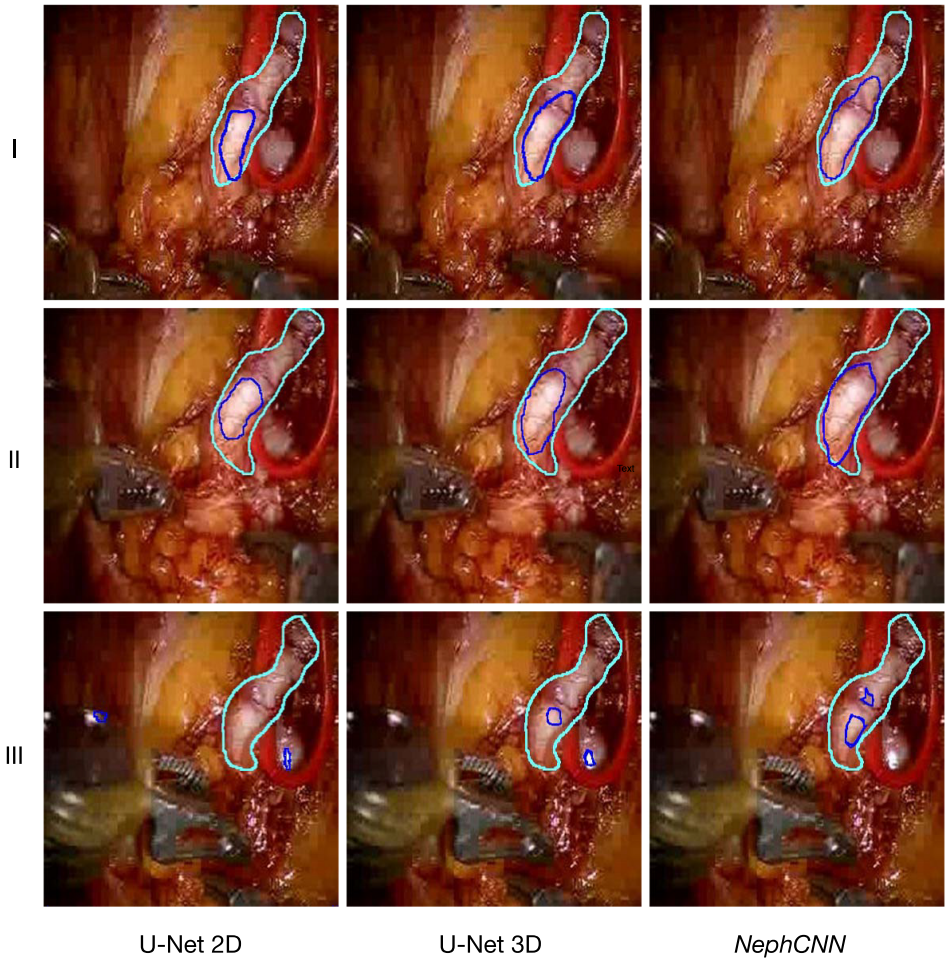


Fig. 4: Examples of the segmentation results on image frames of the test set produced by, respectively: 2D U-Net, 3D U-Net and *NephCNN*. The blue contour represents the segmented part of the renal artery, while the light blue contour represents the ground truth. In row *I* the same test set frame, corresponding to the highest DSC value, is shown for each tested architecture. In row *II* an intermediate case is shown. In row *III* one of the most challenging case is shown: here is evident how the state-of-the-art networks completely failed the vessel segmentation, while the proposed *NephCNN* segmented a portion of the renal artery.

NephCNN was trained for 200 epochs with a batch size of 4 temporal clips and an initial learning rate of 0.0001. The best model was selected as the one minimising the L loss on the augmented validation set.

III. Experimental Setup

A. Dataset

The dataset used to train and test *NephCNN* was selected from the *Nephrec9* dataset [26], which consists of 741573 frames (frame size = 283x218 pixels) extracted from 8 RAPN videos of 8 different patients. The frames are divided according to the RAPN surgical workflow steps. The renal artery was visible only in frames from the dissection step. A total of 240 consecutive frames was selected for each patient, being 240 the minimum number of frames in which the vessel was visible for a patient. Frames with heavy motion blur (due to quick changes in camera position) and with largely occluded vessels were removed from the analysis. Sample frames are shown in Fig. 3. Table I summarizes the number of frames for each patient. The final dataset consisted of 1871 frames.

Since the *Nephrec9* dataset was built for surgical context recognition, manual vessel annotation was performed. Images and corresponding masks were resized to 256x256 pixels for reducing memory constraint and training computational effort.

Six patients (for a total of 1391 frames) were used for training, while one patient were kept as validation set (240 frames) and the remaining patient (240 frames) were used for testing purpose. On-the-fly data augmentation was performed on the training set applying a random set of flipping, zooming ($[-0.1, 0.1]$) and shearing ($[-0.05, 0.05]$) to every epoch.

For the sliding-window algorithm, $W_d = 4$, as a trade off between memory requirements and training speed, and $W_s = 1$ to enlarge the training data size.

B. Ablation study

In this work we performed an ablation study to provide a comprehensive comparison of *NephCNN* architecture with state-of-the-art segmentation FCNNs, performing the following experiments:

a) 2D U-Net: To understand how the temporal information affected the segmentation performance, we implemented a U-Net [24] with 2D kernels as baseline for our comparisons. This network was designed to have the same depth of the *segmentor* network for fair comparison. The network was trained minimizing the loss function reported in Eq. 2, but without the MAE term, (*i.e.*, in a non-adversarial fashion).

b) 3D U-Net: To asses the impact of the adversarial training, we trained a 3D U-Net [13], which is the *segmentor* described in Sec.II-A, without the *critic*.

All experiments were implemented using TensorFlow 2 on an Intel Core i5-8400 CPU with 32GB of RAM and NVIDIA RTX2080TI GPU with 11GB memory.

C. Performance assessment

For performance assessment, we computed, using the test set, the Recall (Rec), defined in Eq. 4, Precision ($Prec$) and Dice Similarity Coefficient (DSC), (Eq. 5).

$$Prec = \frac{TP}{TP + FP} \tag{3}$$

$$Rec = \frac{TP}{TP + FN} \tag{4}$$

$$DSC = \frac{2 \cdot TP}{FP + FN + 2 \cdot TP} \tag{5}$$

where TP (true positive) is the number of vessel pixels correctly classified, FN (false negative) is the number of misclassified vessel pixels and FP (false positive) is the number of misclassified not-vessel pixels.

The Wilcoxon Signed Rank Test was used to test if statistical differences existed both between the DSC and Rec obtained with *NephCNN* and the other tested architectures, considering a significance level (α) equal to 0.05.

IV. Results

The *NephCNN* training lasted about 20 hours. Overall performance metrics for the state-of-the-art networks and the proposed method are illustrated in Table II. Median DSC of the 2D U-Net, the 3D U-Net and the *NephCNN* was 59.70%, 66.33% and 71.76% with Inter-Quartile Range (IQR) of 7.71%, 9.05% and 9.31%, respectively.

The boxplots in Fig. 5 show the performance comparison between 2D U-Net, 3D U-Net and *NephCNN* in terms of DSC. The Wilcoxon signed rank test (significance level α equal to 5%) confirmed that there is a significant statistical difference between the tested architectures.

In Fig. 4 segmentation examples obtained by the tested CNNs are shown, overlaid with the corresponding ground truth.

TABLE II: Evaluation of different frameworks tested on test set images. The average performance, in terms of Dice Similarity Coefficient (*DSC*), Precision (*Prec*) and Recall (*Rec*), with its corresponding standard error is reported. Optimal results are highlighted in bold.

Framework	*DSC*	*Prec*	*Rec*
2D U-Net	0.5970 ± 0.0771	0.8731 ± 0.0626	0.4593 ± 0.0879
3D U-Net	0.6633 ± 0.0905	$\mathbf{0.9377 \pm 0.0774}$	0.4547 ± 0.1113
NephCNN	$\mathbf{0.7176 \pm 0.0931}$	0.8976 ± 0.1132	$\mathbf{0.5066 \pm 0.1377}$

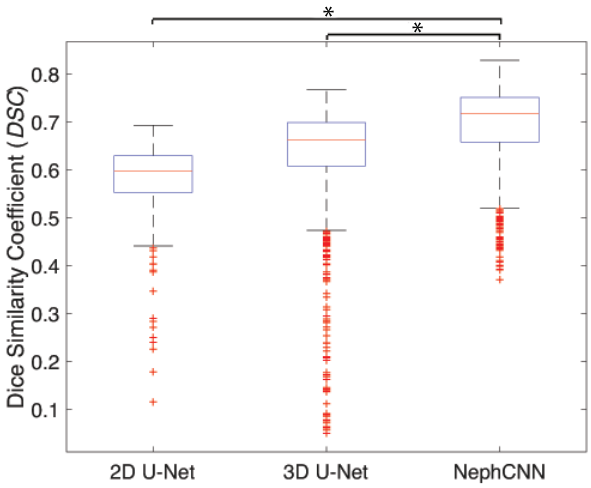


Fig. 5: Boxplots representing segmentation performances for all the analyzed networks, the metric shown is the DSC. Distributions tested with Wilcoxon signed rank are coupled with brackets above each graph, asterisks indicate statistical significance of the difference (* $p < 0.05$).

V. Discussion and Conclusions

This paper represents the first attempt, to the best of our knowledge, that combines shape-constrained adversarial training strategy with spatio-temporal features to perform the automatic segmentation of the renal artery from intraoperative nephrectomy videos.

In order to assess the capability of the proposed network in accomplishing this task, 1871 frames from 8 RAPN videos were extracted and manually annotated.

The results achieved by testing the proposed *NephCNN* showed to successfully increase performances in vessel segmentation with respect to state of the art approaches, providing the highest median DSC value (71.76%) among the ablation studies. These results supported our research hypotheses that shape-constraining and spatio-temporal features can tackle the peculiar challenges, listed in Sec. I and shown in Fig. 3, of nephrectomy images.

Although widely employed for medical tasks, the 2D U-Net showed the lowest performance due to the high complexity and variability of intraoperative laparoscopic images, as already reported in [12]. The introduction of 3D kernels in the U-Net framework, as proposed in [8], enforced the temporal connectivity of features, thanks to the processing of the temporal

information naturally encoded in the laparoscopic videos. In line with [12], further improvements were obtained with the adversarial training, which preserved shape consistency across frames leading to better results.

Specularities on surgical tools and tissues, as shown in Fig. 4 (*III*), hampered the segmentation performance of all the networks. However, the shape constraining effect of adversarial training discouraged the *segmentor* from wrongly identify pixels outside the vessel region.

To improve the segmentation performance in these challenging cases, possible solutions could be provided by exploiting Dense CNNs, which preserve features at different complexity levels, and introducing Atrous Spatial Pyramid Pooling, useful in multi-scale object detection.

Despite our efforts, the complexity of the task and the dataset size remains a strong limitation of our experimental protocol. A more comprehensive validation of the proposed framework will be addressed in the future by investigating extensions of this framework supported by a broader dataset and advanced data augmentation techniques.

To conclude, the encouraging results achieved suggest that the proposed approach could be a first step towards applications able to improve the actual clinical workflow, helping surgeons to reduce intraoperative complications due to unwanted vessel resections and bleeding in robotic-assisted surgery (e.g., automatically implementing virtual fixtures). This method may have a positive impact on RAPN practice, by lowering the surgery duration and surgeons' mental workload and, as a consequence, increasing surgical safety.

REFERENCES

[1] J. Ferlay, E. Steliarova-Foucher, J. Lortet-Tieulent, S. Rosso, J. W. W. Coebergh, H. Comber, D. Forman, and F. Bray, "Cancer incidence and mortality patterns in Europe: Estimates for 40 countries in 2012." *European Journal of Cancer*, vol. 49, no. 6, pp. 1374–1403, April 2013.

[2] J. Ferlay, M. Colombet, I. Soerjomataram, T. Dyba, G. Randi, M. Bettio, A. Gavin, O. Visser, and F. Bray, "Cancer incidence and mortality patterns in Europe: Estimates for 40 countries and 25 major cancers in 2018." *European Journal of Cancer*, vol. 103, pp. 356–387, November 2018.

[3] S. MacLennan, M. Imamura, M. C. Lapitan, M. Imran Omar, T. B. L. Lam, A. M. Hilvano-Cabungcal, P. Royle, F. Stewart, G. MacLennan, S. J. MacLennan, S. E. Canfield, S. McClinton, T. R. Leyshon Griffiths, B. Ljungberg, and J. N'Dow, "Systematic review of oncological outcomes following surgical management of localised renal cancer." *European Urology*, vol. 61, no. 5, pp. 972–993, May 2012.

[4] J. E. Abel, S. H. Culp, M. Meissner, S. F. Matin, P. Tamboli, and C. G. Wood, "Identifying the risk of disease progression after surgery for localized renal cell carcinoma." *BJU International*, vol. 106, no. 9, pp. 1227–1283, January 2010.

[5] G. Novara, S. La Falce, A. Kungulli, G. Gandaglia, V. Ficarra, and A. Mottrie, "Robot-assisted partial nephrectomy." *International Journal of Surgery*, vol. 36, no. Part C, pp. 554–559, December 2016.

[6] L. Maier-Hein, S. S. Vedula, S. Speidel, N. Navab, R. Kikinis, A. Park, M. Eisenmann, H. Feussner, G. Forestier, S. Giannarou *et al.*, "Surgical data science for next-generation interventions," *Nature Biomedical Engineering*, vol. 1, no. 9, pp. 691–696, 2017.

[7] S. Moccia, S. J. Wirkert, H. Kenngott, A. S. Vemuri, M. Apitz, B. Mayer, E. De Momi, L. S. Mattos, and L. Maier-Hein, "Uncertainty-aware organ classification for surgical data science applications in laparoscopy," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 11, pp. 2649–2659, 2018.

[8] E. Colleoni, S. Moccia, X. Du, E. De Momi, and D. Stoyanov, "Deep learning based robotic tool detection and articulation estimation with spatio-temporal layers," *IEEE Robotics and Automation Letters*, vol. 4, no. 3, pp. 2714–2721, 2019.

[9] V. Penza, A. S. Ciullo, S. Moccia, L. S. Mattos, and E. De Momi, "Endoabs dataset: Endoscopic abdominal stereo image dataset for benchmarking 3D stereo reconstruction algorithms," *The International Journal of Medical Robotics and Computer Assisted Surgery*, vol. 14, no. 5, p. e1926, 2018.

[10] B. Gibaud, G. Forestier, C. Feldmann, G. Ferrigno, P. Gonçalves, T. Haidegger, C. Julliard, D. Katić, H. Kenngott, L. Maier-Hein, K. März, E. de Momi, D. Á. Nagy, H. Nakawala, J. Neumann, T. Neumuth, J. R. Balderrama, S. Speidel, M. Wagner, and P. Jannin, "Toward a standard ontology of surgical process models," *International Journal of Computer Assisted Radiology and Surgery*, vol. 13, no. 9, pp. 1397–1408, Jul. 2018.

[11] S. Moccia, S. Foti, A. Routray, F. Prudente, A. Perin, R. F. Sekula, L. S. Mattos, J. R. Balzer, W. Fellows-Mayle, E. De Momi *et al.*, "Toward improving safety in neurosurgery with an active handheld instrument," *Annals of Biomedical Engineering*, vol. 46, no. 10, pp. 1450–1464, 2018.

[12] A. Casella, S. Moccia, E. Frontoni, D. Paladini, E. De Momi, and L. S. Mattos, "Inter-foetus membrane segmentation for TTTS using adversarial networks," *Annals of Biomedical Engineering*, vol. 48, no. 2, pp. 848–859, 2020.

[13] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-net: Learning dense volumetric segmentation from sparse annotation," in *International conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2016, pp. 424–432.

[14] S. Moccia, E. De Momi, S. El Hadji, and L. S. Mattos, "Blood vessel segmentation algorithms—review of methods, datasets and evaluation metrics," *Computer Methods and Programs in Biomedicine*, vol. 158, pp. 71–91, 2018.

[15] V. Penza, X. Du, D. Stoyanov, A. Forgione, L. S. Mattos, and E. De Momi, "Long term safety area tracking (LT-SAT) with online failure detection and recovery for robotic minimally invasive surgery," *Medical Image Analysis*, vol. 45, pp. 13–23, 2018.

[16] S. Wang, Y. Yin, G. Cao, B. Wei, Y. Zheng, and G. Yang, "Hierarchical retinal blood vessel segmentation based on feature and ensemble learning," *Neurocomputing*, vol. 149, pp. 708–717, 2015.

[17] Y. Ganin and V. Lempitsky, "N4-fields: Neural network nearest neighbor fields for image transforms," in *Asian Conference on Computer Vision*. Springer, 2014, pp. 536–551.

[18] Q. Li, B. Feng, L. Xie, P. Liang, H. Zhang, and T. Wang, "A cross-modality learning approach for vessel segmentation in retinal images," *IEEE transactions on medical imaging*, vol. 35, no. 1, pp. 109–118, 2015.

[19] A. Dasgupta and S. Singh, "A fully convolutional neural network based structured prediction approach towards the retinal vessel segmentation," in *IEEE 14th International Symposium on Biomedical Imaging*. IEEE, 2017, pp. 248–251.

[20] J. Mo and L. Zhang, "Multi-level deep supervised networks for retinal vessel segmentation," *International Journal of Computer Assisted Radiology and Surgery*, vol. 12, no. 12, pp. 2181–2193, 2017.

[21] J. Merkow, A. Marsden, D. Kriegman, and Z. Tu, "Dense volume-to-volume vascular boundary detection," in *International conference on medical image computing and computer-assisted intervention*. Springer, 2016, pp. 371–379.

[22] J. Son, S. J. Park, and K. H. Jung, "Retinal vessel segmentation in fundoscopic images with generative adversarial networks." *ArXiv e-prints*, vol. abs/1706.09318, Jun 2017.

[23] A. Lahiri, K. Ayush, P. K. Biswas, and P. Mitra, "Generative adversarial learning for reducing manual annotation in semantic segmentation on large scale miscroscopy images: Automated vessel segmentation in retinal fundus image as test case." *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 794–800, July 2017.

[24] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *Medical Image Computing and Computer-Assisted Intervention*, p. 234–241, 2015.

[25] Y. S. Aurelio, G. M. de Almeida, C. L. de Castro, and A. P. Braga, "Learning from Imbalanced Data Sets with Weighted Cross-Entropy Function," *Neural Processing Letters*, vol. 50, no. 2, pp. 1937–1949, oct 2019.

[26] H. Nakawala, "Nephrec9 (version 0.1)." *Zenodo*, November 2017.