Tissue segmentation for workflow recognition in open inguinal hernia repair training

Elizabeth Klosa[a], Rebecca Hisey[a], Tahmina Nazari[b], Theo Wiggers[c], Boris Zevin[d], Tamas Ungi[a], Gabor Fichtinger[a]

[a]Laboratory for Percutaneous Surgery, School of Computing, Queen's University, Kingston, Canada
[b]Department of Surgery, Erasmus University Medical Center, Rotterdam, The Netherlands
[c]Incision Academy, Amsterdam, The Netherlands
[d]Department of Surgery, Queen's University, Kingston, Ontario, Canada

ABSTRACT

PURPOSE: As medical education adopts a competency-based training method, experts are spending substantial amounts of time instructing and assessing trainees' competence. In this study, we look to develop a computer-assisted training platform that can provide instruction and assessment of open inguinal hernia repairs without needing an expert observer. We recognize workflow tasks based on the tool-tissue interactions, suggesting that we first need a method to identify tissues. This study aims to train a neural network in identifying tissues in a low-cost phantom as we work towards identifying the tool-tissue interactions needed for task recognition. **METHODS:** Eight simulated tissues were segmented throughout five videos from experienced surgeons who performed open inguinal hernia repairs on phantoms. A U-Net was trained using leave-one-user-out cross validation. The average F-score, false positive rate and false negative rate were calculated for each tissue to evaluate the U-Net's performance. **RESULTS:** Higher F-scores and lower false negative and positive rates were recorded for the skin, hernia sac, spermatic cord, and nerves, while slightly lower metrics were recorded for the subcutaneous tissue, Scarpa's fascia, external oblique aponeurosis and superficial epigastric vessels. **CONCLUSION:** The U-Net performed better in recognizing tissues that were relatively larger in size and more prevalent, while struggling to recognize smaller tissues only briefly visible. Since workflow recognition does not require perfect segmentation, we believe our U-Net is sufficient in recognizing the tissues of an inguinal hernia repair phantom. Future studies will explore combining our segmentation U-Net with tool detection as we work towards workflow recognition.

KEYWORDS: Segmentation, U-Net, open inguinal hernia repair, workflow recognition, surgical training

1. INTRODUCTION

Competency based education is becoming the standard for medical training. With this training method, trainees are deemed competent in a skill by the approval of an expert observer. This puts a large time burden on experts as they must continually instruct and assess trainees to determine when they have reached proficiency in a skill. A computer-assisted training platform can take the place of an expert observer in providing trainees with instruction and feedback while they practice in a simulated environment. This would reduce the time burden on experts while still ensuring proper training. As more procedures transition to minimally invasive approaches, there has been reduced experience for surgical residents in open surgical procedures. In recent studies by Bingmer *et al.* and Potts *et al.*, they suggest that this decrease in surgical experience may lead to surgeons who lack the technical skills to safely complete some open procedures[1,2]. Although the minimally invasive techniques can result in faster recovery times and reduced pain and scarring, this approach is not always feasible. It is important that surgeons still possess the necessary skills and experience to perform an open procedure. Specifically, this study suggests that the large reduction in volume of open inguinal hernia repairs being performed may lead to decreased experience and technique of residents in this open repair[1].

An inguinal hernia occurs when there is a weakness in the abdominal wall, allowing contents of the abdomen to protrude out creating a hernia sac. Inguinal hernias are relatively common, and thus they are one of the primary procedures learned in general surgery residency[3]. One common open repair method is the Lichtenstein open inguinal hernia repair (LOIHR). This is a widely used method which involves either removing or pushing back the hernia sac, and placement of synthetic mesh for reinforcement of the posterior abdominal wall to reduce the risk of recurrence. The complicated nature of a LOIHR makes it important that residents acquire the proper technique during training. The likelihood of developing chronic pain after this procedure is relatively high, occurring in 8-16% of patients 6 months postoperatively[4]. Improper

technique can increase this chance of chronic pain, as well as the risk of recurrence and post-operative complications[5]. In order to limit these risks, it is important to ensure competency of graduating surgery residents in this procedure.

In a survey done by T. Nazari *et al*., residents estimated the number of surgical procedures needed for proficiency in open inguinal hernia repairs to be around 30-40 procedures[6]. With each hernia repair taking between 30-45 minutes, this is already 15-30 hours spent by an expert observer in training and assessing only one individual resident. A computer-assisted training platform to guide residents through a LOIHR would reduce this time requirement of experts, while still ensuring that residents receive sufficient training.

A computer assisted training platform for LOIHR would replace an expert observer in providing instruction and formative feedback to residents. To determine whether a resident is properly performing each procedural step, the system first needs a method to recognize which step is occurring. This can be done through workflow recognition. If the tasks can be recognized, we can automate the provision of workflow instructions and formative feedback to residents learning this procedure. Although multiple methods exist to recognize workflow tasks, in this study we decided to use video-based recognition. A video-based approach is advantageous as it allows the simulated procedure to feel realistic to trainees. It does not require costly sensors affixed to the tools, such as in electromagnetic tracking, that may hinder tool function and decrease the realism of the simulated setting. Previous studies use video-based recognition of tools as a method of task recognition. In one of these studies, tasks in central venous catheterization were recognized based on tools entering and leaving the field of view. This method works since each tool is only used for one task in the procedure and each task only uses that one tool. Thus, in central venous catheterization, the tool used defines the task[7]. However, workflow recognition of a LOIHR is more complicated. For one, the tools are not always entering and leaving the field of view. Even when not being used, the tools are often visible. As well, the same tools are used in multiple steps of the procedure and hence the procedure instead relies on the interactions of the tools with the tissues. For inguinal hernia repairs we must modify this approach to task recognition by incorporating knowledge of the tool-tissue interactions. This suggests that we first need methods to separately identify the tools and the tissues.

Identifying the tissues of an inguinal hernia phantom is necessary before we can identify which tissue a tool is interacting with. As such, this is the first step towards recognizing workflow and creating a LOIHR computer-assisted training platform. In this study, we segment the tissues of the inguinal hernia phantom for use in training a U-Net in tissue identification. If we can sufficiently identify the tissues, we are part way towards task recognition of LOIHR.

2. METHODS

2.1 Inguinal Hernia Repair Phantom

In developing the synthetic inguinal hernia repair phantom to be used for training of surgical residents, Nazari *et al.* sought to provide low-cost and easily reproducible phantoms to allow for widespread usage and training of LOIHR. This was important as each phantom could only be used once. The cost of materials for a phantom was less than five USD and a phantom could be produced in under 30 minutes[3]. The phantom represented the male groin region and included the skin, subcutaneous tissue, superficial epigastric vessels, Scarpa's fascia, external oblique aponeurosis, spermatic cord, hernia sac, and nerves, some of which can be seen in Figure 1. Table 1 shows the average pixel percentage of each tissue in a frame, demonstrating the size imbalance between the tissues.

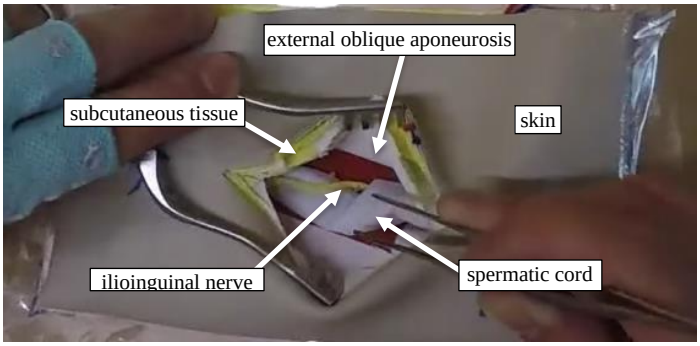

Figure 1: Inguinal hernia repair phantom used in performing a LOIHR procedure

Table 1: Average pixel percentage of each tissue per frame

Tissue	Average pixel percentage
Skin	32%
External oblique aponeurosis	2.5%
Subcutaneous tissue	1.37%
Scarpa's fascia	1.5%
Spermatic cord	1.2%
Hernia sac	0.44%
Superficial epigastric vessels	0.21%
Nerve	0.20%

2.2 Data Acquisition

Five experienced surgeons performed a LOIHR on a synthetic inguinal hernia repair phantom previously developed and validated by Nazari *et al*[3]. Each surgeon wore a head-mounted GoPro Hero 5 Black (GoPro Inc. San Mateo, California) with 720p resolution. The videos were recorded at 20 frames per second and were between 20 to 40 minutes long. Only the phantom and the surgeon's hands were visible in the videos to preserve anonymity of the surgeons.

2.3 Segmentation

Visible tissues in the videos were manually segmented using 3D Slicer. 3D Slicer (www.slicer.org) is a free, open-source platform used for medical imaging analysis and visualization. The videos were first partitioned into individual frames to segment using the 3D Slicer Module Deep Learn Live (https://github.com/SlicerIGT/aigt/tree/master/DeepLearnLive). To segment the tissues, we used the *SlicerAIGT* extension of 3D Slicer (https://github.com/SlicerIGT/aigt). An example segmentation can be seen in Figure 2. The segmented tissues were assigned a class label from one of the tissue types, or nothing. For this study, the ilioinguinal nerve and iliohypogastric nerve were segmented as one. The two nerves were made out of the same material and thus were identical to the network. The nerves can later be distinguished based on which step of the procedure is occurring. Both the original colour image along with the corresponding ground truth segmentation were exported as individual images to be used in training the U-Net.

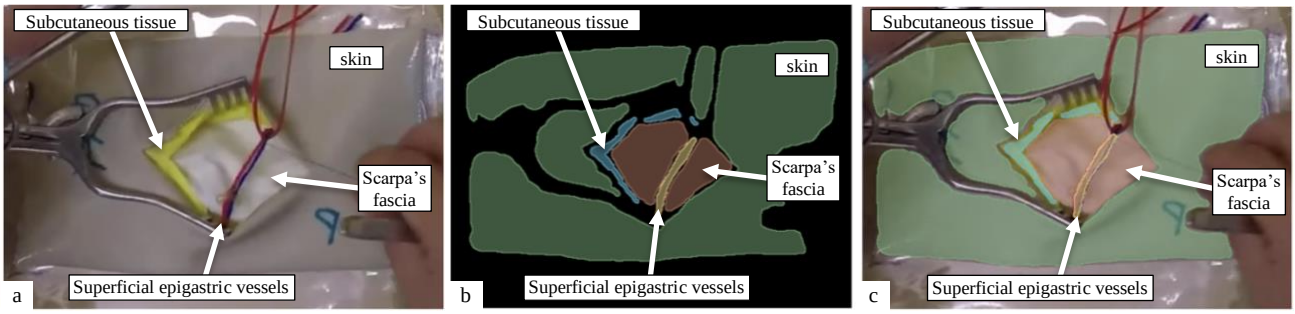

Figure 2: (a) Original image of the inguinal hernia repair phantom with the skin, subcutaneous tissue, Scarpa's fascia and superficial epigastric vessels visible (b) Segmentations of these tissues (c) Segmentation of the tissues overlaid on original image

2.4 Segmentation U-Net

For this study we trained a convolutional neural network with a U-Net architecture. U-Nets have symmetric contracting and expanding paths and skip connections that allow for precise localization[9]. The particular U-Net architecture that we used contained four contracting steps and four expansion steps. The U-Net was evaluated using leave-one-user-out cross validation. Five folds were used since we used five participant videos. For each of the folds, one video was reserved for testing, another was reserved for validation, and the 3 remaining videos were used for training. We evaluated the U-Net's performance using the F-score, which is the harmonic mean of precision and recall. Recall measures how well a model predicts true positives (TP) as positives whereas precision measures how many of the predicted positives were actually positive. We computed the mean F-score of each tissue to assess the U-Net's performance on the individual tissue types.

To determine the rates of incorrect tissue predictions, we computed the false positive rate (FPR) and false negative rate (FNR) with the formulas in Equation 1 and 2.

$$FPR = \frac{FP}{FP + TP} \tag{1}$$

$$FNR = \frac{FN}{FN + TP} \tag{2}$$

For the FPR, we chose to get the number of false positives (FP) predictions compared to the number of true and false positive predictions. Normally, the FPR would compare the ratio of false positives to the false positives and true negatives. However, due to the unbalanced classes, this would not be an effective measure of the false predictions due to the abundance of true negatives.

3. RESULTS AND DISCUSSION

The values in Table 2 show the U-Net's performance in recognizing each individual tissue type. Two example ground truth segmentations and corresponding U-Net predictions can be seen in Figure 4 and Figure 5.

Table 2: F-score, FPR and FNR values for each tissue type

Class	F-score	FPR	FNR
Nothing	0.97	0.02	0.01
Skin	0.69	0.16	0.46
Subcutaneous tissue	0.39	0.07	0.59
Superficial epigastric vessels	0.35	0.004	0.64
Scarpa's fascia	0.37	0.17	0.64
External oblique aponeurosis	0.43	0.21	0.49
Spermatic cord	0.61	0.32	0.32
Hernia sac	0.82	0.00	0.18
Nerve	0.80	0.02	0.01

Figure 3 shows the normalized confusion matrix for the tissues, representing the true and false positive and negative predictions. From this figure, we can see that most of the errors come from the FN predictions. The first column of Figure 3 highlights that the most common false prediction for each tissue was nothing as opposed to another tissue. When we look at the tool-tissue interactions, we would prefer that the U-Net not predict a tissue rather than misclassify it as another tissue. If the wrong tissue is recognized, the training system could falsely detect that the resident is on a different step. This would cause issues when trying to track surgical workflow.

The U-Net performed well in recognizing where no tissues were visible with an F-score of 0.97 and 99% true positive predictions. The U-Net also performed acceptably on the skin and spermatic cord with F-score values of 0.69 and 0.61, respectively. Varying factors may have led to the U-Net recognizing these tissues better than others. For example, the skin was visible in majority of the frames and composed large portions of the frame, providing many examples to train the U-Net. This, along with the distinct colour of the skin compared to other tissues, likely accounted for the high F-score value. As seen in Figure 3, 82% of the U-Net's predictions for the skin were correct, confirming the U-Net's sufficient performance on the skin. Similarly, 80% of the U-Net's predictions for the spermatic cord were correct. Since the spermatic cord was visible for large spans of the videos, success in distinguishing it from the other tissues was likely due to the large training sample that this provided. As well, this tissue was distinguishable in shape and in contrast to the red background inside the inguinal hernia repair phantom.

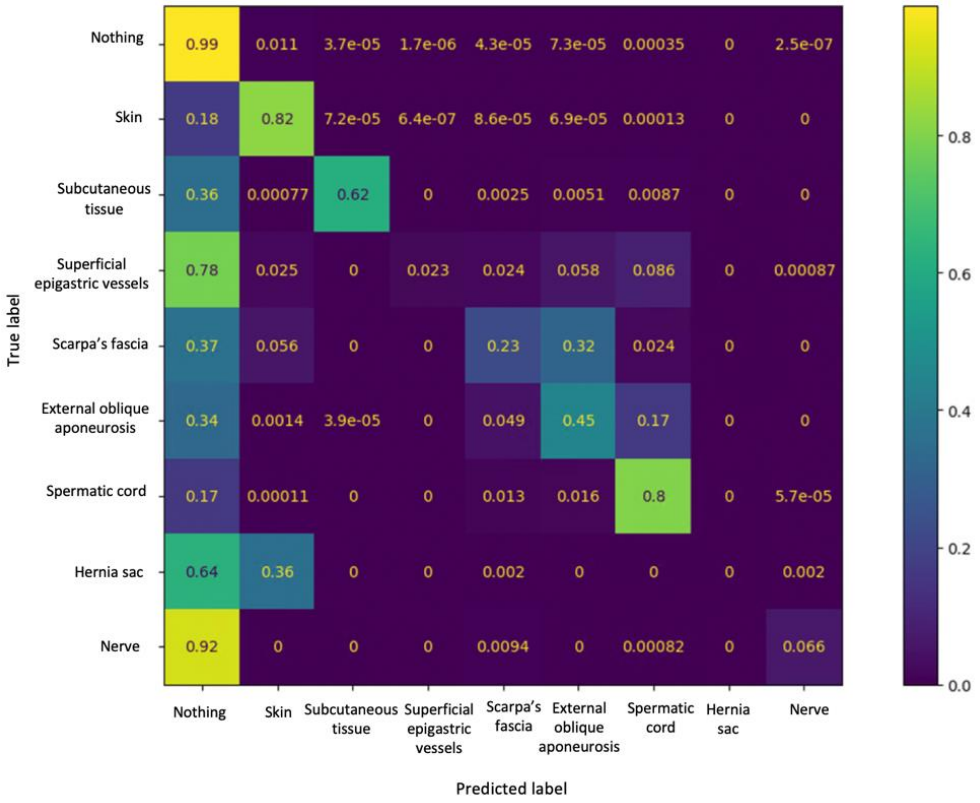

Figure 3: Normalized confusion matrix for the 9 classes comparing the true and predicted labels

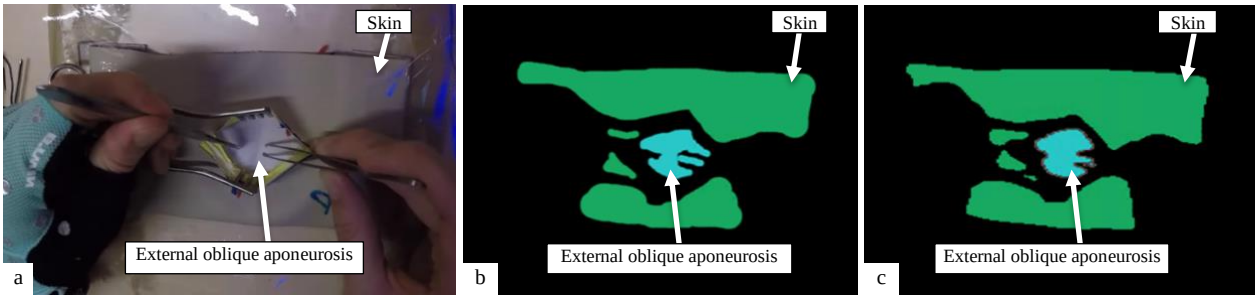

Figure 4: (a) Original image of open inguinal hernia repair phantom with the skin and external oblique aponeurosis visible (b) Ground truth segmentations of these tissues (c) Good U-Net prediction of both the skin and external oblique aponeurosis

The U-Net had F-score values for the Scarpa's fascia and external oblique aponeurosis of 0.37 and 0.43, respectively. Figure 4 shows an example where the U-Net correctly classified both the skin and external oblique aponeurosis. However, the Scarpa's fascia and external oblique aponeurosis were sometimes mistaken for each other in the U-Net's predictions. In these cases, the U-Net correctly identified the tissue region but classified it as the wrong tissue, explaining the lower F-scores of these tissues. This was presumably because both tissues were made of white material and located in the same place on the phantom. Since the Scarpa's fascia must be exposed and divided before the external oblique aponeurosis can be seen, we can mitigate this misclassification by incorporating knowledge of the procedural steps. Thus, these errors can be reduced once the segmentation U-Net is incorporated into recognizing workflow. Similarly, we will use this procedural knowledge to distinguish between the ilioinguinal and iliohypogastric nerve. In this study, both nerves were segmented together. However, knowledge of the workflow will help to determine which nerve is visible based on which step is currently being completed and which steps have already been completed.

The hernia sac and nerve had high F-score values of 0.82 and 0.80, respectively. However, in Figure 3 we can see that the U-Net only correctly predicted the nerve 6.6% of the time, while incorrectly predicting it to be nothing 92% of the time.

This indicates that the nerve was rarely recognized by the U-Net. Similarly, the hernia sac was never actually correctly predicted and was most commonly labelled as nothing or skin. As seen in Table 1, on average the hernia sac composed 0.44% of the frame and the nerve composed 0.20% of the frame. These tissues were so small relative to the image and only briefly present, giving the U-Net little training data to recognize them. Because the classes were so unbalanced, the F-score values were high just due to the abundance of true negatives (TN), where the U-Net correctly predicted that those tissues were not present. We believe that once we have obtained more labelled data to include in training, this will provide the U-Net with sufficient data to aid in recognition of these tissues.

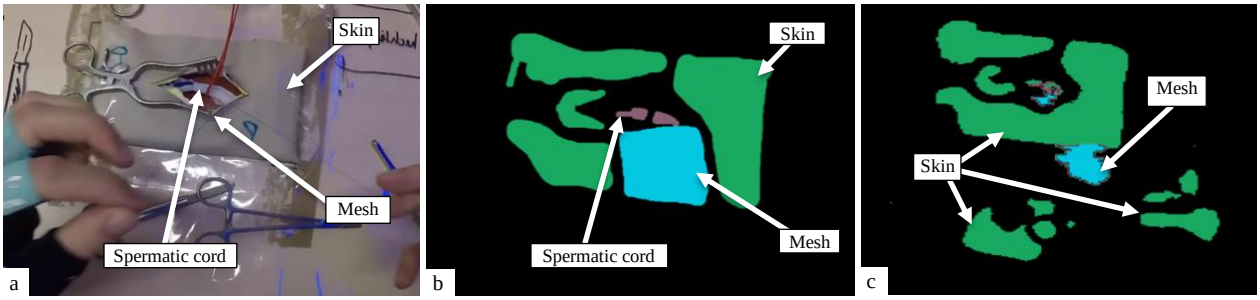

Figure 5: (a) Original image of open inguinal hernia repair phantom with the skin, spermatic cord, and mesh visible. (b) Ground truth segmentations of these tissues. (c) Poor U-Net prediction where the background was predicted as the skin

The use of head-mounted cameras to record the procedures was a factor in the misclassifications of this study. This meant that the camera was constantly moving and thus the background was often in the field of view. In several instances, parts of the background were mislabeled as tissues. This can be seen in Figure 5 where some of the background was predicted to be skin, which would account for the skins FPR of 0.16. These false predictions increase the FPR and lower F-score values of the tissue. We have begun using stationary cameras which should help to minimize these errors by ensuring the phantom is always centered in the frame. Although this error lowers the U-Net's metrics, it should not greatly affect the performance of the U-Net in recognizing tool-tissue interactions. We will be looking to identify the tissue near the tip of each tool, thus false positives outside of this region should not be a factor.

Although our U-Net misclassified some tissues, these misclassifications were acceptable for our purposes. Our end goal was not to recognize the tissues in each individual frame, but rather to recognize tasks in the workflow. As the videos used were recorded at 20 frames per second and each task occurred for longer than a second, an interaction between a tool and a tissue lasts for multiple frames. In identifying these interactions, we can take several frames and pixels into account and eliminate outlier misclassifications. This allows for discrepancies in the predictions where single or short sequences of misclassifications can simply be removed as long as the majority of the frames within a task were correctly classified.

Additionally, a completely accurate segmentation of tissues was not required for the development of a computer-assisted training platform. When combined with tool detection, it will be important that a given tissue can be recognized near the tip of a tool, and thus only small sections of each frame will need to be sufficiently recognized. An inexact boundary segmentation will not hinder the ability to recognize tool-tissue interactions. As well, the layout of the phantom used makes it advantageous for reducing misclassifications and identifying workflow tasks. As the steps of the procedure are completed, different layers of the phantom become visible. Thus, the tasks not only rely on the tool-tissue interactions but on the visible tissues as well. This makes the different steps easier to differentiate and can reduce false tissue predictions by limiting which tissues can be present in each step. Comparatively, an open procedure in which the view never changed would heavily rely only on those tool-tissue interactions and may have all tissues visible at all times. As such, we will incorporate knowledge of the procedural order of tasks to narrow down tissue types that may possibly be present, thus mitigating misclassifications.

The tissue segmentation process posed limitations as it was subject to slight ambiguities in the segmentations due to shadows, lighting and low resolution that often made it difficult to distinguish between tissue types. The inguinal hernia phantom also used the same materials for multiple tissues. For example, both the Scarpa's fascia and external oblique aponeurosis were made out of white felt, and both the ilioinguinal nerve and iliohypogastric nerve were made out of yellow yarn. In future studies we have begun revaluating the materials of the phantom to hopefully reduce misclassifications. Any changes to the materials, such as changing the felt colour or texture of one of the tissues, should have no effect on the face validity or realism of the simulated setting. The colour of the materials has no significance in regard to the tissues they represent and would not affect the way the repair is performed by the surgeon. However, it would likely help in making

the tissues more distinguishable to the U-Net. Another limitation was the small number of LOIHR videos used for this study. This provided only five videos for training the U-Net. With more recorded procedures, and thus more labelled data for use in training, the performance of the U-Net would likely improve.

Despite some misclassifications, our U-Net performed sufficiently in identifying simulated tissues for use in workflow recognition of a LOIHR. In future work, we will look to train neural networks in both tool recognition through object detection, and recognition of the general steps of the procedure. These networks will be used in combination with our segmentation U-Net in identifying the tissue-tool interactions and recognizing workflow. Recognition of the tasks being performed is needed for providing instruction and assessment to residents performing a LOIHR.

4. CONCLUSION

The U-Net was able to sufficiently recognize the simulated tissues during a LOIHR. The U-Net was more successful in recognizing the simulated tissues that were larger and provided more training samples, while having more trouble recognizing tissues that were not as prevalent. Tissues made of similar materials were not well distinguished by the U-Net. Upon dealing with these issues, we believe that our U-Net will be able to recognize the simulated tissues sufficiently well to be used in workflow recognition. When paired with tool detection, the networks will work in unison to identify the tool-tissue interactions that define each task. Therefore, recognition of the tissues will allow for task recognition of a LOIHR as we work to create a training platform that provides real-time instructions and objective feedback to trainees.

5. NEW OR BREAKTHROUGH WORK TO BE PRESENTED

We present a method for identifying tissue types of an open inguinal hernia repair phantom to be used in LOIHR training. A network capable of detecting tissue types will aid in recognition of tool-tissue interactions to determine the task being performed. Task recognition can be used to identify procedural workflow to develop a training platform that provides real-time instructions and feedback for trainees learning LOIHR.

ACKNOWLEDGEMENTS

E. Klosa is funded by the NSERC Undergraduate Summer Research Award. This work was 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 was also financially supported as a Collaborative Health Research Project (CHRP #127797), a joint initiative between the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Canadian Institutes of Health Research (CIHR). R. Hisey is supported by NSERC as a Canada Graduate Scholar. G. Fichtinger is supported as a Canada Research Chair in Computer-Integrated Surgery. Financial support was received from the Southeastern Ontario Academic Medical Association (SEAMO), Educational Innovation and Research Fund.

REFERENCES

[1] Bingmer, K., Ofshteyn, A., Stein, S.L. et al., "Decline of open surgical experience for general surgery residents," Surgical Endoscopy 34(2), 967–972 (2020).

[2] Potts III, J. R., & Valentine, R. J., "Declining resident experience in open vascular operations threatens the status of vascular surgery as an essential content area of general surgery training," Annals of surger, 268(4), 665-673 (2018).

[3] Nazari, T., Simons, M.P., Zeb, M.H., et al., "Validity of a low-cost Lichtenstein open inguinal hernia repair simulation model for surgical training," Hernia 24, 895–901 (2020).

[4] Andresen, K., Rosenberg, J., "Management of chronic pain after hernia repair," Dove Medical Press Limited (2018).

[5] Kulacoglu, H. "Current options in inguinal hernia repair in adult patients," Hippokratia 15(3), 223-231 (2011).

[6] Nazari, T., Dankbaar, M.E.W., Sanders, D.L. et al., "Learning inguinal hernia repair? A survey of current practice and of preferred methods of surgical residents," Hernia 24, 995-1002 (2020).

[7] Hisey, R. J., "Computer-assisted workflow recognition for central venous catheterization," thesis (2019).

[8] Ronneberger O, Fischer P, Brox T., "U-net: Convolutional networks for biomedical image segmentation," Medical Image Computing and Computer Assisted Interventions, 234-241 (2015).