2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Ultrasound Probe Pose Classification for Task Recognition in Central Venous Catheterization

C. Barr[1], R. Hisey[1], T. Ungi[1] and G. Fichtinger[1], *Fellow, IEEE*

Abstract—**Central Line Tutor is a system that facilitates real-time feedback during training for central venous catheterization. One limitation of Central Line Tutor is its reliance on expensive, cumbersome electromagnetic tracking to facilitate various training aids, including ultrasound task identification and segmentation of neck vasculature. The purpose of this study is to validate deep learning methods for vessel segmentation and ultrasound pose classification in order to mitigate the system's reliance on electromagnetic tracking. A large dataset of segmented and classified ultrasound images was generated from participant data captured using Central Line Tutor. A U-Net architecture was used to perform vessel segmentation, while a shallow Convolutional Neural Network (CNN) architecture was designed to classify the pose of the ultrasound probe. A second classifier architecture was also tested that used the U-Net output as the CNN input. The mean testing set Intersect over Union score for U-Net cross-validation was 0.746 ± 0.052. The mean test set classification accuracy for the CNN was 92.0% ± 3.0, while the U-Net + CNN achieved 92.7% ± 2.1%. This study highlights the potential for deep learning on ultrasound images to replace the current electromagnetic tracking-based methods for vessel segmentation and ultrasound pose classification, and represents an important step towards removing the electromagnetic tracker altogether. Removing the need for an external tracking system would significantly reduce the cost of Central Line Tutor and make it far more accessible to the medical trainees that would benefit from it most.**

I. Introduction

A medical student's path from learning to mastering a new skill is one paved by deliberate practice, expert guidance and real-world experience. The high stakes nature of medicine also necessitates that students reach a minimum level of competency before practicing on patients. Raising students to this level via medical simulation has gained traction in recent years, particularly as the technologies to support these systems have become cheaper and more widely available [1]. Simulation has the advantage of mitigating risk to patients while also improving patient outcomes and student confidence in clinical scenarios [2]. Feedback from expert instructors and clinicians is of equal importance to practicing a new skill [3]. Self-assessment among medical students is a poor means of evaluating performance and informing future practice [4], and as a result providing meaningful feedback in a simulation environment is critical to trainee learning outcomes.

Central venous catheterization (CVC) is a clinical skill taught in numerous residency programs and involves the cannulation of a major vessel for high-throughput venous access. With a long-term complication rate of more than 15% and over 5 million performed in the United States each year, the morbidity associated with CVC is substantial [5]. Furthermore, the risk of complications associated with CVC is up to 35% higher when the procedure is performed by a novice, highlighting the significant learning curve associated with correctly performing the procedure [6]. To curb this high complication rate, the standard of care for CVC now requires the use of ultrasound (US) to navigate the procedure. While this technique has been proven to reduce the morbidity associated with CVC [7], the hand-eye coordination it requires may increase the training necessary to reach a high level of competency.

Hisey et al. (2018) developed the Central Line Tutor system to provide students with a safe, realistic, and interactive training environment for learning CVC [8]. This platform combines an industry standard venous access phantom with an electromagnetic (EM) tracker and a webcam. The EM tracker supports various visualization and task recognition functionalities by monitoring the pose of the needle, US probe, and phantom. These EM-based features include vessel segmentation in the US images, recognition of distinct US poses, and 3D visualization of the needle and US probe during training. The RGB camera is primarily used for workflow recognition, which is the key concept behind Central Line Tutor that facilitates step-by-step instructions and real-time analysis of performance.

While the EM tracking system does serve several important roles in the Central Line Tutor system, it also has substantial disadvantages related to cost and complexity. The use of EM tracking effectively doubles the cost of the existing Central Line Tutor system. It also introduces cumbersome tracking elements and requires frequent recalibration of tracked tools to ensure optimal performance. A critical long-term goal in the development of this training system is to eliminate the need for EM tracking altogether. A commercial optical tracker would not be suitable for this system either, since securing bulky optical markers would alter the geometry of the tools. To eliminate the reliance of Central Line Tutor on EM tracking, alternative methods for vessel segmentation, US pose classification, and 3D tool visualization must be developed that strictly make use of RGB video and US data.

In this study, we focus on implementing techniques for vessel segmentation and US pose classification that do not require an external tracking system. The existing method for vessel segmentation uses the EM system to track the position of the US probe relative to 3D models of the vessels. In the absence of tracking, there may be sufficient information in the US images themselves for a deep learning system to directly

[1] C. Barr, R. Hisey, T. Ungi and G. Fichtinger are with the School of Computing, Queen's University, Kingston, ON, Canada.

Corresponding author: C. Barr (c.barr@queensu.ca).

978-1-7281-1178-0/21/$31.00 ©2021 IEEE

segment the vessels. Furthermore, the use of deep learning for segmentation of neck vasculature in US images has been demonstrated previously in the literature [9].

To recognize tasks that use the US probe, Central Line Tutor analyzes the pose of the probe provided by the EM tracker. This probe pose is used to distinguish between “long-axis” scans, when the plane of the ultrasound image is parallel to the direction of the vessel, and “cross-section” scans, when the probe is perpendicular to these vessels. The appearance of the vessels in the US images also correlates strongly with the current type of scan being completed. We hypothesize that, in the absence of EM tracking, a system capable of recognizing these differences in vessel appearance from the US images would be able to classifying the type of scan being performed.

The purpose of this project is to assess the viability of replacing EM-based vessel segmentation and US pose classification with deep learning solutions that strictly require US images.

II. MATERIALS AND METHODS

A. Model Architecture

For vessel segmentation we decided to use the U-Net architecture. The U-Net is a deep architecture proposed by Ronneberger et al. (2015) for segmentation of biomedical images that consists of a contracting path followed by an expanding path [10]. It has quickly become a popular segmentation algorithm across much of computer vision, with many different variations of the architecture available online. The specific implementation of the U-Net that was selected for this task was created by Ungi et al. within the AIGT repository of the SlicerIGT toolkit [11, 12]. It has been previously shown to perform well with US segmentation and is largely optimized for use with US images, making it an ideal candidate for this project.

For the classification task, we decided to try two different network architectures. The first, referred to here as “CNN”, was a shallow architecture with 2 convolutional layers, a max pooling layer and a dense layer followed by a softmax activation layer (Figure 1). The second architecture, referred to as U-Net + CNN, started by segmenting the US image using the trained U-Net before passing the segmented image to the shallow CNN. The idea behind the U-Net + CNN network was to minimize the extraneous noise that the classifier had to deal with and provide only the salient information in the US images. Both networks output a vector of size 3, indicating which of the 3 possible pose classes a given US image belonged to. The first two classes correspond with specific scans performed during central line insertion, namely long-axis and cross-section scans. The third class identifies when the probe is in neither of the previous two poses, and is classified as an “undefined” pose.

B. Data

The dataset was obtained from 40 tracked US sequences collected by 4 medical students and 4 anesthesiologists using the Central Line Tutor system. To train the U-Net, a segmentation ground truth image had to be generated for each US image. This was accomplished by taking advantage of the spatial tracking data associated with the US images, as well as a previously generated 3D model of the vessels in the phantom. After manually fine-tuning the placement of the vessel models for each tracked sequence, the intersection of the vessel models with each tracked US image was extracted.

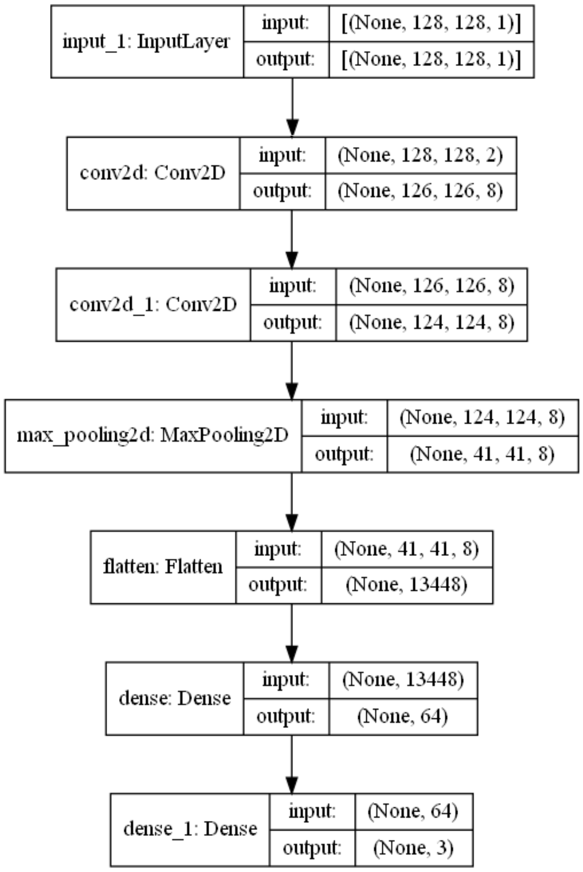
input_1: InputLayer
input: [(None, 128, 128, 1)]
output: [(None, 128, 128, 1)]
conv2d: Conv2D
input: (None, 128, 128, 2)
output: (None, 126, 126, 8)
conv2d_1: Conv2D
input: (None, 126, 126, 8)
output: (None, 124, 124, 8)
max_pooling2d: MaxPooling2D
input: (None, 124, 124, 8)
output: (None, 41, 41, 8)
flatten: Flatten
input: (None, 41, 41, 8)
output: (None, 13448)
dense: Dense
input: (None, 13448)
output: (None, 64)
dense_1: Dense
input: (None, 64)
output: (None, 3)

Fig. 1. Classification network architecture.

Following generation of the vessel segmentations, each image had to be further identified as either "cross-section", "long-axis" or "undefined" based on the probe position at the time of image capture. Pose labelling was performed by annotating each sequence with the timestamp of each pose transition, followed by labeling each frame with its corresponding US probe pose. Modules within the SlicerIGT DeepLearnLive extension were used to facilitate this labelling as well as generate the ground truth segmentations (github.com/SlicerIGT/aigt/tree/master/DeepLearnLive). The result was a set of 32,101 US images with corresponding segmentations and pose labels. The pose label distribution within this dataset was approximately 60% "undefined", 33% "cross-section" and 7% "long-axis". The class imbalance in the input data was preserved since this relative proportion of each class type is expected in real-world datasets this classifier will encounter.

C. Training

A cross-validation testing pipeline was used to ensure the testing results reflected the architecture's performance on unseen images. This approach used a leave-two-users-out scheme, wherein the images from one trainee and one expert were reserved for testing in each round while the rest of the data was used for training and validation. This scheme yielded testing data that matched the variation expected in real-world use, since a single participant will capture many consecutive sequences over the duration of a training session and the performance of each user will vary. Furthermore, an even split of novices and experts ensured that a broad spectrum of user skill level was captured in each testing set.

E. Evaluation

The performance of the U-Net was evaluated using Intersect over Union (IoU). This metric is a standard means of evaluating segmentation and object detection performance, and tends to be the preferred method over using per-pixel accuracy as it better captures overall structural similarities [13]. A custom loss function was created to optimize for IoU directly during U-Net training. Both the U-Net and classifier were trained for 20 epochs. Classification performance on the testing set was evaluated using accuracy, precision and recall, while the loss function used during classifier training was categorical cross-entropy.

III. Results and Discussion

A. U-Net Performance

The test set IoU score for each of the folds was 0.731, 0.817, 0.693 and 0.743, respectively. This yielded a mean test set IoU across the 4 folds of 0.746 ± 0.052. Figure 3 shows an example of the U-Net segmentation performance in fold 0 compared to the input segmentation and original US images across all 3 probe pose classes.

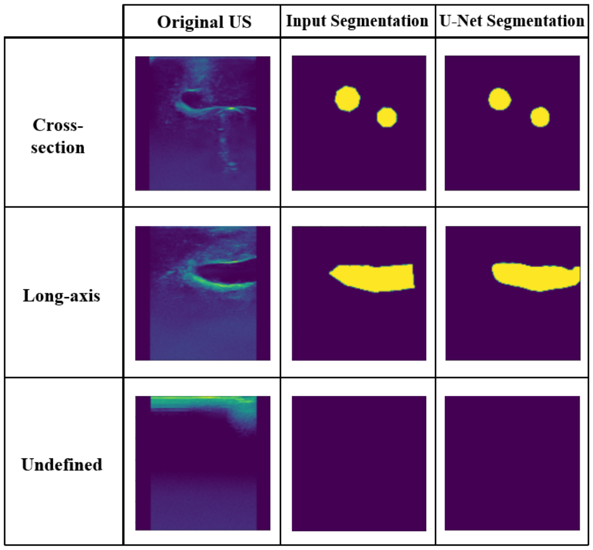

Fig. 3. Performance of U-Net tested on participant 4 compared to input segmentation and original US image. Rows are labelled according to their corresponding pose label. Note that the U-Net segmentation is a probability map that has not been thresholded.

B. Classifier Performance

The mean test set classification performance for U-Net + CNN was an accuracy of 92.7% ± 2.1%, while CNN achieved an accuracy of 92.0% ± 3.0. Table 1 shows the values for mean accuracy, mean weighted precision, and mean weighted recall for both networks. Tables 2 and 3 show the confusion matrices for the U-Net + CNN architecture and CNN architecture, respectively.

TABLE I. Overall Classifier Performance

Network Architecture	Mean Accuracy (%)	Mean Weighted Precision (%)	Mean Weighted Recall (%)
U-Net + CNN	92.7 ± 2.1	92.6 ± 1.3	92.6 ± 1.8
CNN	92.0 ± 3.0	92.9 ± 2.0	92.0 ± 3.0

TABLE II. U-Net + CNN Confusion Matrix

	Cross-section	Long-axis	Undefined
Cross-section	91.5%	4.7%	3.8%
Long-axis	25.1%	52.2%	22.7%
Undefined	1.2%	1.0%	97.8%

TABLE III. CNN Confusion Matrix

	Cross-section	Long-axis	Undefined
Cross-section	88.5%	17.1%	5.7%
Long-axis	29.1%	68.4%	13.8%
Undefined	1.3%	2.3%	96.4%

C. Discussion

The performance of the U-Net in vessel segmentation is encouraging based on the IoU scores and qualitative accuracy of the predicted segmentations. The low standard deviation in IoU performance between test sets suggest that the network performs well on unseen data. The extremely high contrast probability map output by the network suggests a high degree of confidence about the predicted segmentations, meaning minimal thresholding is necessary to refine the output. US images taken in the "cross-section" and "undefined" orientation tend to be of a qualitatively higher accuracy than those of the "long-axis" category. This is likely due to the occasionally ambiguous and noisy nature of images taken from this probe pose, with vessels appearing at unusual orientations and cut off at extreme angles.

The classifier results suggest that using the U-Net to segment input images before performing classification does confer a slight accuracy advantage. Analysis of the confusion matrices for these networks suggests that the CNN approach is more prone to falsely predicting the long-axis or cross-section class. The U-Net + CNN architecture is more likely to predict the undefined class when it is incorrect, meaning that it defaults to the base state of "no US scan occurring" when the scan type is unclear. This could be a result of the U-Net correctly returning blank segmentations for noisy images that do not show the vessels, and therefore simplifying the classification task for the CNN. As a result, the U-Net + CNN approach may be better suited to task recognition, since false positives that predict the undefined class are far less likely to throw off an overall task classification than an erroneous long-axis or cross-section prediction. This problem of occasional misclassifications can be further mitigated by taking several of the previous frames into account and using a majority voting technique. Since US tasks tend to occur in large consecutive blocks of frames, analyzing the neighbourhood around a given frame would have the effect of smoothing over sporadic classification errors.

A major limitation of this study is the use of a single venous access phantom to capture the full dataset. In practice, this system will need to generalize to any standard central line insertion phantom. Variation in material properties and vessel geometry among different phantoms may pose challenges for deep learning networks trained on a single phantom. Furthermore, performance may suffer on phantoms designed for other central line insertion sites like the subclavian and femoral veins. Testing this method on multiple central line phantoms will be an important future study to understand how generalizable this deep learning approach is.

The main next step for this research will involve integrating the vessel segmentation and US pose classification networks into Central Line Tutor and evaluating their performance. Of primary interest will be the task recognition accuracy compared to the existing EM gold standard. Completely removing the EM tracker will also require development of a new method for needle-based skills assessment. Comparing the predictive power of different quantitative metrics for trainee evaluation is a well-defined problem [14], and future research will explore how best to extract needle-based performance metrics from RGB and US images. While the long-term goal of replacing all EM tracking in Central Line Tutor is a non-trivial task, the potential benefits in terms of reduced cost and system complexity are well worth the effort.

IV. Conclusion

Providing automated feedback and instruction to medical trainees via task recognition has the potential to improve access to quality medical education, and reducing the cost of such systems is a critical consideration. In Central Line Tutor, the use of deep learning for vessel segmentation and US probe pose classification is an important step towards removing the external tracking system altogether. This study suggests that both tasks can be accomplished using deep learning at a level of accuracy sufficient for training purposes, and future work will be aimed at validating the performance of this classifier within the Central Line Tutor system.

Acknowledgments

C. Barr was supported by an NSERC CGS M award. R. Hisey is supported by the QEII-GSST award. G. Fichtinger is supported as a Canada Research Chair. Work funded, in part, by NIH/NIBIB and NIH/NIGMS (via grant 1R01EB021396-01A1 - Slicer+PLUS: Point-of-Care Ultrasound) and by CANARIE's Research Software Program. This work also was financially supported as a Collaborative Health Research Project (CHRP #127797).

References

[1] Bradley, P., "The history of simulation in medical education and possible future directions". Medical Education, 40(3), 254-262 (2006). doi:10.1111/j.1365-2929.2006.02394.x

[2] W. C. McGaghie, et al., "Does simulation-based medical education with deliberate practice yield better results than traditional clinical education? A meta-analytic comparative review of the evidence," Academic Medicine, vol. 86, no. 6, pp. 706-11, (Jun. 2011).

[3] Kornegay, J. G., Kraut, A., Manthey, D., Omron, R., Caretta-Weyer, H., Kuhn, G., . . . Yarris, L. M., "Feedback in Medical Education: A Critical Appraisal. AEM Education and Training", 1(2), 98-109 (2017). doi:10.1002/aet2.10024

[4] Langendyk, V., "Not knowing that they do not know: Self-assessment accuracy of third-year medical students," Medical Education, 40(2), 173-179 (2006). doi:10.1111/j.1365-2929.2005.02372.x

[5] Mcgee, D. C., and Gould, M. K., "Preventing Complications of Central Venous Catheterization," New England Journal of Medicine, 348(12), 1123-1133 (2003). doi:10.1056/nejmra011883

[6] Kumar, A., and Chuan, A., "Ultrasound guided vascular access: Efficacy and safety," Best Practice & Research Clinical Anaesthesiology, 23(3), 299-311 (2009). doi:10.1016/j.bpa.2009.02.006

[7] Froehlich, C. D., Rigby, M. R., Rosenberg, E. S., Li, R., Roerig, P. J., Easley, K. A., and Stockwell, J. A., "Ultrasound-guided central venous catheter placement decreases complications and decreases placement attempts compared with the landmark technique in patients in a pediatric intensive care unit," Critical Care Medicine, 37(3), 1090-1096 (2009). doi:10.1097/ccm.0b013e31819b570e

[8] Hisey, R., Ungi, T., Holden, M., Baum, Z., Keri, Z., McCallum, C., Howes, D. and Fichtinger, G. "Real-time workflow detection using webcam video for providing real-time feedback in central venous catheterization training," Proc. SPIE 10576, Medical Imaging 2018: Image-Guided Procedures, Robotic Interventions, and Modeling, 1057620 (2018).

[9] Groves, L.A., VanBerlo, B., Veinberg, N. et al., "Automatic segmentation of the carotid artery and internal jugular vein from 2D ultrasound images for 3D vascular reconstruction," Int J CARS 15, 1835–1846 (2020). https://doi.org/10.1007/s11548-020-02248-2

[10] Ronneberger, O., Fischer, P., and Brox, T., "U-Net: Convolutional Networks for Biomedical Image Segmentation." Lecture Notes in Computer Science Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, 234-241(2015). doi:10.1007/978-3-319-24574-4_28

[11] Ungi, T., Lasso, A., & Fichtinger, G. (2016). Open-source platforms for navigated image-guided interventions. Medical image analysis, 33, 181–186. https://doi.org/10.1016/j.media.2016.06.011

[12] Ungi, T., Greer, H., Sunderland, K. R., Wu, V., Baum, Z. M., Schlenger, C., . . . Fichtinger, G. (2020). Automatic Spine Ultrasound Segmentation for Scoliosis Visualization and Measurement. IEEE Transactions on Biomedical Engineering, 67(11), 3234-3241. doi:10.1109/tbme.2020.2980540

[13] Beers, F. V., Lindström, A., Okafor, E., & Wiering, M. Deep Neural Networks with Intersection over Union Loss for Binary Image Segmentation. Proceedings of the 8th International Conference on Pattern Recognition Applications and Methods (2019). doi:10.5220/0007347504380445

[14] Holden, M. S. (2018). Computer-Assisted Assessment and Feedback For Image-Guided Interventions Training (Doctoral dissertation, Queen's University, Kingston, ON). Retrieved from https://qspace.library.queensu.ca/jspui/handle/1974/25901