Robotic and Neurosurgical Instrument Segmentation for Development of Intelligent Surgical Assistant

Niveditha Kalavakonda
Blake Hannaford
nkalavak@uw.edu
blake@uw.edu
University of Washington
Seattle, WA

Zeeshan Qazi
Laligam Sekhar
zeeshan@uw.edu
lsekhar@neurosurgery.washington.edu
Harborview Medical Center
Seattle, WA

ABSTRACT

Technology for Robot-assisted Neurosurgery is currently limited in application. Neurosurgery procedures require a surgeon and a surgical assistant who assists with tasks. A surgical assistant with robotic precision and awareness of a human will result in effective treatment of patients. As a step towards the development of an Intelligent Surgical Assistant, surgical instruments need to be tracked to understand the response of a human surgical assistant to verbal and nonverbal cues from surgeons. For this purpose, an instrument tracking algorithm was developed and tested using two datasets: a labeled dataset from Intuitive Surgical, Inc. from MICCAI 2017 for the Endoscopic Vision Challenge and a custom dataset developed and released with this paper, called NeuroID [1]. In this paper, three different instrument segmentation approached are evaluated and compared to a hand-crafted heuristic baseline. The source code of our methods is also made publicly available to facilitate reproducibility.

KEYWORDS

context-aware surgical systems, collaborative surgery, datasets, instrument segmentation, microsurgery, MICCAI 2017, neurosurgery, NeuroID, robotic surgery

1 INTRODUCTION

Robotic surgery is being adopted more commonly by surgeons and its use in hospitals is growing at a steady rate[10]. The most commonly performed robotic procedure is Minimally Invasive Surgery for laparoscopic applications. The da Vinci Surgical System is a laparoscopic robot that has seen a steady growth and adoption, with over 700,000 surgeries performed in 2015 using the system with a 15% increase in 2016[1]. The fields of orthopedics, gynecology, cardiac surgery and retinal surgery are some others that benefit from the surgical application of robotics.

On the contrary, robotics is not commonly employed in the field of neurosurgery due to the complexity of procedures. Neurosurgery is a field that would benefit most from robotic surgery due to improved precision, consistency in surgical technique, enhanced safety, and their minimally invasive nature. [12] provides a consolidated report on the innovations in robotics for endovascular and cerebrovascular neurosurgery. The applications are mostly for imaging-related modalities, guidance systems or master-slave type control by the surgeon. Depending on the level of autonomy in operation, the surgical robotic systems can be classified into three categories: supervisory controlled robots, teleoperated robots and shared control type systems [13][4]. The work done by collaborative mechanisms depict the reliability of surgical robots for less critical tasks while relying on the surgeon to be the dominant operator. This boosts the procedure with robotic speed and precision while circumventing the problem of complete autonomy and decision making[7] [11].

[1] http://brl.ee.washington.edu/robotics/surgical-robotics/neurosurgical-instrument-segmentation/

2 PROBLEM MOTIVATION

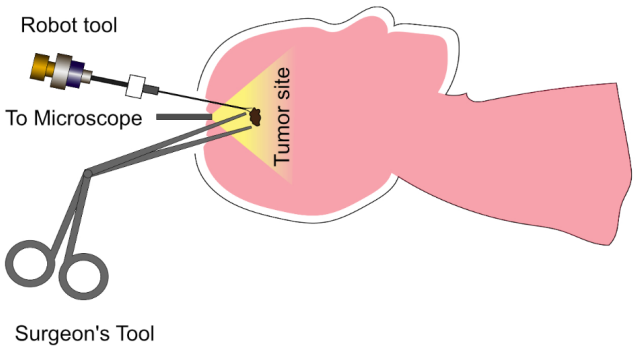

Figure 1: Proposed human-robot collaboration in neurosurgical workspace.

Our goal is to develop an assistive robot for neurosurgery that would execute three important but simple tasks that a surgical assistant would perform. A sample setup of the desired system is shown in Fig. 1. The first step in the direction of developing a human-robot collaborative surgical assistant is to model the surgical protocol for neurosurgery. This requires the system to identify and analyze how different surgical instruments move in a given surgical workspace. This project identifies and segments different surgical instruments present in image frames of an in-vivo surgical procedure to achieve this goal.

3 BACKGROUND AND RELATED WORK

Instrument tracking has been a topic of interest over the last decade in the field of surgery. In addition to benefiting collaborative robotics, the position information of instruments helps in increasing the context-awareness of the surgeon and helps reduce human errors.

There are two categories of approach for instrument detection - marker-based and marker-less. Marker-based approached would add to additional costs for manufacturing and sterilization. It also adds to the complexity of a procedure by requiring constant calibration for trackers to obtain position information relative to the global coordinate frame. Scalability of the solution is hence more applicable to marker-less solutions.

Bouget et. al [8] detect surgical tools by learning the local appearance and global shape from the training data. A random forest model is trained over ten feature channels and the shape model is learned using linear Support Vector Machine. The parameters are searched exhaustively and hence slow for runtime implementation. This work is also one of two cases that develops a neurosurgical dataset for identifying instruments. Recent research is more oriented towards using machine learning and deep learning for surgical tool segmentation. [5] uses Convolutional Neural Networks with auto encoder-decoder architectures. Pakhomov et. al [14] modified a deep residual learning network `ResNet-101` and used dilated convolutions with stride to perform a binary segmentation task for differentiating surgical instruments and tissue. The output image dimensions are restored by incorporating a deconvolutional layer. Both these methods were trained on datasets with pixel-wise segmentation, restricted to the field of Minimally Invasive Surgery. These require significant memory and hardware for implementation due to the size of the networks. The other challenge for using deep learning on current datasets is the large number of parameters in the network used in proportion with the smaller size of the dataset, which may lead to overfitting.

4 APPROACH AND UNIQUENESS

The current collaborative robotic assistants for surgery alternate with the surgeon to perform tasks and do not work beside them in the surgical workspace. To the best of our knowledge, [6] is the only project in the literature to develop a surgical robot working alongside a surgeon, for an invasive laparoscopic procedure. Research in the field works on automating the surgical procedure by replacing the surgeon while ignoring potential benefits of automating tasks performed by the surgical assistant. Effective implementation of such systems will need three important attributes: modeling the surgical protocol, a mode of communication between surgeon and robotic assistant, and autonomous navigation of surgical instruments to perform assistive tasks. We are addressing the first stage of the project in this paper.

This work compares four different marker-less methods for binary segmentation of neurosurgical instruments towards identifying and tracking tools used by surgeons and surgical assistants. We also generated a new, labeled dataset for neurosurgical instrument segmentation, which will be made publicly available with the code. Our approaches were also evaluated on a public robotic instrument dataset for comparison.

4.1 Dataset Generation

The datasets for instrument segmentation need to account for the variability in challenges posed. These include heterogeneous environments with diversity in optical interactions such as blur, shadows, occlusions, specular reflection, fast motion, smoke, and blood. For the Intelligent Surgical Assistant, there are two datasets that the algorithm is tested on to account for the differences between robotic instruments and instruments used during neurosurgery. For robotic instruments, the MICCAI 2017 Endoscopic Vision Challenge Instrument Segmentation [3] dataset was used.

The instruments used in robotic surgery are different from those in a traditional surgical procedure. These instruments do not possess the distinct features of a robotic instrument, which usually consists of shaft, wrist and clasper parts. Due to the lack of a standard dataset for neurosurgical tools, a dataset was generated from surgical videos recorded in Harborview Medical Center, where the Intelligent Surgical Assistant will be tested on completion. The procedures selected for data collection have a higher level of assistant involvement, using up to five instruments simultaneously in the surgical field.

Each video has a resolution of 720 x 480 px and runs at 29.97 frames per second. The images for the surgical tools dataset were collected every 14th frame to record information at an approximate rate of 2 frames per second. We annotated 300 images from each video, of which the first 225 were used for training and the rest were used for testing. Additionally, 727 images from across the videos were annotated to increase the diversity of visual conditions. These were split randomly into training (581 images) and testing (146 images) sets. The target distribution in the training and testing sets were evaluated to be similar. This reduces overfitting to a particular surgery or set of images. The total dataset of 2227 images provides pixel-wise labels for binary and instance segmentation (i.e. instrument type ID) tasks. A sample annotated image is shown in 2.

4.2 Tool Segmentation

In this work, three different deep architectures for binary segmentation were evaluated against a hand-crafted heuristic procedure (Described in our publication [2]). The first architecture is a Vanilla U-net (composed of an encoder-decoder structure)[15] for the tool vs non-tool identification task. The encoder network consists of successive convolutional layers, pooling layers and Rectified Linear Unit (ReLU) activations, capturing a compact feature map in an encoded latent representation. The pooling layers from the encoder are replaced with upsampling layers in the symmetric decoder network for recovering spatial information. The concatenation of higher resolution features from the downsampled path with the features in the upsampled section (Fig. 3) provides precise localization. The performance of the default encoder and decoder were rather limited. By leveraging the benefits of transfer learning [19] we evaluated two variations in U-net – one with a VGG16 encoder network[17] and another using the lighter, more versatile MobileNetV2[16]. The U-net with VGG16 network consists of 3x3 kernels and uses 1x1 convolutions intermittently to change the dimensionality in the filter space whereas the MobileNetV2 network presents an inverted residual structure.

Training:

The VGG-UNet and MobileUNet networks were pre-trained on the ImageNet dataset[9]. For upsampling in the decoder network, we used bilinear interpolation for the Vanilla U-net and fractionally

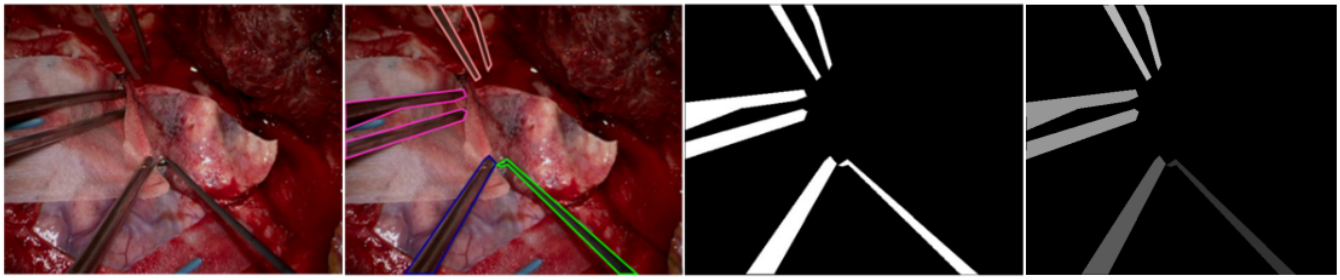

Figure 2: Manual annotation of a collected image frame to detect and identify different Neurosurgical instruments (grasper, peach), (suction, blue), (curette, green) and (pickup, magenta).The binary segmentation ground truth and instance segmentation ground truth from the annotation are included.

strided convolutions/transposed convolution with the for VGG-UNet network.

MobileNetv2 uses depthwise separable convolutions, linear bottleneck between layers and shortcut connections between the bottlenecks. It uses fewer parameters and runs faster in comparison to VGG-16, without a large loss in accuracy. The obtained pre-trained weights were initially trained for faster inference speeds as opposed to higher accuracy. This initially resulted in relatively lower performance. We updated our decoder for MobileUNet based on results from [18] to create a computationally efficient solution. This uses a data-dependent upsampling technique called DUpsampling that incorporates better feature aggregation and downsamples the fused features to the lowest resolution before merging them.

The energy function is minimized to maximize probability of accurate predictions. We perform K-fold cross-validation to avoid overfitting. The dataset was augmented using horizontal flip, vertical flip, normalize, padding and random crop to increase its size and learn invariance properties. A threshold of 0.3 was used to determine if the pixel was tool or background. All pixel values below the specified threshold were set to 0, while all values above the threshold were set to 255 to produce final prediction mask. The networks were trained for 10 epochs, with a batch size of 4. The networks used Adam optimizer with an *alpha* of 0.0001.

5 RESULTS

On applying the instrument segmentation algorithm detailed above, it was possible to segment instruments from tissue. A sample output image is shown in Figure 2.

To show the relative performance of the neural networks with respect to the hand-crafted heuristic baseline[2], the Dice coefficient and Intersection over Union (IoU) metrics were used to calculate quality of binary segmentation (Table 1). By incorporating a lighter network and modifying the downsampling technique, we were able to improve performance while generating a faster and lighter network. The performance variations between the two datasets stem from the level of variations in the chosen surgical scenarios (more in NeuroID) and relatively smaller dataset used from NeuroID. We will incorporate the larger dataset in the next training iteration.

The deep-learning methods used were in an end-to-end pipeline, performing efficient analysis on the full resolution images. Post-processing techniques such as conditional random fields or grabcut could be applied to further improve performance. The results can be extended with a categorical cross-entropy (H) for instrument identification.

Table 1: Evaluation of performance - Dice Coefficient and IoU

Analysis	Robotic Instruments		Neurosurgical Instruments	
	Dice	IoU	Dice	IoU
Baseline Method	0.516	0.461	0.339	0.312
Vanilla U-net	0.813	0.724	0.6740	0.653
U-net w/ VGG-16	0.887	0.80	0.736	0.7102
MobileUNet	0.921	0.882	0.769	0.748

On the data collection front, obtaining data for neurosurgical procedures was difficult due to personnel availability, mechanics of obtaining patient consent, and fewer surgeries actively involving tools held by a surgical assistant. Additionally, due to idiosyncratic depth of tumor (between 1 to 5cm) and its location in the frame, the level of focus blur and extent of blurred area was also highly variable among the videos. We will be increasing the number of videos and size of the dataset.

Table 2: Evaluation of performance with NeuroID- (Inference on NVIDIA Titan Xp)

Analysis	Inference Time (in sec)
Heuristic method	0.271
Vanilla U-net	0.089
U-net w/ VGG-16	0.176
Mobile-UNet	0.026

We were also able to evaluate the inference time for the different networks. This was performed for neurosurgical instrument segmentation. The results are shown in Table 2..

6 ACKNOWLEDGMENTS

We would like to thank the Amazon Catalyst Grant for funding this project. We also acknowledge the helpful contribution from the NVIDIA GPU Grant.

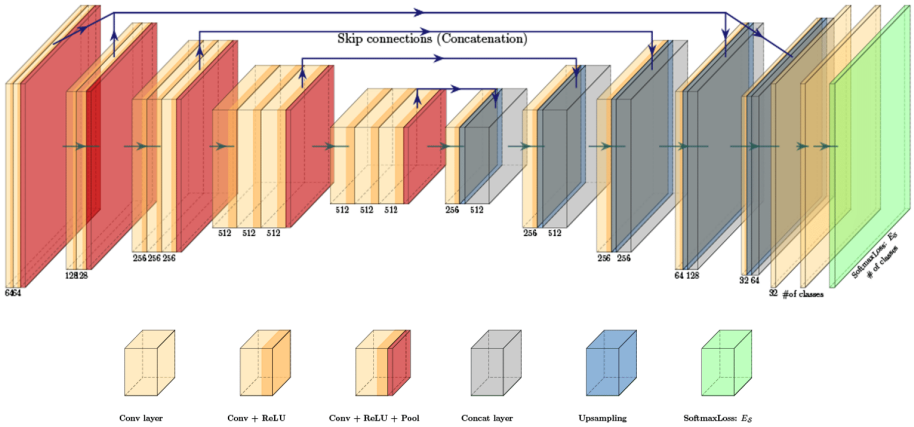

Figure 3: Encoder-Decoder architecture used for instrument segmentation. The number of channels is indicated below the box for a VGG encoder. The height of the box represents a feature map resolution.The yellow boxes represent convolution blocks, red represents pooling, blue represents upsampling and gray represents concatenation.

REFERENCES

[1] [n. d.]. Intuitive Surgical Annual Report 2017. http://www.annualreports.com/HostedData/AnnualReports/PDF/NASDAQ_ISRG_2017.pdf. Accessed: 2018-03-05.

[2] [n. d.]. Robotic Instrument Segmentation Sub-Challenge. https://endovissub2017-roboticinstrumentsegmentation.grand-challenge.org/. Accessed: 2017-07-02.

[3] Max Allan, Alexey Shvets, Thomas Kurmann, Zichen Zhang, Rahul Duggal, Yun-Hsuan Su, Nicola Rieke, Iro Laina, Niveditha Kalavakonda, Sebastian Bodenstedt, Luis Herrera, Wenqi Li, Vladimir I. Iglovikov, Huoling Luo, Jian Yang, Danail Stoyanov, Lena Maier-Hein, Stefanie Speidel, and Mahdi Azizian. 2019. 2017 Robotic Instrument Segmentation Challenge. *CoRR* abs/1902.06426 (2019).

[4] and J. T. Wen. 2000. Autonomous suturing using minimally invasive surgical robots. In *Proceedings of the 2000. IEEE International Conference on Control Applications. Conference Proceedings (Cat. No.00CH37162)*. 742–747. https://doi.org/10.1109/CCA.2000.897526

[5] M. Attia, M. Hossny, S. Nahavandi, and H. Asadi. 2017. Surgical tool segmentation using a hybrid deep CNN-RNN auto encoder-decoder. In *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. 3373–3378. https://doi.org/10.1109/SMC.2017.8123151

[6] E. Bauzano, B. Estebanez, I. Garcia-Morales, and V. F. MuÃśoz. 2016. Collaborative HumanâĂŞRobot System for HALS Suture Procedures. *IEEE Systems Journal* 10, 3 (Sep. 2016), 957–966. https://doi.org/10.1109/JSYST.2014.2299559

[7] P. Berthet-Rayne, M. Power, H. King, and G. Yang. 2016. Hubot: A three state Human-Robot collaborative framework for bimanual surgical tasks based on learned models. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*. 715–722. https://doi.org/10.1109/ICRA.2016.7487198

[8] David Bouget, Rodrigo Benenson, Mohamed Omran, Laurent Riffaud, Bernt Schiele, and Pierre Jannin. 2015. Detecting Surgical Tools by Modelling Local Appearance and Global Shape. *IEEE Transactions on Medical Imaging* 34, 12 (dec 2015), 2603–2617. https://doi.org/10.1109/tmi.2015.2450831

[9] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. 2009. ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR09*.

[10] R. Elek, T. DÃąniel Nagy, D. ÃĄ. Nagy, G. Kronreif, I. J. Rudas, and T. Haidegger. 2016. Recent trends in automating robotic surgery. In *2016 IEEE 20th Jubilee International Conference on Intelligent Engineering Systems (INES)*. 27–32. https://doi.org/10.1109/INES.2016.7555144

[11] K. E. Kaplan, K. A. Nichols, and A. M. Okamura. 2016. Toward human-robot collaboration in surgery: Performance assessment of human and robotic agents in an inclusion segmentation task. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*. 723–729. https://doi.org/10.1109/ICRA.2016.7487199

[12] Simon A Menaker, Sumedh S Shah, Brian M Snelling, Samir Sur, Robert M Starke, and Eric C Peterson. 2018. Current applications and future perspectives of robotics in cerebrovascular and endovascular neurosurgery. *Journal of NeuroInterventional Surgery* 10, 1 (2018), 78–82. https://doi.org/10.1136/neurintsurg-2017-013284 arXiv:https://jnis.bmj.com/content/10/1/78.full.pdf

[13] N. Padoy and G. D. Hager. 2011. Human-Machine Collaborative surgery using learned models. In *2011 IEEE International Conference on Robotics and Automation*. 5285–5292. https://doi.org/10.1109/ICRA.2011.5980250

[14] Daniil Pakhomov, Vittal Premachandran, Max Allan, Mahdi Azizian, and Nassir Navab. 2017. Deep Residual Learning for Instrument Segmentation in Robotic Surgery. *CoRR* abs/1703.08580 (2017).

[15] O. Ronneberger, P.Fischer, and T. Brox. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI) (LNCS)*, Vol. 9351. Springer, 234–241. http://lmb.informatik.uni-freiburg.de/Publications/2015/RFB15a (available on arXiv:1505.04597 [cs.CV]).

[16] Mark B. Sandler, Andrew G. Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. 2018. MobileNetV2: Inverted Residuals and Linear Bottlenecks. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2018), 4510–4520.

[17] Karen Simonyan and Andrew Zisserman. 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition. *CoRR* abs/1409.1556 (2014). arXiv:1409.1556 http://arxiv.org/abs/1409.1556

[18] Zhi Tian, Tong He, Chunhua Shen, and Youliang Yan. 2019. Decoders Matter for Semantic Segmentation: Data-Dependent Decoding Enables Flexible Feature Aggregation.

[19] Jason Yosinski, Jeff Clune, Yoshua Bengio, and Hod Lipson. 2014. How transferable are features in deep neural networks? *CoRR* abs/1411.1792 (2014). arXiv:1411.1792 http://arxiv.org/abs/1411.1792