

Weakly supervised segmentation for real-time surgical tool tracking

Eung-Joo Lee^{1,2}, William Plishker², Xinyang Liu³, Shuvra S. Bhattacharyya¹, Raj Shekhar^{2,3} ✉

¹Department of Electrical and Computer Engineering and the Institute for Advanced Computer Studies, University of Maryland, College Park, MD, USA

²IGI Technologies, Inc., College Park, MD, USA

³Sheikh Zayed Institute for Pediatric Surgical Innovation, the Children's National Medical Center, Washington DC, USA

✉ E-mail: rshekhar@childrensnational.org

Published in Healthcare Technology Letters; Received on 18th September 2019; Accepted on 2nd October 2019

Surgical tool tracking has a variety of applications in different surgical scenarios. Electromagnetic (EM) tracking can be utilised for tool tracking, but the accuracy is often limited by magnetic interference. Vision-based methods have also been suggested; however, tracking robustness is limited by specular reflection, occlusions, and blurriness observed in the endoscopic image. Recently, deep learning-based methods have shown competitive performance on segmentation and tracking of surgical tools. The main bottleneck of these methods lies in acquiring a sufficient amount of pixel-wise, annotated training data, which demands substantial labour costs. To tackle this issue, the authors propose a weakly supervised method for surgical tool segmentation and tracking based on hybrid sensor systems. They first generate semantic labellings using EM tracking and laparoscopic image processing concurrently. They then train a light-weight deep segmentation network to obtain a binary segmentation mask that enables tool tracking. To the authors' knowledge, the proposed method is the first to integrate EM tracking and laparoscopic image processing for generation of training labels. They demonstrate that their framework achieves accurate, automatic tool segmentation (i.e. without any manual labelling of the surgical tool to be tracked) and robust tool tracking in laparoscopic image sequences.

1. Introduction: Surgical tool segmentation and tracking of pose (i.e. position and orientation) in the endoscopic camera coordinate system is essential for various surgical visualisation and navigation applications [1]. For instance, tracking the laparoscope and the laparoscopic ultrasound (LUS) transducer enables augmented reality visualisation [2]. Tracking the LUS transducer and the ablation needle helps create a virtual environment to guide needle targeting in laparoscopic ablative therapy [3]. In addition, 2D pose estimation with tool tracking can be utilised to assess the surgical operative skill of minimally invasive or robot-assisted surgical procedures [4].

The use of real-time tracking hardware has been shown to be effective in tracking tools within complex surgical environments. Compared with optical tracking, which has the line-of-sight requirement, electromagnetic (EM) tracking is widely used for tracking flexible tools, such as LUS transducers with articulating imaging tips. In addition, computer-vision-based approaches have been presented for tool segmentation and tracking. Such methods utilise handcrafted features that capture gradient, colour, and texture information to obtain binary segmentation masks [5] and track surgical tools [6]. In vision-based solutions, a fiducial marker on the tool of interest or auxiliary devices, such as a stereo laparoscope, are also used for tool tracking [6, 7].

With the advent of deep convolutional neural networks (DCNNs), deep learning-based approaches have been proposed for surgical tool segmentation and tracking. These approaches have demonstrated promising results when sufficient volumes of training data are available [8, 9]. Transfer learning methods have been shown to provide competitive performance in binary segmentation tasks while using smaller training datasets [10, 11]. However, such transfer learning approaches still require laborious, time-consuming effort to derive training data that is labelled at the pixel-level.

To address this problem, weakly supervised methods using image-level rather than pixel-level annotation have been proposed for localisation and segmentation tasks. These methods use weak supervision of the discriminative regions generated by the classification network [12] and by additional information such as that

provided by kinematic models [13]. A major challenge in this approach is to acquire precise segmentation masks from the feature map, as the discriminative region of the tool of interest is often sparse.

In this work, we present a new weakly supervised framework for segmentation and tracking of surgical tools based on a hybrid sensor system that provides integrated EM tracking and processing of visual data from a laparoscopic imaging system. First, we use both EM and visual data for automatic seed selection. Then we relabel seeds using a feature map created from a DCNN, and apply a Random Walks framework for accurate, binary semantic labelling.

The proposed framework enables fast, robust, and fully automated generation of annotated datasets. This work is significant, as it can (a) generate training data for surgical instruments of any size with an EM tracker setup and (b) alleviate the limitations of EM tracking using a complementary, vision-based tracking method.

The main contributions in this work are two-fold: first, we develop methods to generate semantic labelling of surgical tools without any manual interaction, and secondly, we develop methods for weakly supervised segmentation with a light-weight DCNN for tool tracking. For evaluation, we obtain laparoscopic image datasets using EM-tracked surgical tools over an anatomical phantom. We also acquire in-vivo datasets from an Institutional Animal Care and Use Committee (IACUC)-approved animal study, and then apply our proposed framework to these datasets.

The results of our experiments demonstrate the capability to provide robust surgical tool segmentation and tracking using DCNNs that are configured from automatically-generated training data.

2. Methods: The overall structure of our proposed framework is illustrated in Fig. 1. The subsystem labelled Semantic Labelling is used to generate a labelled dataset that is used to train the DCNN in the subsystem labelled Weakly Supervised Segmentation. The Semantic Labelling subsystem takes as input both EM and laparoscopic image data that has not yet been

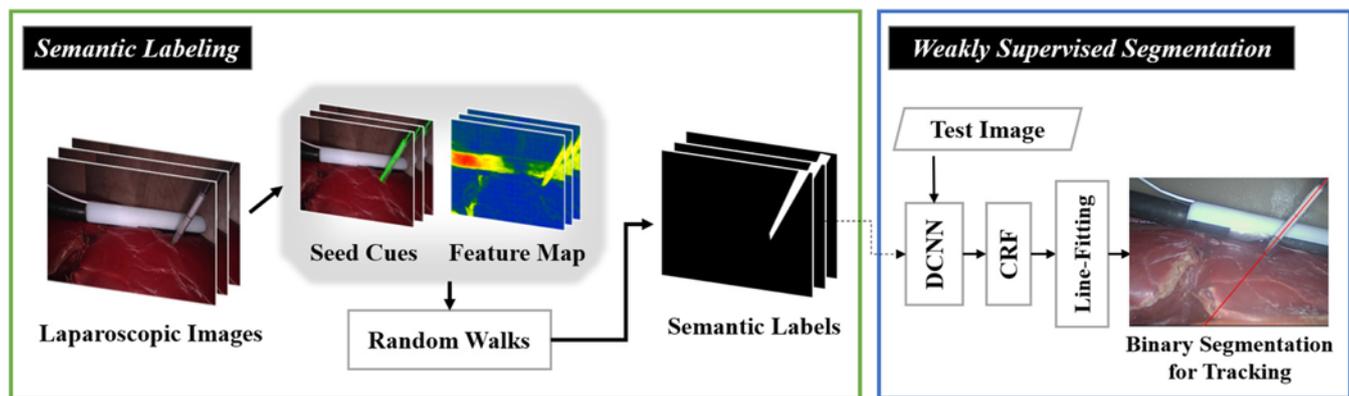


Fig. 1 Overall structure of our proposed framework. Semantic labels generated in the Semantic Labelling subsystem are used to train a DCNN for the Weakly Supervised Segmentation subsystem

labelled. From this multimodal data, localisation is performed to generate ‘seed cues,’ which are used to initialise processes for automatically labelling the laparoscopic image data illustrated. From the input data, the Semantic Labelling subsystem also computes a set of features that provides confidence values of binary classes, foreground and background, for seed refinements. The seed cues and features supply input to a module that applies random walks to generate semantic labels. The output of the Semantic Labelling subsystem is a labelled version of the set I_L of laparoscopic images that were provided as input to the subsystem. Each pixel of each image in I_L is labelled as either foreground (part of a tool) or background.

The labelled version of I_L is used to train the DCNN for the Weakly Supervised Segmentation subsystem (see Fig. 1), which is used for real-time tool tracking. In the overall framework, the Semantic Labelling subsystem enables high-accuracy configuration of the DCNN through a fully-automated process of labelled dataset generation for network training.

2.1. Coarse seed selection using EM tracking: Using EM tracking, we obtain coarse seed cues, which can be viewed as subsets of approximate labels in the image domain. This process is the first phase of two computational phases represented by the block labelled ‘Seed Cues’ in Fig. 1. Fig. 2a shows the setup of the kind of hybrid (bimodal) sensor system that our proposed framework is designed to work with. In our experiments, we used a commercial EM tracking system, called Aurora (developed by NDI Medical), which includes a table top field generator. Custom-designed EM tracking mounts, each containing a six degrees-of-freedom sensor, were attached to the handle of the laparoscope, the imaging tip of the LUS transducer, and the handle of the laparoscopic needle.

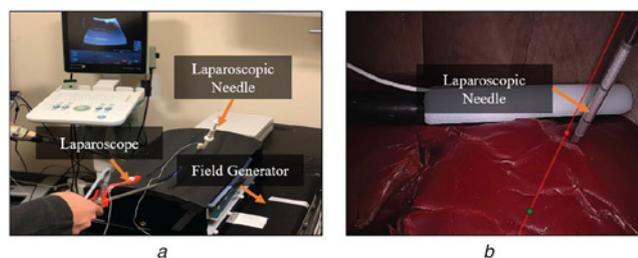


Fig. 2 Coarse seed generation using EM tracking
a Hybrid sensor system setup
b Coarse seed cues derived from EM tracking: tip point (red point) and intersection point (green point)

The diameter of the EM sensor is 1.3 mm, and the sensor on the needle was placed so that its z-axis was in parallel with the needle’s longitudinal axis.

The needle tip location in the sensor coordinate system was obtained using an original equipment manufacturer stylus.

By touching the needle tip with the stylus, the needle tip location can be acquired in the coordinate system of the sensor attached on the needle. A projected needle trajectory was calculated as an extension line from the needle tip along the needle’s longitudinal axis. The intersection point between the needle trajectory and the LUS image plane can be obtained in the EM tracking space. As illustrated in Fig. 2b, the needle tip (red dot), projected needle trajectory (red line) and intersection point (green dot) were overlaid on the laparoscopic video frame through camera calibration.

As can be seen in Fig. 2b, seed cues may deviate from the target object due to errors in EM tracking and calibration. However, two points of coarse seed data can be used to calculate the approximate orientation of the tool in the image space. Such approximate information is useful for refining the coarse seed data into more accurate fine seeds, as described in the following section.

2.2. Fine seed selection using laparoscopic image processing: Fig. 3 depicts fine seed derivation, which is the second phase of Seed Cue

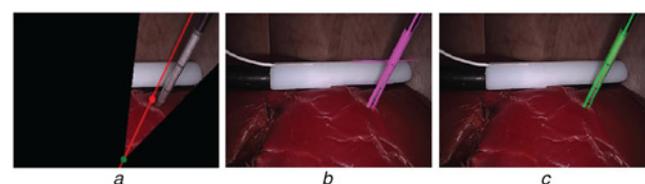


Fig. 3 Fine seed derivation
a ROI generation
b Line detection using the Probabilistic Hough Transform
c Orientation-based line filtering

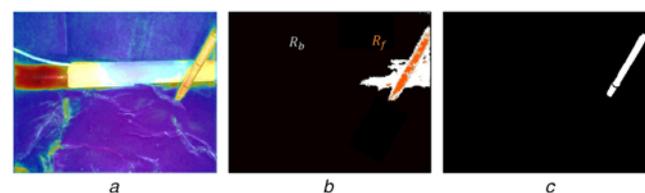


Fig. 4 Semantic labelling using the Random Walks framework
a Feature map overlaid onto the input image
b Seed refinement
c Binary semantic labelling

computation. Using orientation with respect to x -axis determined by the coarse seed points, which we refer to as θ_c , we create a region of interest (ROI) sector that encompasses a candidate region of the surgical tool, as depicted in Fig. 3a.

As θ_c along with the coarse seed points present coarse 2D pose of the surgical tool of interest, we generate the ROI sector by assigning

offset angles to the upper and lower sides with respect to θ_c from the intersection point illustrated as the green dot in Fig. 3a.

Construction of the ROI allows a more reliable and stable localisation of the tool of interest. Since surgical tools are represented as straight line segments in the image, we apply the Probabilistic Hough Transform, a well-known line detection

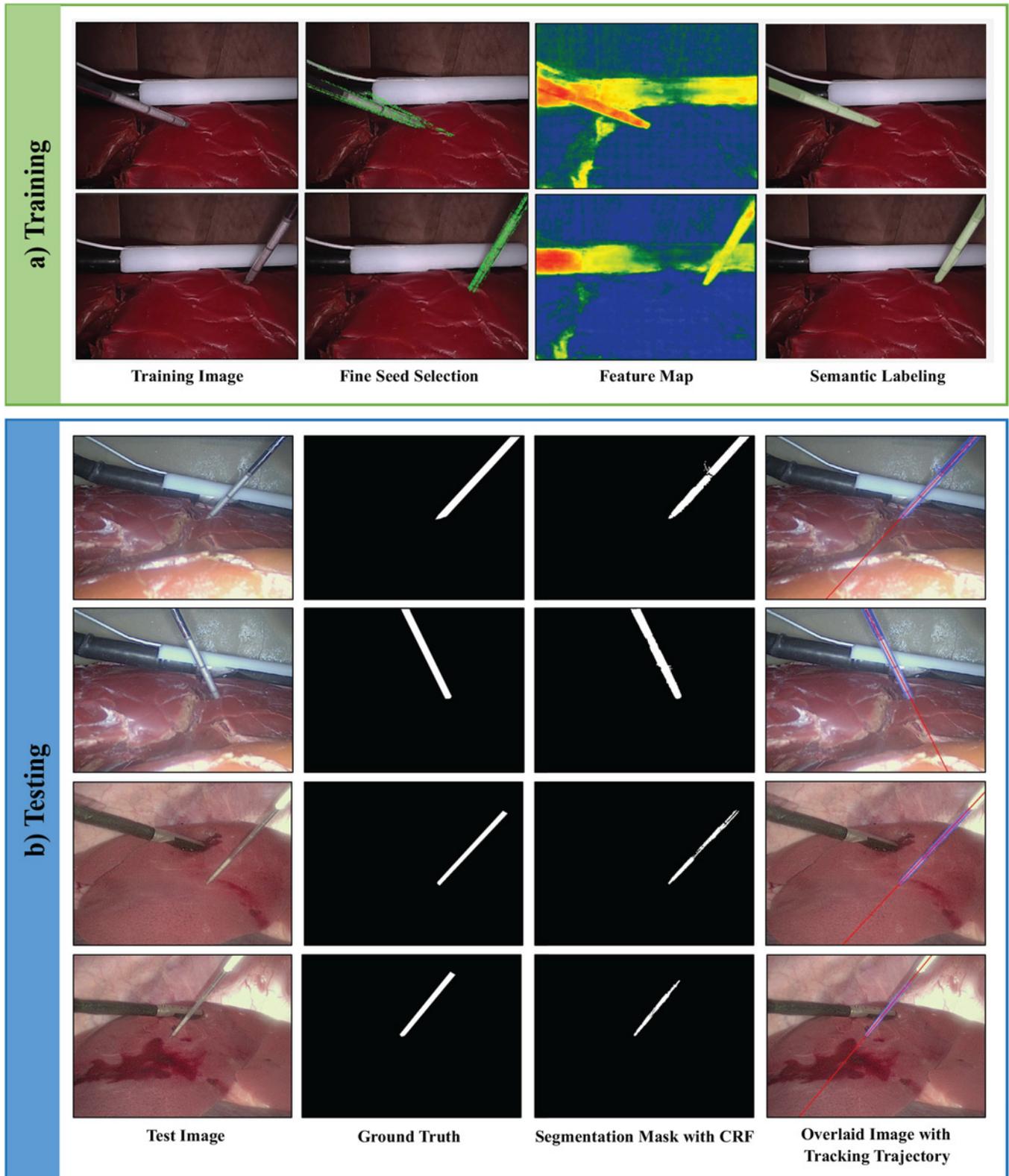


Fig. 5 Illustration of results obtained from
a Semantic labelling from the training dataset

b Binary segmentation on the testing dataset. In the last column of the testing illustration, the red trajectory indicates the pose of tool

method [14], as a basis for fine seed selection. More specifically, we use a bilateral filter for smoothing and enhancement, and then we employ the Canny edge detector followed by the Probabilistic Hough Transform for edge and line detection.

Seed selection based on the applied line detection method, which is sensitive to image noise, lacks accuracy due to complex and redundant features generated by organ texture, miscellaneous objects, and the image background. Thus, for robustness, we group line segments using agglomerative hierarchical clustering and extract desired line segments of the tool according to orientation, θ_c .

This is applied as a post-processing step to line detection. As illustrated in Fig. 3c, this clustering process enhances the precision with which the tool position is computed. The set S of fine seeds derived from clustering consists of the union of the disjoint pixel subsets S_f and S_b , which represent the estimated foreground and background, respectively. Note that the pixels in S_b are all external to the ROI sector.

2.3. Feature-based seed refinement and random walk segmentation: To aid in semantic labelling, we make use of a feature map, as illustrated in Fig. 1. The feature map is created before the Semantic Labelling subsystem is executed.

The feature map is obtained from an segmentation DCNN N_{map} that is trained using fully annotated labels of surgical tools from the Endoscopic Vision challenge of MICCAI 2017 [15]. To generate the feature map, we use a segmentation network for the DCNN N_{map} , which is based on an encoder-decoder architecture.

After training N_{map} , we extract its last layer, which is a softmax layer, as the feature map. The extracted feature map F generates as output a probability value $F(z) \in [0, 1]$ for each pixel z of the input image. The value $F(z)$ represents the probability of the given pixel being a foreground pixel.

The feature map F is used to further improve the accuracy of the fine seed sets S_f and S_b whose derivation was discussed in Section 2.2. Fig. 4 illustrates the pipeline for semantic labelling using the Random Walks framework.

In Fig. 4a, red-yellow pixel regions represent a high probability of the foreground, while blue-green pixel regions represent high probability of the background.

The ROI helps to restrict the region of analysis in the image so that the surgical tool of interest can be determined in a robust manner. The improved seed sets R_f and R_b for foreground and background, respectively, in Fig. 4b are derived by:

$$\begin{aligned} R_f &= \{z | (z \in S_f) \cap (F_z > \theta_f)\}, \\ R_b &= \{z | (z \in S_b) \cup (F_z < \theta_b)\}, \end{aligned} \quad (1)$$

where θ_f and θ_b ($0 \leq \theta_b \leq \theta_f \leq 1$) denote empirically-determined thresholds for the seed refinement process.

In the feature map F , the pixels that have probabilities close to 1 and 0 can be classified as foreground and background seeds, respectively. For the seed refinement process, the thresholds θ_f and θ_b are used to select the foreground and background seeds.

This approach enables the generation of globally distributed seed cues for the background while eliminating mislabelled seeds in the foreground.

Using the relabelled seed sets R_f and R_b , we employ a Random Walks algorithm that labels all of the pixels in a given laparoscopic input image. The Random Walks algorithm that we apply is a graph-based segmentation method that shows robustness to weak boundaries and image noise [16]. We use a Conditional Random Field (CRF) to post-process the result of the Random Walk module so as to refine the boundary, thereby acquiring semantic labelling for training the deep segmentation network.

2.4. Weakly supervised segmentation and tool tracking: Through design and integration of the modules described in Section 2.1 through Section 2.3 for semantic labelling, our framework is capable of automatically generating labelled datasets, which can subsequently be used to train various kinds of DCNNs for high accuracy surgical tool segmentation. As in this work we are specifically interested in real-time tool tracking, we employ an efficient DCNN structure N_{inf} as the inference engine, as illustrated by the block labelled DCNN in the right side of Fig. 1.

We demonstrate our framework using different DCNN structures that are plugged in as N_{inf} . In particular, we use models with light-weight encoder structures, thereby enabling fast processing at inference time.

For the final step of real-time tool tracking, we enhance the output of N_{inf} by postprocessing using a CRF module for boundary delineation. We downsample the input image and the acquired segmentation mask by two to reduce the computational burden for testing. Then, we extract the centre line of the surgical tool by means of OpenCV's fitline function [17], which utilises least squares regression. The resulting straight line is used to estimate the pose of the tool in the input image.

3. Experiments

3.1. Datasets and experimental setup: We collected three frame-sequences using phantom datasets with 500 frames each: two datasets were for training and the other dataset was for testing. We used EM-tracked surgical tools over an anatomical phantom and generated a sufficient number of image sequences that have varying tool positions and orientations. We also acquired a laparoscopic image sequence, consisting of 100 frames, from an IACUC-approved animal study. We used this dataset for testing. The dataset contains complex surgical scenes that have high variability, including effects of complexity and variability due to blood and specular reflection. Each image sequence has a resolution of 1280×1024 pixels.

For the DCNN N_{inf} , we used U-Net [18]; TeraNet-11 [10], which utilises a pre-trained VGG11 network; and two different LinkNet-type [19] DCNN structures: LinkNet-34 and LinkNet-152.

As a loss function, we used binary cross entropy, which is commonly used for binary pixel classification. We used both of these DCNNs separately and averaged the results. For the integration of the feature map, we set $\theta_f = 0.8$ and $\theta_b = 0.4$.

We used the scikit-image package for Random Walks segmentation and OpenCV 3.4 for line-fitting. For training and testing of the segmentation networks, we used a computer equipped with an Intel Core i9-9820X CPU, 64 GB RAM and an 11 GB NVIDIA RTX 2080 Ti GPU.

3.2. Evaluation: To approximate ground truth, we created manually segmented labels by using VGG Image Annotator (VIA) [20] and then calculated both the Jaccard Index and the DICE similarity coefficient (DSC) for quantitative evaluation of semantic labelling and the binary segmented mask.

Table 1 Performance comparison of semantic labelling with different thresholds used for the seed refinement procedures

$\theta_b \setminus \theta_f$	0.6	0.7	0.8	0.9
0.1	64.1	70.3	73.8	68.4
0.2	72.2	75.2	82.1	75.9
0.3	76.3	82.6	86.0	79.5
0.4	82.9	88.1	90.8	85.3

Here, θ_f and θ_b represent thresholds for foreground and background seeds, respectively. The table reports mean DSC values with respect to manually segmented labelling.

Table 2 Quantitative results of the proposed framework for semantic labelling and segmentation tasks based on 200 manually segmented labelling

Segmentation DCNN	Semantic labelling (Training)		Binary segmentation (Phantom)		Binary segmentation (In-vivo)	
	DSC	Jaccard Index	DSC	Jaccard Index	DSC	Jaccard Index
model						
LinkNet-34	89.28% (3.37)	85.62% (4.62)	89.53% (4.20)	86.36% (5.72)	87.45% (5.02)	84.32% (5.75)
LinkNet-152	90.14% (2.21)	88.35% (5.17)	91.46% (3.67)	89.76% (4.83)	88.86% (6.22)	85.37% (6.52)
TernausNet-11	87.05% (3.82)	86.15% (4.92)	87.31% (3.90)	85.61% (4.43)	85.45% (6.21)	83.67% (5.12)
U-Net	75.28% (6.37)	73.47% (8.62)	72.67% (7.21)	69.61% (9.16)	71.45% (8.02)	68.67% (9.30)

The table reports mean values with standard deviations shown in parentheses.

Table 3 Tracking accuracy in degrees acquired from validation datasets for the segmentation task

EM tracking	Proposed tracking			
	LinkNet-34	LinkNet-152	TernausNet-11	U-Net
baseline				
8.39° (2.61)	5.29° (3.22)	4.39° (2.19)	5.68° (2.74)	10.29° (4.23)

The table reports mean values with standard deviations shown in parentheses.

For each task, we created 100 frames of manually segmented labels from phantom datasets and another 100 frames from clinical datasets. We used the resulting set of 200 frames for validation. To derive this set of 200 frames, we manually performed pixel-wise annotation, as there is no publicly available ground truth data for surgical instruments that have been used for training. Using such manually-derived labels, we employed the Jaccard index as an evaluation metric. This metric is a similarity measure of overlap that is used to assess the accuracy of segmentation results. The metric is defined as:

$$\text{Jaccard Index}(A, B) = \frac{|A \cap B|}{|A \cup B|}, \quad (2)$$

where A represents the set of foreground pixels in ground truth, B represents the set of foreground pixels in the predicted segmentation mask, and $|S|$ indicates the cardinality of the set S . We also used the DSC, which is commonly used as a quantitative metric for segmentation performance. This metric can be expressed as:

$$\text{DSC}(A, B) = \frac{2|A \cap B|}{|A| + |B|}. \quad (3)$$

We report tracking accuracy by obtaining the orientation of lines fitted to the manually segmented labels and the binary segmented masks. For orientation accuracy, we calculate the angle at the intersection of the two lines (e.g. an angle of 0° indicates a perfect orientation). In addition, to establish a baseline, we compare the tracking accuracy of the EM-based tracking system relative to the line fitted to the manually segmented result.

3.3. Experimental results: Fig. 5 illustrates results obtained by the proposed framework for semantic labelling and binary segmentation. Fig. 5b shows test images; the first two are from phantom datasets, while the others, which contain occlusion, blood, and light reflection, are from in-vivo datasets. As presented in Fig. 5, segmentation using generated semantic labelling, followed by CRF post-processing, has the capability to delineate the boundary even in surgical scenes, and this capability can be used for pose estimation with the line-fitting method.

Table 1 presents a performance comparison involving semantic labelling using threshold selection for seed refinement. The evaluation is performed using the ground truth data available for the semantic labelling task. Here, θ_f and θ_b represent thresholds for the foreground and background seeds, respectively, that were discussed in Section 2.3. As presented in Table 1, higher values

of θ_b provide higher average DSC values. We anticipate that this is because higher values of θ_b help to remove background features from the feature map. On the other hand, high θ_f values place greater emphasis on localisation cues for the foreground of the surgical tool of interest. The threshold in bold, which produced the highest mean DSC value, was selected for the seed refinement process.

Table 2 summarises quantitative results derived from our experiments on semantic labelling and segmentation using segmentation DCNN models. Owing to the higher level of complexity and blurriness observed in the images, the accuracy of segmentation obtained from in-vivo datasets is lower than that from anatomical phantom datasets.

Table 3 presents the tracking accuracy, reported as an angular error in degrees, of the proposed method. The validation datasets for the segmentation task are used. These results indicate that the proposed approach with the LinkNet-type and TernausNet segmentation models have significantly improved tracking accuracy relative to EM-based tracking alone.

4. Conclusion: In this paper, we developed a new method, using data from both EM tracking and laparoscopic imaging, which generates semantic labelling of surgical tools without any human intervention. Using labelled data generated from this method, we developed a system for real-time tool tracking based on weakly supervised segmentation with a light-weight, DCNN.

This work could be generalised for multiple tools tracking in two manners. By collecting a set of training samples of multi-tools based on the proposed labelling procedures using the EM sensor, we could perform multi-class segmentation which enables segmentation of different tools. Using a set of training samples, we could also construct multiple segmentation DCNNs and obtain segmentation of multiple tools independently in each frame.

Our method for semantic labelling addressed a major bottleneck in the development of high accuracy tool tracking systems, which is that of providing sufficient labelled data for the DCNN training. We have demonstrated the accuracy of our proposed methods using a relevant manually segmented dataset, and two different DCNN structures that were trained using the automatically-generated training data.

5. Funding and Declaration of Interests: The work described in this paper was funded by the National Institutes of Health Grant 2R42CA192504. Conflict of interest: None declared.

6 References

- [1] Bouget D., Allan M., Stoyanov D., *ET AL.*: 'Vision-based and markerless surgical tool detection and tracking: a review of the literature', *Med. Image Anal.*, 2017, **35**, pp. 633–654
- [2] Liu X., Kang S., Plishker W., *ET AL.*: 'Laparoscopic stereoscopic augmented reality: toward a clinically viable electromagnetic tracking solution', *J. Med. Imaging*, 2016, **3**, (4), p. 045001
- [3] Sastry A.V., Swet J.H., Murphy K.J., *ET AL.*: 'A novel 3-dimensional electromagnetic guidance system increases intraoperative microwave antenna placement accuracy', *HPB. (Oxford)*, 2017, **19**, (12), pp. 1066–1073
- [4] Jin A., Yeung S., Jopling J., *ET AL.*: 'Tool detection and operative skill assessment in surgical videos using region-based convolutional neural networks'. 2018 IEEE Winter Conf. on Applications of Computer Vision (WACV), Lake Tahoe, Nevada, USA, 2018, pp. 691–699
- [5] Speidel S., Benzko J., Krappé S., *ET AL.*: 'Automatic classification of minimally invasive instruments based on endoscopic image sequences'. Proc. SPIE, Medical Imaging 2009: Visualization, Image-Guided Procedures, and Modeling Int. Society for Optics and Photonics, Orlando, FL, USA, 2009, vol. 7261, p. 72610A
- [6] Allan M., Thompson S., Clarkson M.J., *ET AL.*: '2D-3D pose tracking of rigid instruments in minimally invasive surgery'. Int. Conf. on Information Processing in Computer-assisted Interventions, Fukuoka, Japan, 2014, pp. 1–10
- [7] Zhao T., Zhao W., Halabe D.J., *ET AL.*: 'Fiducial marker design and detection for locating surgical instrument in images', Google Patents, 2016. US Patent 9,526,587
- [8] García-Peraza-Herrera L.C., Li W., Fidon L., *ET AL.*: 'Toolnet: holistically-nested real-time segmentation of robotic surgical tools'. 2017 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS), Vancouver, Canada, 2017, pp. 5717–5722
- [9] Twinanda A.P., Shehata S., Mutter D., *ET AL.*: 'Endonet: a deep architecture for recognition tasks on laparoscopic videos', *IEEE Trans. Med. Imaging*, 2017, **36**, (1), pp. 86–97
- [10] Shvets A.A., Rakhlin A., Kalinin A.A., *ET AL.*: 'Automatic instrument segmentation in robot-assisted surgery using deep learning'. 2018 17th IEEE Int. Conf. on Machine Learning and Applications (ICMLA). IEEE, Orlando, Florida, USA, 2018, pp. 624–628
- [11] Lee E.J., Plishker W., Liu X., *ET AL.*: 'Segmentation of surgical instruments in laparoscopic videos: training dataset generation and deep-learning-based framework'. Proc. SPIE Medical Imaging, 2019, San Diego, CA, USA, 2019, vol. 10951, p. 461
- [12] Vardazaryan A., Mutter D., Marescaux J., *ET AL.*: 'Weakly-supervised learning for tool localization in laparoscopic videos'. Intravascular Imaging and Computer Assisted Stenting and Large-Scale Annotation of Biomedical Data and Expert Label Synthesis, 2018, pp. 169–179
- [13] Rocha C.D.C., Padoy N., Rosa B.: 'Self-supervised surgical tool segmentation using kinematic information', arXiv preprint arXiv:190204810, 2019
- [14] Kiryati N., Eldar Y., Bruckstein A.M.: 'A probabilistic Hough transform', *Pattern Recognit.*, 1991, **24**, (4), pp. 303–316
- [15] Allan M., Shvets A., Kurmann T., *ET AL.*: '2017 robotic instrument segmentation challenge', arXiv preprint arXiv:190206426, 2019
- [16] Grady L.: 'Random walks for image segmentation', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2006, **28**, (11), pp. 1768–1783
- [17] Bradski G.: 'The OpenCV library', Dr Dobb's Journal of Software Tools, 2000
- [18] Ronneberger O., Fischer P., Brox T.: 'U-net: convolutional networks for biomedical image segmentation'. Int. Conf. on Medical image computing and computer-assisted intervention, Munich, Germany, 2015, pp. 234–241
- [19] Chaurasia A., Culurciello E.: 'Linknet: exploiting encoder representations for efficient semantic segmentation'. 2017 IEEE Visual Communications and Image Processing (VCIP). IEEE, 2017, pp. 1–4
- [20] Dutta A., Zisserman A.: 'The VIA annotation software for images, audio and video', arXiv preprint arXiv:190410699, 2019