System for central venous catheterization training using computer vision-based workflow feedback

Rebecca Hisey,
Laboratory for Percutaneous Surgery at the School of Computing within Queen's University, Kingston, ON, Canada.

Daenis Camire,
Department of Critical Care Medicine, Queen's University, Kingston, ON, Canada.

Jason Erb,
Department of Critical Care Medicine, Queen's University, Kingston, ON, Canada.

Daniel Howes,
Department of Critical Care Medicine, Queen's University, Kingston, ON, Canada.

Gabor Fichtinger [Fellow, IEEE],
Laboratory for Percutaneous Surgery at the School of Computing within Queen's University, Kingston, ON, Canada.

Tamas Ungi
Laboratory for Percutaneous Surgery at the School of Computing within Queen's University, Kingston, ON, Canada.

Abstract

Objective: To develop a system for training central venous catheterization that does not require an expert observer. We propose a training system that uses video-based workflow recognition and electromagnetic tracking to provide trainees with real-time instruction and feedback.

Methods: The system provides trainees with prompts about upcoming tasks and visual cues about workflow errors. Most tasks are recognized from a webcam video using a combination of a convolutional neural network and a recurrent neural network. We evaluated the system's ability to recognize tasks in the workflow by computing the percent of tasks that were recognized and the average signed transitional delay between the system and reviewers. We also evaluated the usability of the system using a participant questionnaire.

Results: The system was able to recognize 86.2% of tasks in the workflow. The average signed transitional delay was –0.7s. The average usability score on the questionnaire was 4.7 out of 5 for the system overall. The participants found the interactive task list to be the most useful component of the system with an average score of 4.8 out of 5.

rebecca.hisey@queensu.ca .
D. Camire, J. Erb, and D. Howes are with the Department of Critical Care Medicine, Queen's University, Kingston, ON, Canada.

Conclusion: Overall, the participants' response to the system was positive. Participants perceived that the system would be useful for central venous catheterization training. Our system provides trainees with meaningful instruction and feedback without needing an expert observer to be present.

Significance: We are able to provide trainees with more opportunities to access instruction and meaningful feedback by using workflow recognition.

Keywords

Computer aided instruction; Catheterization; Artificial neural networks

I. INTRODUCTION

CENTRAL venous catheterization (CVC), otherwise known as a central line insertion, is the procedure of inserting a catheter into a major vein that provides a near direct path to the heart. Common sites for catheter insertion are the internal jugular, subclavian and femoral veins [1]. CVC is often performed in emergency situations and on critically ill patients. It is also used in non-emergency situations for patients that require frequent venous access, such as those undergoing chemotherapy. Since CVC has such a wide range of applications, it is considered as an essential skill that is taught in residency of most medical specialties.

This procedure, while common, is not without risks. Typical complications include hematoma, hemorrhage and sometimes stroke [2]. A major influence on the risk of complications is the experience of the physician performing the procedure. Studies have shown that patients are 35% more likely to experience complications from a central line if the procedure is performed by a novice compared to an expert physician [3]. Clearly, trainees need an environment to learn mastery of the procedure without posing risks to patients.

Training in a simulated setting is one way that risk to patients can be mitigated. Students trained on simulators have lower complication rates than those trained on live patients [4][5]. In addition, students trained in a simulated setting also typically require less assistance, make fewer errors and report higher confidence when they perform the procedure on live patients [6][7]. In general, training in a simulated setting has been shown to lead to a reduced complication rate and better patient outcomes than traditional training methods [8]. For CVC, simulated training involves performing the procedure on a part-task trainer which is comprised of a phantom with fluid filled vessels.

Unfortunately, static phantoms are missing a key component that can make them effective as standalone trainers. Feedback is widely regarded to be an essential component of trainee learning and frequent feedback has been shown to improve patient outcomes [9]. Most often this feedback comes from an expert observer and requires a large time commitment from instructors and evaluators. Studies have shown that attending physicians commonly cite lack of time as their most significant obstacle to providing trainees with feedback [10]. When physicians are unable to find time, this limits trainees' opportunity to receive meaningful feedback. Research in medical education has shown that trainees are discontent with the

quantity of feedback that they receive [11]. Furthermore, research has shown that trainees cannot accurately assess their own performance without expert feedback [12]. This evidence suggests that trainees not only need an environment to practice that is safe for patients, but one that is also able to provide them with meaningful, actionable feedback on their performance.

Workflow recognition is one way to provide trainees with performance feedback without needing an expert physician to be present. In this way, a computer system can be used to provide instruction for upcoming tasks, or give feedback about errors that have been made. By using a computer to provide this feedback, this alleviates the time burden on instructors and proctors. Workflow recognition has been a substantial area of research in recent years. A common method of recognizing tasks in surgical workflow is to monitor the movements of tools using optical or electromagnetic (EM) tracking. Recently, approaches have focused on the recognition of tasks from video [13][14]. This is because video is widely available, inexpensive, and does not affect the use of various surgical tools by requiring bulky markers. Many of the studies that recognize workflow tasks from video rely on deep learning.

Recent methods of surgical workflow recognition have begun combining convolutional neural networks (CNNs) with long-short term memory (LSTM) networks and have shown good success [15]. LSTMs have been shown to be very successful at recognizing events that are separated by long periods of time [16]. This makes them ideal for workflow recognition. This approach allows the network to make decisions based on the information contained in a series of images in a video, rather than basing the decision on a single frame.

In this paper, we demonstrate how the need for an expert observer during practice can be eliminated using a training system that incorporates workflow recognition to provide real-time instruction and feedback. This system, which we have called Central Line Tutor, uses a neural network on webcam video data along with EM tracking to recognize the tasks in CVC. We evaluate Central Line Tutor on both its ability to recognize the tasks in the workflow, and its usability for training novices in CVC.

II. METHODS

A. System Setup

Central Line Tutor is a system designed to provide instruction and feedback to trainees learning CVC without needing an expert observer to be present. Central Line Tutor currently uses a combination of video-based workflow recognition and recognition based on EM tracking to provide instruction. Though we are working hard to eliminate the need for the EM tracking system, it is still currently necessary to evaluate critical tasks in the workflow. The system is comprised of a computer, webcam, a central venous access phantom, an EM tracker and an ultrasound machine (Figure 1). EM sensors are placed on the ultrasound probe, needle and the phantom. The hardware interface is provided by PLUS-toolkit (https://plustoolkit.github.io), which synchronizes and streams all data to the main application [17].

The software component of the system is also built on the platform 3D Slicer which is an open-source platform for medical image informatics (www.slicer.org). The system built

upon the Perk Tutor extension within 3D Slicer (www.perktutor.github.io). Perk Tutor is an open-source extension for training ultrasound-guided needle interventions. Perk Tutor uses the EM tracking information to measure performance metrics such as path length, translational movements and procedure completion time [18]. The Perk Tutor platform allows us to incorporate these previously validated metrics into our system.

A preliminary description of the system can be found in [19]. The system used in this paper, contains the same hardware setup, however the earlier paper presented only a rudimentary color-based method for recognizing the tasks in the workflow based on surgical tools used in the procedure. The earlier description of the system also did not include a detailed explanation of how the EM-based task recognition was performed and the study conducted in [19] used a dataset that contained only a single novice user. The previous study also utilized a different kit of central line tools compared to this study, therefore there is no overlap in data used between the two studies. In addition to using more sophisticated methods for workflow recognition, and testing on a larger, more varied dataset, this paper also goes beyond what was done previously by conducting face and content validity testing with a diverse group of users.

B. User Interface

The user interface displays both the ultrasound and webcam videos to the users. The interface also displays 3D models of the phantom, ultrasound probe and needle. The EM sensors placed on each of these components allow the trainee to visualize their movements in space via a series of coordinate transformations. Since the vessels are static within the phantom, trainees can see where the ultrasound probe and needle are with respect to the location of the vessels. Holden *et al.* have previously validated that the use of these 3D visualizations is an effective tool that improves trainee learning in ultrasound guided needle interventions [20].

In addition to the 3D visualizations, this system also provides users with three levels of difficulty with progressively fewer aides (Figure 2). The first level has the largest number of aides. Trainees are able to see the position of the ultrasound image plane and needle with respect to the vessels. The vessels are also color-coordinated to clearly identify the target vein. Trainees are also provided with a complete checklist of the tasks in the CVC workflow. This checklist is interactive and each time the trainee completes a task it is both checked and highlighted in green. Missed tasks are highlighted in red to clearly identify workflow errors. Trainees also receive a "current step prompt" that indicates which task the trainee should perform next. The second level reduces the checklist so that instead of a complete list of individual tasks, trainees are only given general categories of tasks. Furthermore, in this level trainees are no longer able to see the vessels and phantom in the 3D viewer. The trainees are only able to see the ultrasound probe and needle. This forces them to rely on the ultrasound video to determine where the ultrasound image is with respect to the vessels. In the final level, trainees are given only the ultrasound video and no instruction.

C.   EM-based Task Recognition

To update the checklists, Central Line Tutor uses two different methods of workflow recognition to identify the tasks in the CVC workflow. The first method uses EM tracking to recognize tasks that involve the ultrasound probe and the needle. These tasks are more critical than others in the workflow, and it is important that we can measure not only that they are completed, but they are completed correctly and efficiently. Using tracking for these tasks allows us to compute performance metrics using the aforementioned Perk Tutor extension. The tasks recognized by this method include cross section and long axis scans of the vessel with the ultrasound probe, and inserting the needle into the correct vessel.

Tasks that involve EM tracked tools are recognized based on the position and orientation of the tools with respect to the vessels in the phantom. Since the vessels are static within the phantom, we can track them by placing a sensor on the exterior of the phantom. The phantom serves as our reference coordinate system. To recognize the two different types of ultrasound scans, we determine if the plane created by the ultrasound image intersects both vessels, or only one. To do this, we first transform the coordinates of the four corners of the ultrasound image plane into our reference coordinate system (Figure 3). Cross section scans are defined as when the ultrasound image plane intersects the center line of both vessels (Figure 4(a)). Long axis scans are defined as being complete when the shortest distance from each of the vertical edges of the ultrasound image plane to the center line of a single vessel is less than the radius of the vessel (Figure 4(b)). The task of inserting the needle into the vessel is recognized in a similar manner. The coordinates of the tip of the needle are first transformed into our reference coordinate system. The needle insertion task is defined as being complete when the shortest distance from the needle tip to the center line of a vessel is less than half of the radius of the vessel. Since these tasks are defined based on the physical location of the US image with respect to the vessels within the phantom, there is no learning process needed for our system to recognize these tasks.

D.   Video-based Task Recognition

The majority of tasks are recognized from the webcam video using a neural network. This is because video is less expensive than EM tracking and does not require additional sensors to be added to the tools. Additional sensors are bulky and can make practicing in a simulated setting feel less realistic. Furthermore, several tools are very small and they cannot perform their function if sensors are attached. For example, the guidewire must be threaded through the center of both the dilator and the catheter, neither of which have openings wide enough to accommodate an EM sensor.

The task recognition network runs in an Anaconda virtual environment and communicates with Central Line Tutor via the OpenIGTLink protocol, as shown in Figure 5. 3D Slicer passes the image to the Anaconda environment where the image is normalized, resized, then classified by the task recognition network. The task label and the confidence are passed from the environment back to 3D Slicer using the OpenIGTLink protocol implemented within the PyIGTLink library. The tasks that are recognized using the webcam include: applying local anesthetic, removing the syringe from the needle, inserting and removing the

guidewire, inserting and removing the dilator, cutting the skin with the scalpel and inserting the catheter.

1) Task Recognition Networks—The network that we use is a combination of a CNN and a network that uses an ensemble of parallel LSTM units. Since each task in the CVC workflow only uses one surgical tool at a time, we can define the various tasks in the workflow based on which tool is currently in the field of view. The CNN that we use for tool recognition is ResNet50 [21]. From now on we will refer to this network as CNN_{Tool}. To reduce the amount of data needed to train CNN_{Tool} we use transfer learning. The weights of CNN_{Tool} are initialized from those obtained by training on the much larger ImageNet dataset. This allows the network to learn basic image features before being trained on our specific dataset [22]. The model and the pre-trained weights are available through Keras' application library (https://keras.io/api/applications/resnet/#resnet50-function). Once the weights have been initialized, we replace the final layer with a new softmax layer whose size corresponds to the number of tools. Finally, we train all layers of the network on our central line dataset. The final tool prediction is given by the class that produces the highest confidence in the softmax output of the network. While this approach works reasonably well, identifying tools in individual video frames does not provide any information about what is occurring within the context of the entire procedure.

To allow for the inclusion of contextual information we include information from previous frames and produce the task classification from a sequence of images, rather than a single snapshot. To create these sequences, we first classify each image with CNN_{Tool} to obtain the tool labels. Each sequence is composed of the tool classification output for the frame that we are currently analyzing and the tool classification output for the 49 previous frames. As the network receives a new image, the tool classification output of the new frame is appended to the end of the sequence and the tool classification of the oldest frame is removed from the start of the sequence. Each time the sequence is updated, it is fed to a second network that produces the task classification.

The first layer of our task classification network is an ensemble of parallel, bidirectional LSTM units. We use one LSTM unit for each of the 7 tasks involved in the procedure. We also add one additional unit for the case of no task. The idea behind this approach was that by incorporating one unit for each task, this would in theory allow each unit to learn a binary response to the sequences that are characteristic of one specific task. Each of the units independently processes the sequence of classifications both forwards and backwards before producing a single output based on the entire sequence of classifications. Using this approach, we allow each LSTM layer to only be responsible for performing a binary classification rather than a multi-class classification. The output from the each of the LSTM layers is then concatenated before going through two dense layers with ReLU activation functions and a final softmax layer to produce the task classification. For each sequence the final task prediction label is determined by the class with the highest confidence from the softmax output. The network structure can be seen in Figure 6. We refer to this combination of networks as CNN+LSTM. The specific implementations for these networks can be found at: https://github.com/SlicerIGT/aigt/tree/master/DeepLearnLive/Networks

2) Dataset—The central line dataset is composed of 40 videos in total. 20 videos were recorded from 4 medical students and 20 videos were recorded from 4 experienced clinicians. Each participant recorded five trials of the procedure using the Central Line Tutor setup. All recordings were done using the system's first difficulty level that displays the entire task list. However, participants were given the opportunity to explore all difficulty levels and attempt the procedure using only US guidance. The students are fourth year medical students who are novices in the procedure and report having performed one CVC or fewer. The expert participants are experienced anesthesiologists, who routinely perform CVC in their practice. The recordings include the color video from the webcam, along with the ultrasound video and the positional information from the EM tracking system. The webcam video shows only the participant's hands, the surgical tools in use and the area of insertion of the phantom. There are no identifying features visible in order to protect the anonymity of the participants. To ensure that the system can perform well in different environments the videos were recorded in different locations, with different lighting conditions and varying camera positions.

To create our dataset, each video is divided into individual frames. Each image is manually assigned two labels. The first label corresponds to the tool that is actively in use when the image was taken. In the rare case that multiple tools were visible in the frame, the active tool was assumed to be the tool that was in the participant's hand. The possible tool labels include anesthetic, catheter, dilator, guidewire, guidewire casing, scalpel, syringe and no tool. These labels are used as the ground truth for training CNN_{Tool}. The second label corresponds to the task that is being performed. The possible labels in this category include apply anesthetic, insert needle, insert guidewire, use scalpel, dilate opening, insert catheter and no task. These task labels serve as the ground truth labels for training the LSTM portion of our CNN+LSTM network. A full breakdown of the number of images per class can be seen in Table 1.

In addition to labelling each frame with both task and tool labels, each video is annotated by three independent reviewers to identify transition points within the procedure. Transition points represent the time in a video when a new task begins. If a transition point is missing, this indicates that the associated task did not occur. For EM recognized tasks, these points occur when the ultrasound probe or needle are in a specific position or orientation. For webcam recognized tasks, these transition points typically occur when a specific tool enters or leaves the scene. A full list of transition points, tasks that they correspond to and their prevalence within the dataset is found in Table 2.

E. Experiments

1) Task Recognition Performance—One of the key features of Central Line Tutor is its ability to provide instructions and feedback in real-time. To demonstrate content validity, we compared the system's ability to recognize transition points in real-time to the transition points recognized by human reviewers.

To test Central Line Tutor's capabilities, the full recordings (consisting of the synchronized recordings of the EM tracking information, ultrasound video and webcam video streams) were replayed on the system to simulate the procedure being performed in real-time. The

system recorded the name and timestamp of each transition point as it occurred in the video. To compare our results to our previous studies, this process was repeated twice for each video. The first time using CNN_{Tool} to recognize the webcam transition points, as in previous studies, and the second time using CNN+LSTM.

To assess the how well Central Line Tutor was able to recognize tasks using each network, we measure the percentage of transition points recognized when the system incorporated CNN_{Tool} into its workflow recognition compared to CNN+LSTM. This metric is defined by the number of transition points that were recognized by the system divided by the total number of transition points that are present in all recordings. Except the task of inserting the needle, each task in the procedure corresponds to a single transition point. As such the percentage of correctly identified transition points provides a measure of the number of tasks that were recognized compared to the total number of tasks that were identified by the reviewers. We choose to measure the percentage of correctly identified transition points as opposed to the classification accuracy on the individual images in the dataset to address the issues of class imbalance. As seen in Table 1 there is a large discrepancy between the number of images in each class, with several classes being heavily underrepresented. Conversely, from Table 2 we can see that there is a more even distribution of transition points across all tasks in the recordings. This metric also allows us to measure the performance of the system on each trial as a whole, rather than the performance on individual images.

For this study, each of the networks were trained using a leave-two-user-out (L2UO) cross validation scheme. For each fold, we reserve all videos from one expert participant and one novice participant. From the remaining six participants, we reserve one video from each participant for validation and all remaining videos are used for training.

To assess how quickly Central Line Tutor can recognize tasks compared to the human reviewers we measured the average signed transitional delay (TD_{signed}) between the timestamps generated by the system and the average timestamp given by the three reviewers. Average signed transitional delay is defined in (1).

$$TD_{signed} = \frac{\sum_{i=1}^{N} \sum_{j=1}^{M} T_{CLT}(i,j) - T_{human}(i,j)}{N * M} \tag{1}$$

In this equation $T_{human}(i,j)$ and $T_{CLT}(i,j)$ represent the average timestamp on video i for transition point j generated by the human reviewers and Central Line Tutor respectively. In these equations N represents the number of videos, and M represents the number of transition points in the procedure. For average signed transitional delay, a negative value indicates that the system recognized the transition point ahead of the human reviewers. Conversely, a positive value indicates that the system recognized the transition point after the human reviewers. This metric allows us to discern whether the system is typically early or late to recognize tasks.

Finally, to establish the statistical significance of our results we measure the absolute intra-class correlation (ICC) for each transition point in the dataset. This metric reports how

closely the times generated by Central Line Tutor compare to the average times given by the reviewers. For completeness, we also measure the agreement between each of the three reviewers using the same absolute ICC metric. In addition to computing the ICC between the reviewers and Central Line Tutor, we also test for significant difference by performing a paired t-test between the average timestamps generated by the reviewers and the system.

2) **Participant Survey**—Finally, to establish face validity we evaluate the usability of Central Line Tutor by asking participants to fill out a survey once they complete their trials. The survey asks questions about the realism of various aspects of the system, and how useful they believe they are. Each of the questions is answered on a 5-point Likert scale, where 1 indicates that the participant strongly disagrees with the statement and 5 indicates that the participant strongly agrees. The questions are divided into three categories: how realistic the system is, how useful the various components of the system are for training, and how useful they believe the system will be for different use cases. We use an independent t-test to determine if there is a significant difference between the ratings given by the novices compared to the experts.

III. RESULTS

A. Task Recognition Performance

In our evaluation of content validity, we found that Central Line Tutor was able to recognize 92.7% of transition points when using CNN_{Tool} and EM tracking. The system was able to recognize 86.2% of transition points when and CNN+LSTM was used in place of CNN_{Tool}. The tasks that were most reliably recognized from their transition points include: inserting the needle and inserting the catheter. The tasks that were least reliably recognized were the tasks involving the anesthetic, scalpel and dilator. The full results for the percent of transition points recognized can be seen in Table 3.

When we look at average signed transitional delay for the entire procedure, the combination CNN+LSTM and EM tracking had an average transitional delay of -0.8 ± 8.7s, whereas CNN_{Tool} and EM tracking had an average transitional delay of -7.4 ± 19.5s. When we compare EM-recognized tasks to the tasks recognized by the webcam, we see that the EM tracked tasks are recognized ahead of the human reviewers with an average signed transitional delay of -4.2 ± 8.4s. For the webcam tasks, CNN+LSTM typically recognized tasks after the human reviewers with an average signed transitional delay of 1.2 ± 10.0s, while CNN_{Tool} recognized tasks ahead of reviewers with an average signed transitional delay of -9.5 ± 26.4s. The breakdown of average signed transitional delay for each individual transition point is shown in Figure 7.

The absolute ICC between the average reviewer timestamps and those generated by Central Line Tutor were 0.97 when using a combination of CNN_{Tool} with EM tracking and 0.99 when the combination of CNN+LSTM with EM tracking was used. Furthermore, the average absolute ICC between the three reviewers was 0.98. The results of the paired t-test indicate that there was no significant difference between the timestamps generated using Central Line Tutor and the reviewers when CNN+LSTM was used to recognize the webcam

transition points ($p = 0.9$). However, there was a significant difference when CNN_{Tool} was used ($p < 0.001$).

B. Participant Survey

The system was favorably received by all participants. The average score across all questions was 4.7 out of 5. On the questions concerning the realism of the system, which is of particular concern for face validity, the average score was 4.5 with participants finding the ultrasound visualization of tissue and vascular space to be the most realistic component. The average score for the usefulness of the various components of the system was 4.6. For this section, participants felt that the task list was the most useful component of the system. In terms of general usability, the average score across questions in this section was 4.8. The situations in which the participants believed the system would be most useful were in the improvement of ultrasound-handling proficiency, improvement of needle insertion and in facilitating preparedness. The results of the survey are shown in Figure 8. There results of the t-test indicate that there was not a significant difference between the ratings of the novice participants and the experts ($p = 0.04$).

IV.   DISCUSSION

We evaluated Central Line Tutor's ability to recognize tasks in the CVC workflow in real time. We compared two different networks for task recognition to determine which would allow Central Line Tutor to have the best performance. Finally, we tested Central Line Tutor's usability through a participant survey.

To evaluate the real-time performance of the networks and establish content validity, we tested the networks' ability to detect transition points in the procedure. These transition points were the time at which a new task began. Even though CNN_{Tool} recognized more transition points than CNN+LSTM, when we look at the average signed transitional delay, we can see that CNN_{Tool} had a large negative transitional delay indicating that it typically recognized tasks well ahead of the human reviewers. The fact that there was a significant difference between the reviewers and CNN_{Tool} indicates that CNN_{Tool} was most likely making erroneous predictions. Conversely, if we look at CNN+LSTM we can see from the positive average signed transitional delay that it tended to recognize tasks after the reviewers. Even though CNN+LSTM recognizes tasks after the reviewers, there was not a significant difference between the two. Furthermore, from the absolute ICC results we can see that the timestamps generated using CNN+LSTM were more closely correlated with the timestamps from the reviewers compared to CNN_{Tool}. These results suggest that using CNN+LSTM in Central Line Tutor for workflow recognition will improve the system's ability to recognize tasks in the procedure over using CNN_{Tool} alone.

From the results we can also see that classes that had the fewest images such as the anesthetic, scalpel and dilator did have lower recognition accuracy for their corresponding tasks compared to larger classes. This is likely due to the lack of balance in our dataset. However, the lack of balance actually benefits a system designed to provide instruction. By ensuring that the training data contained a high percentage of "no tool" images, the system was biased towards false negatives. This is favorable because missing a task has minimal

effect on the instructions that the system provides while many false positives would cause the interactive checklist to move to different tasks in the workflow causing confusion for the user. In future studies, we will evaluate the effects of network accuracy on the detection of transition points.

Finally, we tested the usability of Central Line Tutor by asking participants to fill out a survey. Participants were asked to evaluate Central Line Tutor based on its realism and usability. Overall the system was well received by both expert and novice participants. The results of the survey indicated that participants felt that the system delivered a realistic training experience and that it had many useful features to help with the learning process. Trainees especially liked the interactive task list and the multiple levels of difficulty. In general, the survey results indicate that participants believed the system would be a useful tool for training CVC.

One limitation of this study was that we only have a single setup of the system available to us. We have done our best to ensure that this approach will be generalizable for others who may wish to replicate this work by including variations in lighting, setup location and camera position within our dataset. We also selected a tool kit that is easily available and widely used in many different hospitals. In future studies, we will test the generalizability of our models by testing our system using different phantoms and tool kits. We also cannot guarantee that the performance of this network structure was fully optimized during the training process. In the future we will determine if the accuracy of this model can be improved by testing various combinations of hyper-parameters and network structures.

In the future, we plan to conduct further testing to formally determine the effects of the system on the efficiency of training CVC. We also aim to expand on this work by making Central Line Tutor into a low-cost system that uses a minimal number of hardware components. By using simulated ultrasound and video-based tool tracking methods, we intend to eliminate the need for an ultrasound machine and EM tracking system. The removal of these two components will minimize the amount of maintenance required by the system and significantly reduce the cost of the system, thus making it more accessible.

V. CONCLUSION

We presented a functional system for training central venous catheterization. This system provides instruction and feedback to trainees learning the procedure without needing an expert observer to be present. To test the system's ability to recognize tasks in CVC we compared a network that combines a CNN and an LSTM network to our previous methods of using a CNN alone. We showed that this combined network could recognize task transition points more consistently and with less delay when operating in real-time compared to the CNN alone. Central Line Tutor was well received by all participants who found that it was a realistic, useable training platform for central venous catheterization. Once we can reduce the system to a minimal number of components, we believe that it will become a helpful tool for improving access to quality instruction and feedback for medical trainees in the process of learning central venous catheterization.

Acknowledgments

REFERENCES

[1]. Tse A, Schick MA, "Central Line Placement," https://www.ncbi.nlm.nih.gov/books/NBK470286/, Aug. 2020.

[2]. Randolph AG, et al. , "Ultrasound guidance for placement of central venous catheters," Crit. Care Med, vol. 24, no. 12, pp. 2053–2058, Dec. 1996. [PubMed: 8968276]

[3]. Kumar A, Chuan A, "Ultrasound guided vascular access: efficacy and safety," Best Pract. Res. Clin. Anaesthesiol, vol. 23, no.3, pp. 299–311, Sep. 2009. [PubMed: 19862889]

[4]. Barsuk JH, et al. , "Simulation-based mastery learning reduces complications during central venous catheter insertion in a medical intensive care unit," Crit. Care Med, vol. 37, no. 10, pp. 2697–2701, Oct. 2009. [PubMed: 19885989]

[5]. Gurm HS, et al. , "Using Simulation for Teaching Femoral Arterial Access: A Multicentric Collaboration," Catheter. Cardiovasc. Interv, vol. 87, no. 3, pp. 376–380, Feb. 2016. [PubMed: 26489781]

[6]. McIntosh KS, et al. , "Computer-based virtual reality colonoscopy simulation improves patient-based colonoscopy performance," Can. J. Gastroenterol. Hepatol, vol. 28, no. 4, pp. 203–206, Apr. 2014. [PubMed: 24729994]

[7]. Grantcharov TP, et al. , "Randomized clinical trial of virtual reality simulation for laparoscopic skills training," Br J Surg, vol. 91, no. 2, pp. 146–150, Feb. 2004. [PubMed: 14760660]

[8]. McGaghie WC, 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," Acad. Med, vol. 86, no. 6, pp. 706–11, Jun. 2011. [PubMed: 21512370]

[9]. Shelesky G, et al. , "Does weekly direct observation and formal feedback improve intern patient care skills development? A randomized controlled trial," Fam. Med, vol. 44, no. 7, pp. 486–492, Jul.-Aug. 2012. [PubMed: 22791533]

[10]. Yarris L, et al. , "Attending and resident satisfaction with feedback in the emergency department," Acad. Emerg. Med, vol. 16, no. 12, pp. 76–81, Dec. 2009. [PubMed: 18945243]

[11]. Tuck KK, et al. , "Survey of residents' attitudes and awareness toward teaching and student feedback," J. Grad. Med. Educ, vol. 6, no. 4, pp. 698–703, Dec. 2014. [PubMed: 26140121]

[12]. Watling C, "Resident teachers and feedback: time to raise the bar," J. Grad. Med. Educ, vol. 6, pp. 781–782, Dec. 2014. [PubMed: 25512805]

[13]. Twinanda AP, et al. , "RSDNet: learning to predict remaining surgery duration from laparoscopic videos without manual annotations," IEEE Trans Med Imaging, Dec. 2018.

[14]. Jin Y, et al. , "SV-RCNet: workflow recognition from surgical videos using recurrent convolutional network," IEEE Trans Med Imaging, vol. 37, no. 5, pp. 1114–1126, Dec. 2017.

[15]. Yengera G, et al. , "Less is more: surgical phase recognition with less annotations through self-supervised pre-training of CNN-LSTM networks," arXiv:1805.08569, May 2018.

[16]. Sutskever I, et al. , "Sequence to sequence learning with neural networks," Adv Neural Inf Process Sys, pp. 3104–3112, Sep. 2014.

[17]. Lasso A et al. "PLUS: open-source toolkit for ultrasound-guided intervention systems," IEEE Trans Biomed Eng, vol. 61, no. 10, pp. 2527–37, Oct. 2014. [PubMed: 24833412]

[18]. Ungi T, et al. , "Perk Tutor: An Open-Source Training Platform for Ultrasound-Guided Needle Insertions," IEEE Trans Biomed Eng, vol. 59, no. 12, pp. 3475–3481, Dec. 2012. [PubMed: 23008243]

[19]. Hisey R, et al., "Real-Time Workflow Detection Using Webcam Video for Providing Real-Time Feedback in Central Venous Catheterization Training," SPIE Medical Imaging 2018, Mar. 2018.

[20]. Holden MS, et al. , "Overall Proficiency Assessment in Point-of-Care Ultrasound Interventions: The Stopwatch is not Enough," BIVPCS 2017, POCUS 2017, pp. 146–153, 2017.

[21]. He K, et al., "Deep residual learning for image recognition," CVPR, pp. 770–778, Jun. 2016.

[22]. Dalal N, Triggs B, "Histograms of oriented gradients for human detection," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 886–893, Jul. 2005.
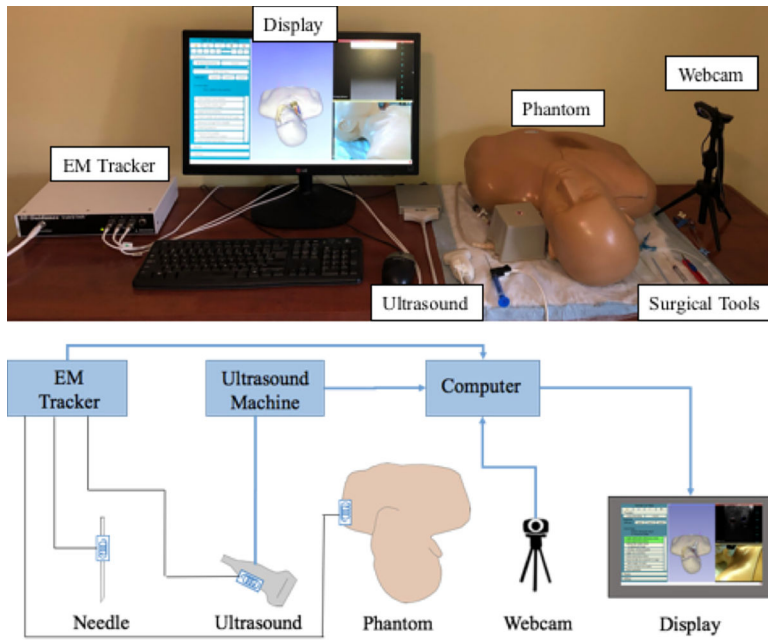
Figure 1.
Central Line Tutor system setup

(a)　　　　　　　　　　　(b)　　　　　　　　　　　(c)

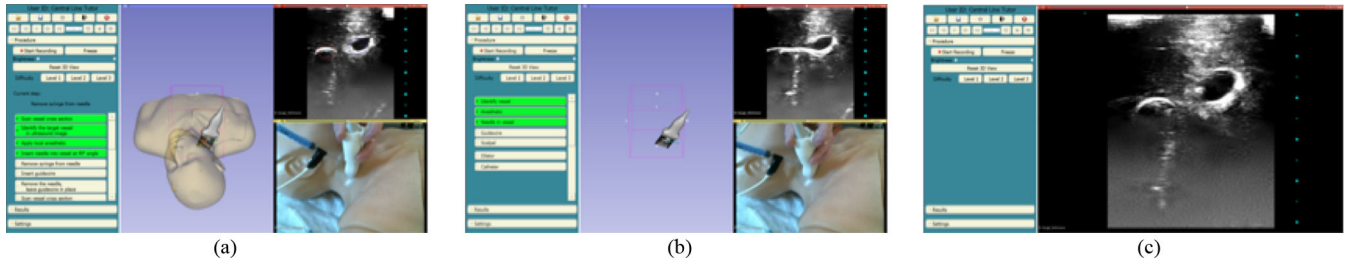**Figure 2.**
Central Line Tutor user interface, showing visual aids for 3 difficulty levels. (a) Level 1, (b) Level 2, and (c) Level 3
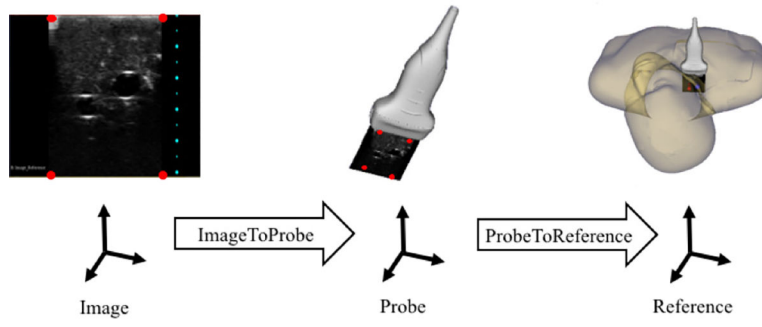
Figure 3.
Coordinate transform from ultrasound image coordinate system to reference coordinate system
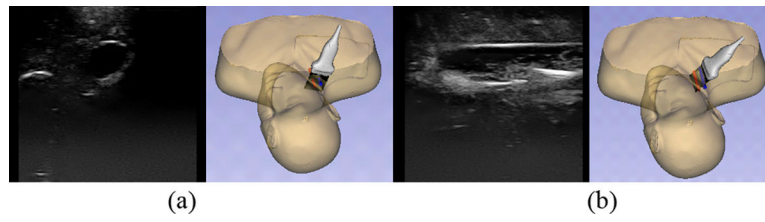
Figure 4.
Ultrasound scans showing both the position of the ultrasound probe and its corresponding ultrasound image. (a) Cross sectional scan, (b) Long axis scan
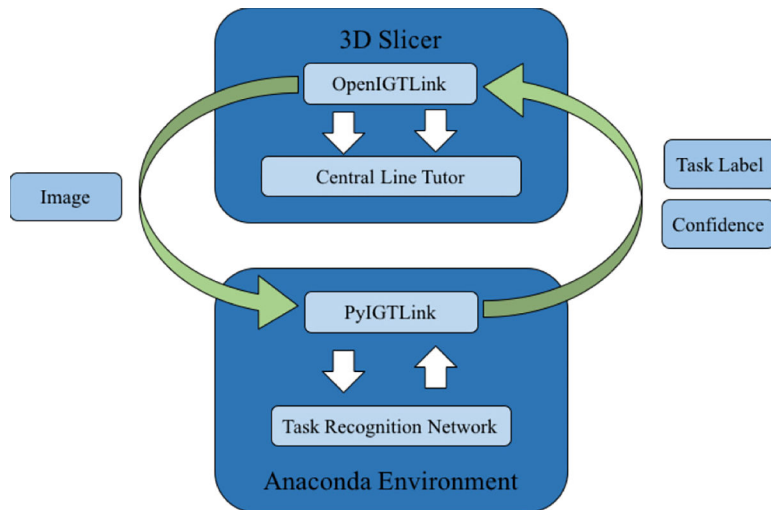
Figure 5.
Illustration of communication between Central Line Tutor, operating in 3D Slicer, and the task recognition network running in a virtual environment
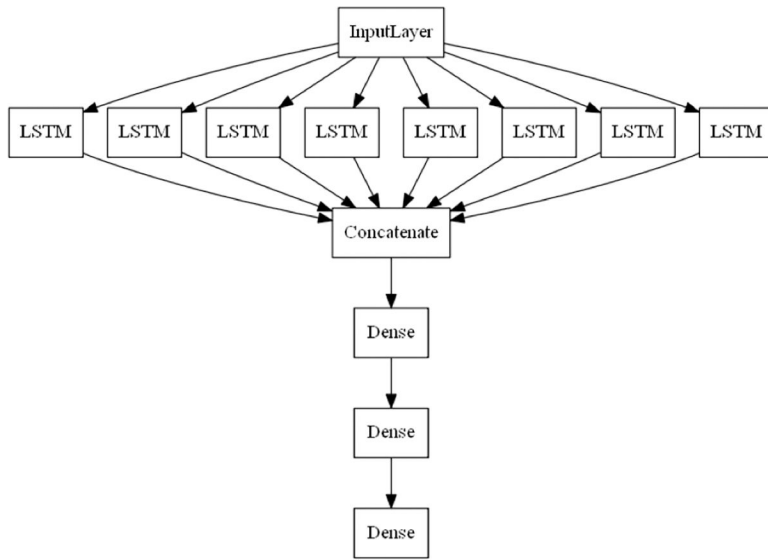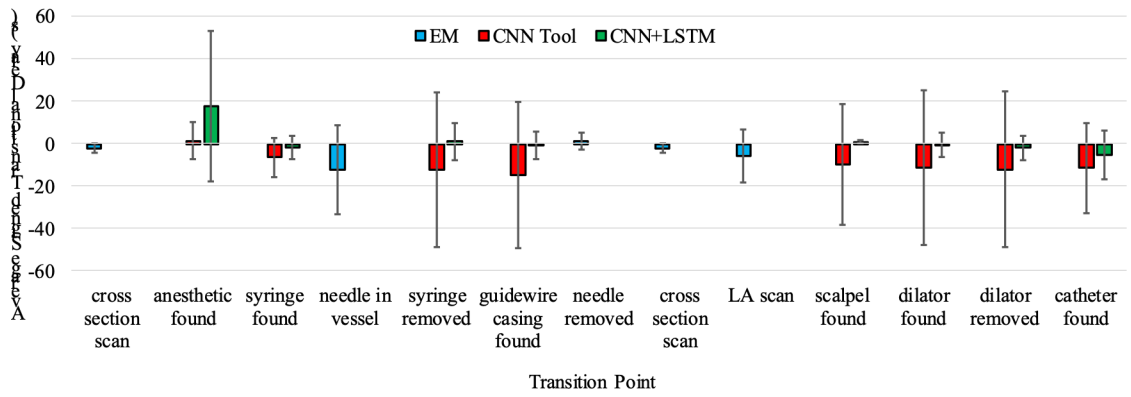
Figure 6.
Parallel LSTM network structure

Figure 7.
Average signed transitional delay results for each transition point. Each transition point corresponds with the start of a single task in the CVC workflow.

Realism:
1. Global impression of realism
2. Ultrasound visualization of tissue and vascular space
3. Training model
4. Handling of equipment

Usefulness:
1. Global impression of usefulness
2. Improvement of ultrasound-handling efficiency
3. Improvement of needle insertion
4. In training residents
5. Individualized learning
6. Performance monitoring
7. Improving efficiency of assessment
8. Improving quality of assessment
9. Facilitating preparedness

Useful components:
1. 3D models
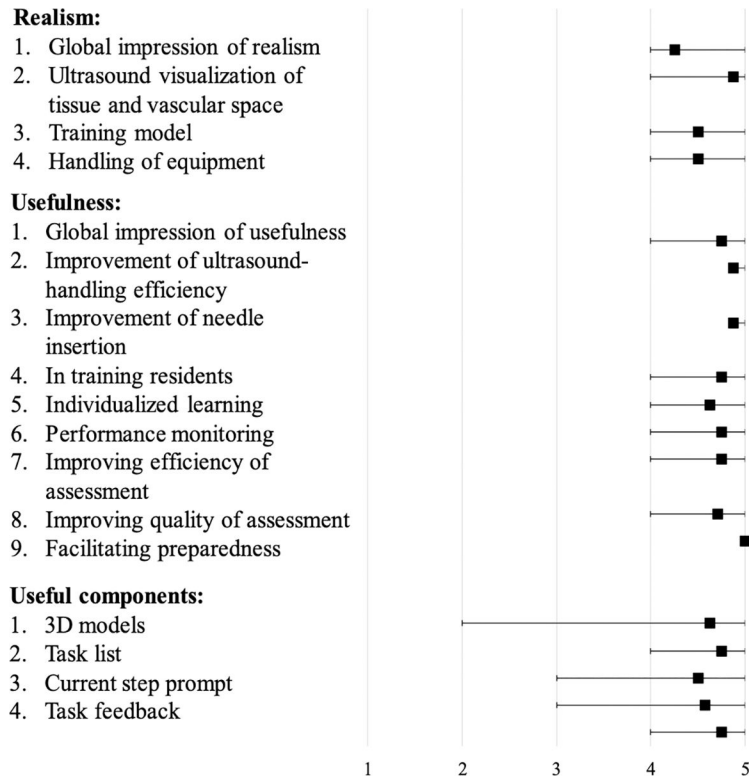2. Task list
3. Current step prompt
4. Task feedback

Figure 8.
Results of participant survey

TABLE I

Breakdown of number of images per class for task and tool labels

Tool Labels		Task Labels	
Label	**Image count**	**Label**	**Image count**
Anesthetic	2084	Apply Anesthetic	2179
Syringe	16282	Insert needle	16288
Guidewire casing	10207	Insert Guidewire	10845
Scalpel	1045	Use Scalpel	1164
Dilator	2204	Dilate Opening	2215
Catheter	13336	Insert Catheter	10910
Guidewire	32384	Remove Guidewire	1584
No Tool	18264	No Task	50621

Table II

List of tasks in the central venous catheterization workflow, along with their corresponding transition points

Task	Transition Point	Recognition Method	Total Count
Scan vessel cross section	Cross section scan 1	EM	39
Apply local anesthetic	Anesthetic found	Webcam	39
Insert needle into vessel	Syringe found	Webcam	40
	Needle in vessel	EM	40
Remove syringe from needle	Syringe removed	Webcam	40
Insert guidewire	Guidewire casing found	Webcam	40
Remove needle	Needle removed	EM	40
Scan vessel cross section	Cross section scan 2	EM	39
Scan vessel long axis	Long axis scan	EM	39
Cut skin with scalpel	Scalpel found	Webcam	39
Insert dilator	Dilator found	Webcam	39
Remove dilator	Dilator removed	Webcam	39
Insert catheter	Catheter found	Webcam	40

Table III

Percentage of tasks and transition points recognized by Central Line Tutor

Task	Transition Point	EM & CNN$_{Tool}$	EM & CNN+LSTM
Scan vessel cross section	Cross section scan 1	89.7%	89.7%
Apply local anesthetic	Anesthetic found	71.8%	66.7%
Insert needle into vessel	Syringe found	100%	100%
	Needle in vessel	97.5%	97.5%
Remove syringe from needle	Syringe removed	100%	100%
Insert guidewire	Guidewire casing found	97.5%	95.0%
Remove needle	Needle removed	100%	100%
Scan vessel cross section	Cross section scan 2	100%	100%
Scan vessel long axis	Long axis scan	94.9%	94.9%
Cut skin with scalpel	Scalpel found	94.9%	43.6%
Insert dilator	Dilator found	79.5%	66.7%
Remove dilator	Dilator removed	79.5%	66.7%
Insert catheter	Catheter found	100%	100%