AWARD NUMBER: W81XWH-14-1-0042

TITLE: See-What-I-Do: Increasing mentor and trainee sense of co-presence in trauma surgeries with the STAR platform

PRINCIPAL INVESTIGATOR: Juan P Wachs

RECIPIENT: Purdue University, 315 N. Grant Street, West Lafayette, 47907, IN

REPORT DATE: Apr-2020

TYPE OF REPORT: Annual Report

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
April 2020	Annual	3 Mar 2019 - 2 MAR 2020

4. TITLE AND SUBTITLE

See-What-I-Do: Increasing Mentor and Trainee Sense of Co-Presence in Trauma Surgeries with the STAR Platform

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-14-1-0042

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Juan P Wachs

Email: jpwachs@purdue.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
.
Purdue University
West Lafayette, IN 47907

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Our primary research objectives are to design, implement, and evaluate a working prototype that enables effective telementoring of a trainee surgeon by a remote mentor. This includes (1) a trainee-site subsystem for augmenting the view of the actual surgical field seamlessly by using a transparent display with illustrations of the current and next steps of the procedure, and (2) a mentor-side patient-size interaction platform with a gesture-based interface.

15. SUBJECT TERMS-

None Listed

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
U	U	U	UU	53	19b. TELEPHONE NUMBER *(include area code)*

TABLE OF CONTENTS

Purpose: Develop a framework that will enable increasing the mentor and trainee sense of co-presence through augmented visualization to facilitate surgical training and performance.

Scope: Optimal trauma treatment integrates different surgical skills not all available in military field hospitals. Telementoring can provide the missing expertise, but current systems require the trainee to shift focus frequently from the operating field to a nearby telestrator, they fail to illustrate the next surgical steps, and they give the mentor an incomplete picture of the ongoing surgery. We are addressing these gaps by developing STAR – System for Telementoring with Augmented Reality.

Major Findings: This year's main focus was to document all the steps for proper installation and execution of the system, as well as completing extra features of our telementoring platform. With respect to documenting the platform, manuals explaining how to download, install, run and use our telementoring platform were written and distributed. With respect of the extra technical developments done over the system, three major components were created: 1) an approach to visualize the vital signs of the patient receiving medical assistance in the augmented reality head-mounted display; 2) an architecture to communicate the remote mentor with the image of an ultrasound device being used by the local mentee; 3) an Artificial Intelligence module to give surgical guidance when no remote mentor is available or the network communication gets compromised. Additionally, the stabilization routines from our system were analyzed more extensively with a new user study. Finally, we validated of our platform as means to provide surgical coaching and confidence in surgery residents and medical students.

1. **INTRODUCTION:** Narrative that briefly (one paragraph) describes the subject, purpose and scope of the research.

Our primary research objectives are to design, implement, and evaluate a working prototype that enables effective telementoring of a trainee surgeon by a remote mentor. This includes (1) a trainee-site subsystem for augmenting the view of the actual surgical field seamlessly by using a transparent display with illustrations of the current and next steps of the procedure, and (2) a mentor-side patient-size interaction platform with a gesture-based interface.

2. **KEYWORDS:** Provide a brief list of keywords (limit to 20 words).

Augmented reality, telementoring, telemedicine, computer vision, future-steps visualization, surgical training, co-presence, simulation, tele-existence.

3. **ACCOMPLISHMENTS:** The PI is reminded that the recipient organization is required to obtain prior written approval from the awarding agency Grants Officer whenever there are significant changes in the project or its direction.

What were the major goals of the project?
List the major goals of the project as stated in the approved SOW. If the application listed milestones/target dates for important activities or phases of the project, identify these dates and show actual completion dates or the percentage of completion.

Specific Aim 3:
STAR specialization for cric in austere environments (03-Set-2017– 03-Mar-2018) 100%
Experimental Design 4: austere environment validation (03-Mar-2018– 02-Apr-2020) 100%

Specific Aim 4:
STAR specialization for fasciotomy on a cadaveric leg (03-Mar-2018– 02-Apr-2020) 100%
Experimental Design 5: Validate STAR in fasciotomies (03-Mar-2017 – 03-Mar-2018) 100%

What was accomplished under these goals?
For this reporting period describe: 1) major activities; 2) specific objectives; 3) significant results or key outcomes, including major findings, developments, or conclusions (both positive and negative); and/or 4) other achievements. Include a discussion of stated goals not met. Description shall include pertinent data and graphs in sufficient detail to explain any significant results achieved. A succinct description of the methodology used shall be provided. As the project progresses to completion, the emphasis in reporting in this section should shift from reporting activities to reporting accomplishments.

Major Activities: ***Research, develop, and assess a transparent-display augmented-reality system that allows the seamless enhancement of a trainee surgeon's natural view of the surgical field with annotations and illustrations of the current and next steps of the surgical procedure.***

Specific Objectives

Task 3.1- Specialize the system for a cric procedure on a patient simulator in an austere environment

Tracking of ultrasound device for mentee-to-mentor pose transmission

As part of the most recent technical developments integrated to the STAR platform, an approach to acquire the image from a portable ultrasound device was developed. In the previous report, we reported on our approach to stream the image of the ultrasound device to the mentor site. The image from the device was displayed in the Mentor System's screen, updating in real-time.

As the final part of this module, an approach to update the position of the ultrasound image in the Mentor System based on the position of the ultrasound on the mentee's site was developed. As reported in the previous report, the real-time pose of the ultrasound probe is acquired by the Mentee System using a 3D-printed structure with ArUco markers. The HoloLens' on-board camera captures the ultrasound probe, and these video frames are fed to a computer vision module that detects if the current video frame contains the pattern of markers that represents the ultrasound probe. The resulting detected markers in the frame can be unprojected into a 3D position and orientation relative to the camera. We combine this with the world-space pose of the camera itself to determine the world-space pose of the tracker.

This pose is streamed to the Mentor System through JSON messages, where it is unpacked and used to calculate the position of the probe in the image of the mentor site. The final position of the probe is calculated by combining the ultrasound probe position and rotation matrix with the image stabilization routines of the Mentor System. This last step is performed to account for the image distortion introduced in the Mentor System to make the image look as stable as possible with the respect to the mentee's point of view. Figure 1a showcases the mentee using the portable ultrasound device to scan the leg of a patient. The designed probe with the ArUco markers is presented in Figure 1b. Figure 1c showcases the view in the Mentor System of the current ultrasound slice being scanned. A close-up of the ultrasound image provided to the mentor is presented in Figure 1d.

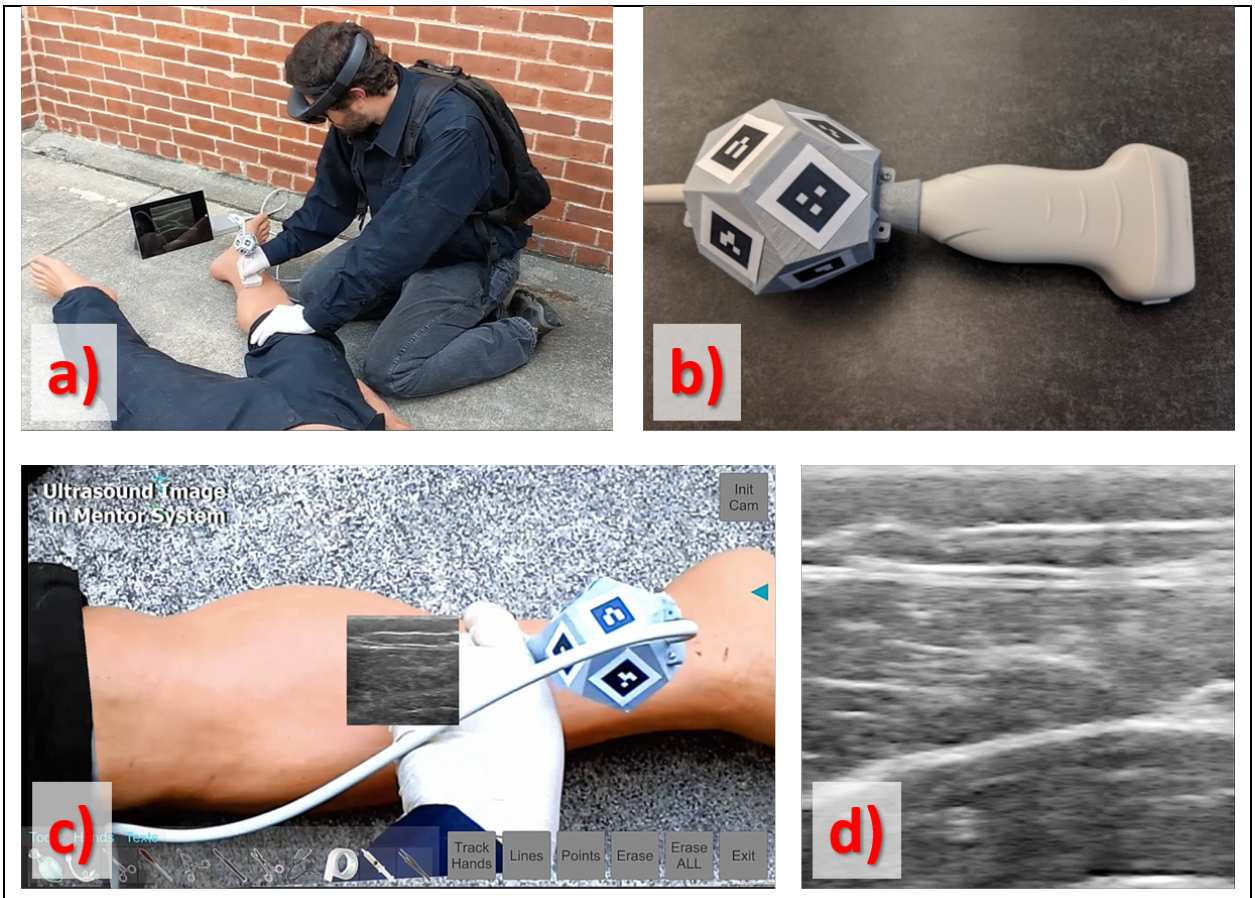


Figure 1. Portable ultrasound visualization, streaming and tracking module.

Figure 2 showcases an example of how the image of the ultrasound in the Mentor System changes with respect to the position of the ultrasound probe. The position of the tip of the probe is obtained from tracking the 3D-printed tracker and is then calculated in the image of the Mentor System for its correct visualization and placement.

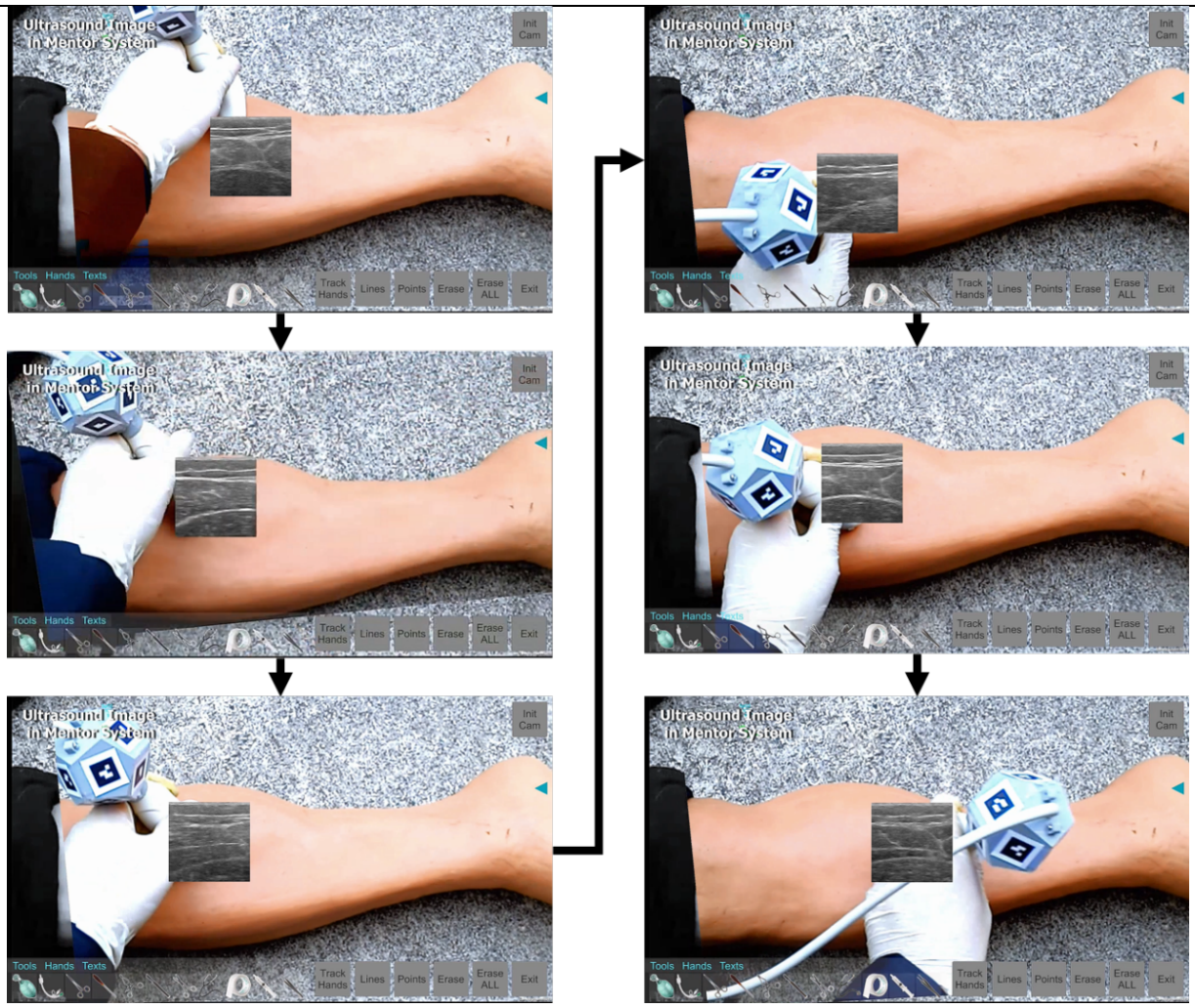


Figure 2. Example of the real-time tracking of the ultrasound probe, visualized from the Mentor System's perspective.

Transmission of patient vital signs data from mentee site to Mentor System

In the last annual report, we commented on the integration of different vital signs sensor to our telementoring system. With this approach, the different vital signs are visualized as a "virtual monitor" for the mentee wearing the head-mounted display. This virtual monitor is anchored to a world position in the mentee's workspace, allowing to visualize it by shifting the focus when necessary. The values displayed in the virtual monitor are updated in real-time, based in the readings from vital signs sensors.

As a final step in this module, the values acquired from the vital signs are transmitted to the Mentor System to be visualized by the remote expert. The visualization of these signals is anchored to the Mentor System and presented in the same format in which it is presented to the mentee. Figure 3 presents an example of the vital signs of a patient being visualized in the Mentor System. Figure 4a showcases the vital signs values being obtain from the platform's oximeter sensor. The acquired values are transmitted through the internet and through Bluetooth, to be visualized in 2D the Mentor System (Figure 4b) and in 3D in the Mentee System (Figure 4c), respectively.

Figure 3. Example of the visualization of the vital signs in the Mentor System.

Figure 4. Vital signs acquisition, transmission and visualization module.

Recording a new video to showcase STAR's new developments

To summarize the latest technical developments integrated to our telementoring system, a video showcasing them in a simulated austere scenario was recorded. The video encompassed the various modules of our telementoring platform, including the real-time ultrasound tracking and visualization (Figure 5a), the acquisition and visualization of the patient's vital signs (Figure 5b), the projection over the patient's body of 3D virtual objects representing the mentor-

authored surgical instructions (Figure 5c), and the image stabilization routines implemented at the Mentor System (Figure 5d). The video is available online at https://www.youtube.com/watch?v=4mtx8869Hi8&t.

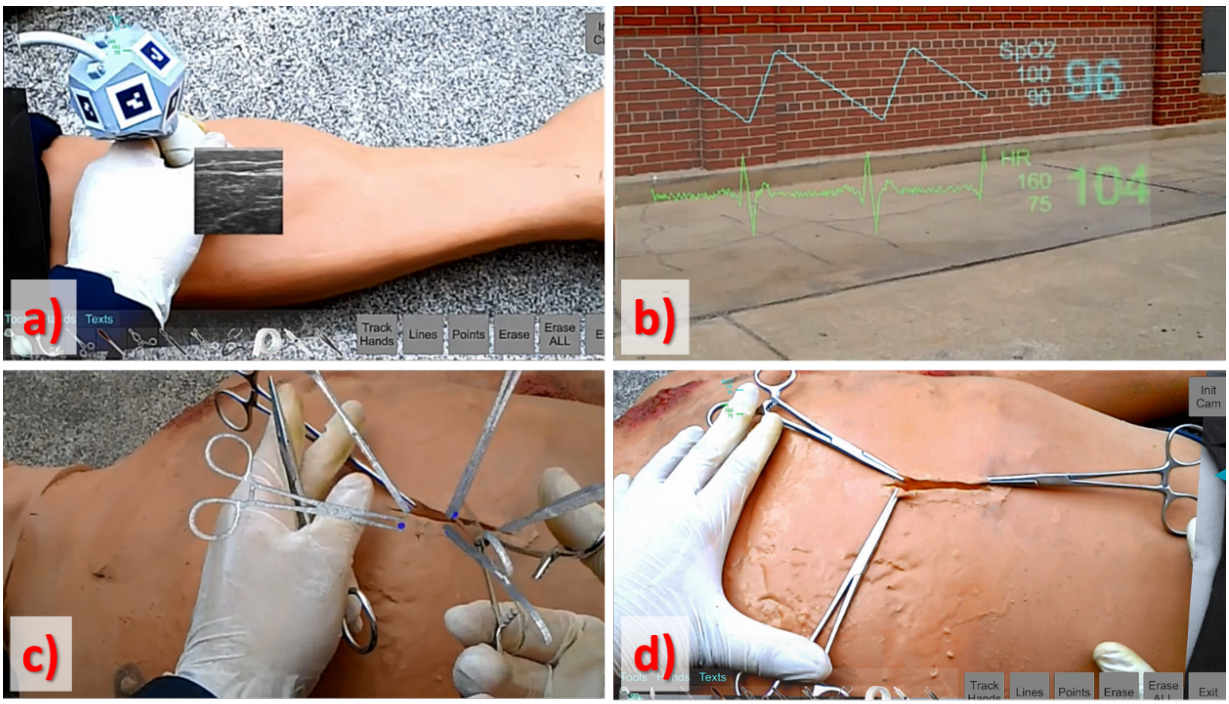


Figure 5. Snapshots of the latest video showcasing the most recent developments in the STAR platform.

Project Documentation and Installation Guide

As part of the dissemination efforts and conde handoff, an installation guide was created. The installation guide includes a detailed, step-by-step explanation of the steps to: 1) download the code from the GitHub repositories; 2) install the code and dependencies in a PC and Microsoft HoloLens device; 3) run each of the subsystems of our telementoring platform; and 4) connect the subsystems between them. Figure 6 presents and schematic of the different modules that were documented and will be explain in the system's installation guide. Additionally, Figure 7 presents screenshots of different sections of the document. The document can be downloaded using the following link: https://engineering.purdue.edu/starproj/wp-content/uploads/STAR_Installation_Guide.pdf.

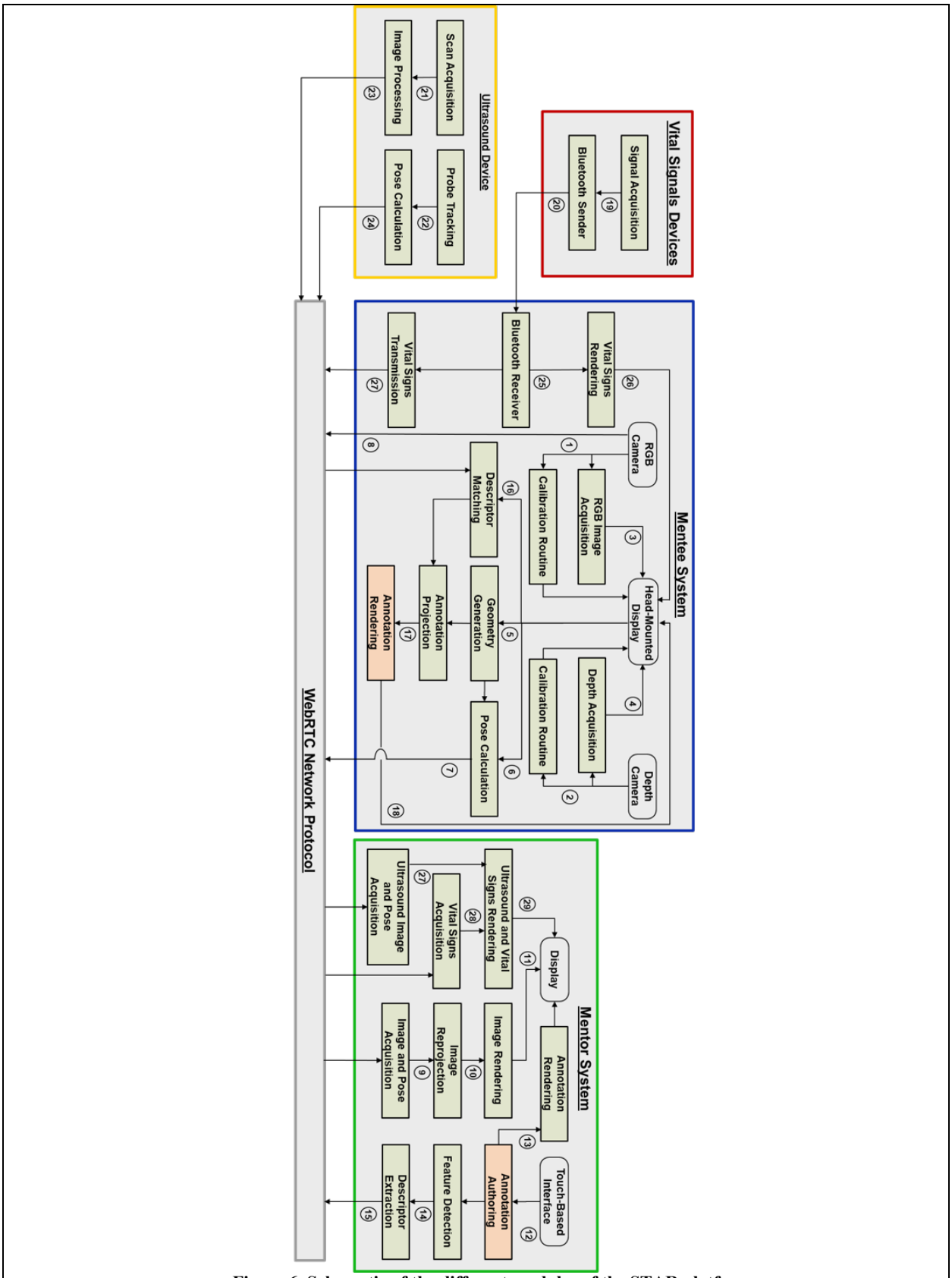


Figure 6. Schematic of the different modules of the STAR platform.

Contents

Introduction

The System for Telementoring with Augmented Reality (STAR; https://engineering.purdue.edu/starproj/) is a novel platform that conveys surgical instructions remotely. Through the platform, mentees requiring assistance can receive surgical instructions from remotely located mentors. The platform leverages an Augmented Reality (AR) Head-Mounted Display (HMD; Microsoft HoloLens, Microsoft Corporation, Washington, USA) worn by a mentee surgeon to display mentor-authored surgical instructions. Mentees wearing the ARHMD can visualize these expert instructions as three-dimensional (3D) overlays directly onto their field of view of the patient's body. Figure 1 presents a schematic of our platform.

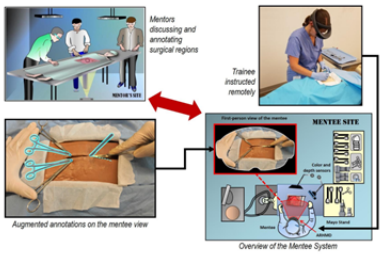


Figure 1. Schematic of the STAR platform. By integrating image stabilization routines with the ability to project mentor-authored surgical instructions onto the operating field, this portable system is designed to convey clinical expertise in both controlled and austere scenarios.

The STAR platform is divided into two subsystems: the Mentee System, leveraging an ARHMD at the mentee site; and the Mentor System, leveraging a computer with a touch-based interface at the mentor site. Figure 2 presents a step-by-step technical description of our platform. First, the Mentee System undergoes initial calibration routines to align the ARHMD view with the operating field (1 and 2). The device's on-board cameras (3 and 4) then acquire the color (RGB) and depth views of the operating field. The ARHMD generates an approximation of the geometry of the operating field based on the RGB and depth images (5). This geometry, along with internal sensors, is then used to calculate the ARHMD's pose (6) with respect to the operating field. The pose and the RGB image are sent to the Mentor System using the WebRTC web socket protocol (7 and 8).

Required Hardware and Software

The STAR platform requires the following components to be installed:

On the mentor site, the platform requires:

- A computer with Windows 10, version 10.0.17763 or equivalent. Additionally, this computer needs to have a touch interface. This can be either an integrated touch display, or an additional touch display connected to the computer.
- The Mentor System codebase, located at: github.com/edkazar/MentorSystemUWPWebRTC.

On the mentee site, the platform requires:

- A Microsoft HoloLens, version 1.
- Either:
  - An Android smartphone.
  - A computer with Windows 10, version 10.0.17763 or equivalent.
- The Mentee System codebase, located at: www.github.com/practisebody/STAR.
- The STAR Controller codebase, located at: www.github.com/practisebody/STARController_UWP.

Installation Steps

The installation steps are also found in respective GitHub repositories, as well as an explanation of how to compile the codebases from scratch.

Mentee System Installation:

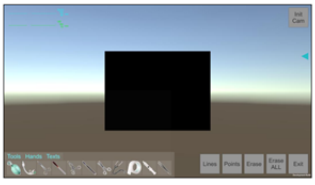
Follow the instruction in https://docs.microsoft.com/en-us/windows/mixed-reality/using-the-windows-device-portal to set up Windows Device Portal for the HoloLens.

2. Connect to HoloLens
   i. Using Windows Device Portal to connect the PC to the HoloLens, either over Wi-Fi or over USB

3. Upload to HoloLens. *Note: This part might be not accurate if Windows Device Portal updates.*
   i. Using Windows Device Portal, click on "Views > Apps > Deploy apps"

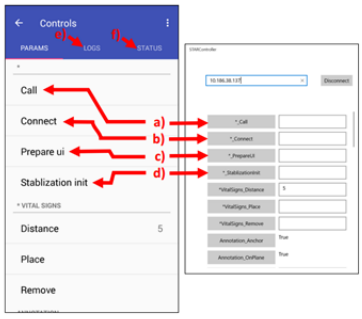
2. After you start the app for the first time, make sure to accept all permissions the app requires to function (Camera, Microphone, Internet).

3. After you accepted all the permissions, try clicking the "WebRTC" button. If the background changes to look like the following image and the WebRTC button disappears, the installation was successful.



STAR Controller App:

This subsection shows both the Android and PC apps functionalities. The images presented are screenshots obtained after scrolling down in the Apps.



a) Call button: Once both the Mentee and Mentor Systems are connected to each other, this button starts a WebRTC call between the two systems.
b) Connect button: Connects the Mentee System to the WebRTC server.
c) Prepare UI button: Removes elements of the User Interface, in preparation for an incoming connection with the Mentor System. This function is also automatically called whenever the Call button is pressed.
d) Stabilization Init button: Once both the Mentee and Mentor Systems are in call, this button starts the stabilization routines in the Mentor System.
e) Logs Tab: Debugging information related to the App's usage will be stored in this tab.
f) Status Tab: Debugging information related to the App's connection to the WebRTC server will be stored in this tab.

Figure 7. Portable ultrasound visualization, streaming and tracking module.

Additionally, the code and installation steps of each of the platform's subsystems are online in GitHub repositories. The following repositories can be used to acquire our platform's subsystems:

- Mentor System: https://github.com/edkazar/MentorSystemUWPWebRTC/
- Mentee System: https://github.com/practisebody/STAR/
- STAR Controller App:
  - PC Version: https://github.com/practisebody/STARController_UWP/
  - Phone Version: https://github.com/practisebody/STARController_Android/

Figure 8 presents a screenshot of one of the GitHub repositories containing our platform's code.

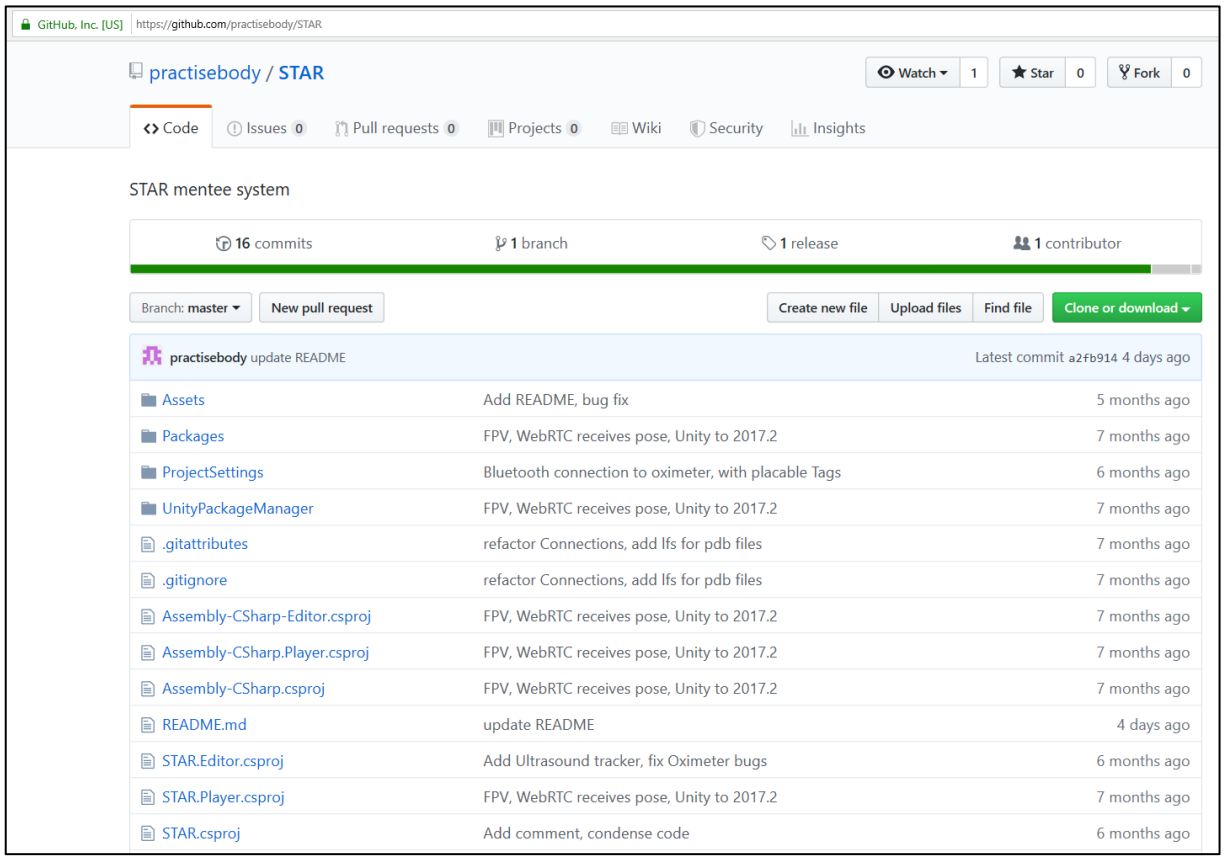


Figure 8. Screenshot of the Mentee System's GitHub repository.

Development of a telementoring platform based on an Augmented Reality Head-Mounted Display

In the previous annual report, we presented the stabilization algorithm leveraged by our Augmented Reality Head-Mounted Display (ARHMD) based telementoring platform to provide the remote mentor with a stabilized view of the mentee's operating field. To refine the algorithm even further, we performed a new user study evaluating our ARHMD-based platform. The objective of the user study was to evaluate if the stabilized view provided by our platform was able to provide an observer with more situation awareness of a workspace without incurring into

additional motion sickness. To perform this evaluation, we recorded the view of a workspace using the on-board camera of our ARHMD platform. Afterwards, we applied the stabilization routines of our platform to this video feed, as saved the result as a video. The stabilized video serves as a visualization of the workspace for a remote collaborator. The original, unstabilized view of the workspace was also saved into a video, and was used as the control condition of our user study. We tested the effectiveness of workspace visualization by asking participants to find matching numbers in both the original and the stabilized videos, for two different workspaces.

Participants

We recruited participants (n = 20, 5 female) from the graduate student population, in the 24 - 30 age group. We opted for a within-subject design, with each participant performing the task in all conditions.

Task

Each participant was seated 2m away from an LCD monitor with a 165cm diagonal. The monitor displayed a video of a workspace annotated with numbers, and the participants were asked to find pairs of matching numbers. When a participant spotted a matching pair, they were asked to call the number out loud. An experimenter, who knew all the matching pairs in each workspace tallied the number of correct matching pairs found by the participants. Numbers called out by the participants that were not part of a matching pair were not counted towards the participant's final score. Figure 9 provides a view of a participant performing one of the tasks.

Figure 9. Participant acting as the mentor, reading number from the acquired workspace video.

Workspaces

Two workspaces were leveraged to perform our user study. To acquire the video of each workspace, an experimenter wore our ARHMD and walked around the workspace, recording it with the front-facing camera of the device (Figure 10). More specifically, the experimenter (who was impersonating a mentee wearing the ARHMD) started out at a default position, then panned the view by rotating the head, and finally moved to the side of the workbench to see it a different perspective (rotated by 90° from the original position). This resulted in a video sequence where the mentee's workspace moved considerably.

Numbers were added to the workspaces using pieces of paper, all facing the mentee in the initial position. Each workspace had 24 numbers. 8 of these numbers were unique, and 16 of these numbers were part of a pair (for a total of 8 pairs). These numbers remained in their initial position of the entire video recording. Nonetheless, some elements of the workspace were relocated during the video recording to introduce additional difficulty and clutter. In some cases, the relocation of these elements was large enough to cover or uncover some of the numbers in the workspace. Furthermore, as the mentee viewpoint translated, some of the numbers would appear and disappear due to occlusions.

The first workspace was comprised of a workbench cluttered with tools (Workbench). In some cases, the clutter was piled up to 30cm above the workbench surface. The second workspace was an engine mounted on a wooden base (Engine). The engine had a height of 80cm, measured from the floor. Numbers were also placed along the surface of the engine. Figure 11 presents a top-front view of both workspaces, captured with the ARMHD's on-board camera

Figure 10. The experimenter, acting as the mentee, acquiring a workspace with HoloLens.

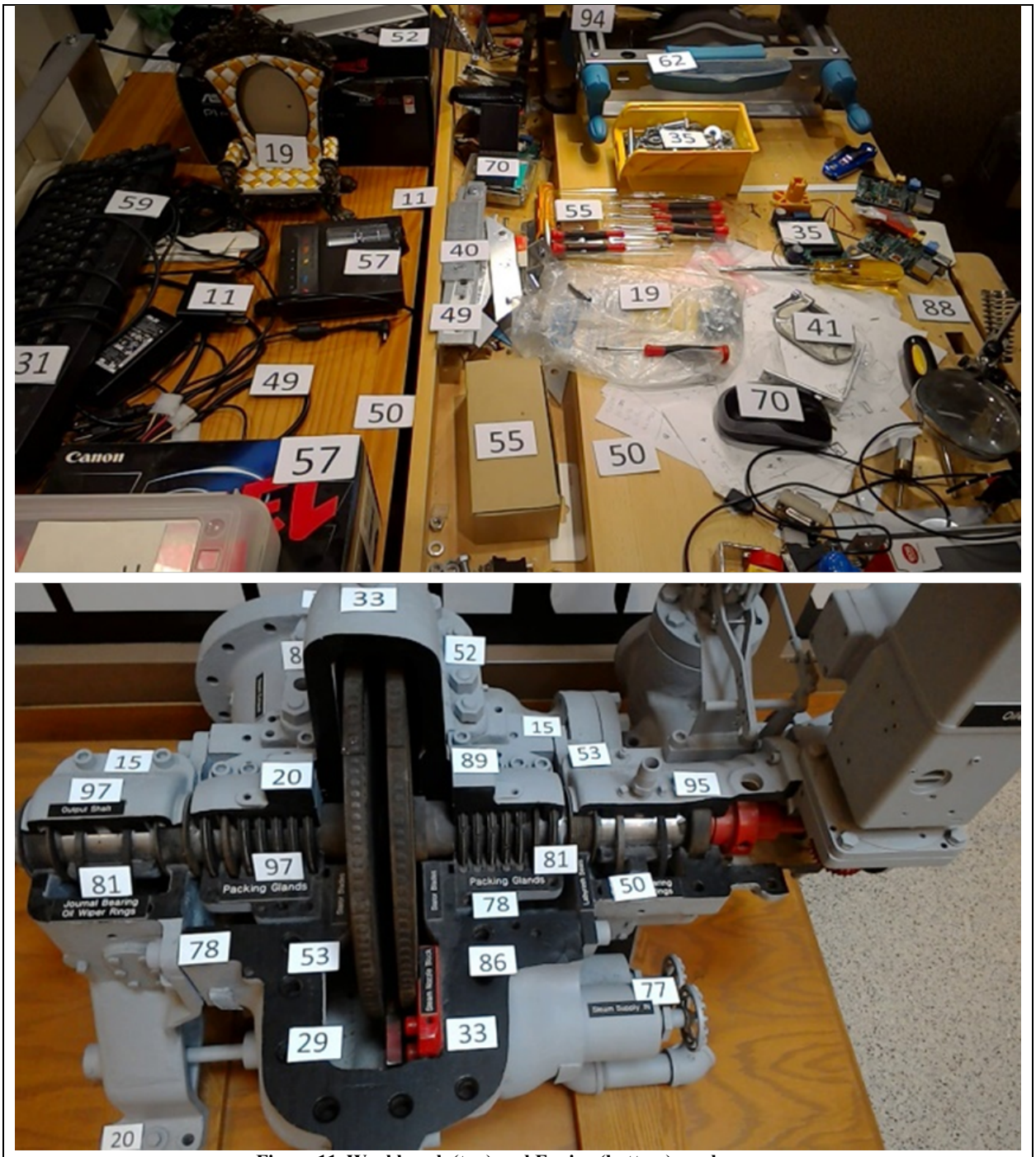


Figure 11. Workbench (top) and Engine (bottom) workspaces.

Conditions

Each of the 20 participants was tested for both the Workbench and the Engine workspaces, and for each of the two conditions (for a total of 4 repetitions per participants). Participants were shown the unstabilized video in the control condition (No Stabilization, NS), and the video after applying the stabilization routines in the experimental condition (Stabilized, S).

Metrics

We measured participant task performance as the number of pairs found. We also measured participant workload using the NASA Task Load Index (NASA-TLX) questionnaire, and participant simulator sickness using the Simulator Sickness Questionnaire (SSQ). Therefore, an overall better performance could be attributed to more matching pairs found, lower cognitive load, or absence of simulator sickness.

Results

A within-subject statistical analysis compared the result of two conditions. The participants and the order of the trials were treated as blocks in the statistical design. The data normality assumption was confirmed with the Shapiro-Wilk test and the data equal-variance assumption was confirmed with the Levene's test. We then ran a repeated measures ANOVA on the data. Table 1 gives the number of pairs found in both workspaces, for each condition. S had a significant advantage for both workspaces. Table 2 compared the NASA TLX scores between the S and NS conditions (i.e. NS-S, as lower NASA TLX scores indicate less demand on the participant). Most S advantages are significant.

Table 1. Comparison between the number of pairs found in the no stabilization (NS) and stabilization (S) conditions. An asterisk (*) indicates a statistically significant difference between conditions.

Workspace	NS	S	S – NS	p-Value
Workbench	5.45±0.83	5.95±1.19	0.50±0.28	0.043*
Engine	5.05±1.57	6.10±1.29	1.05±0.31	0.002*

Table 2. p-values of NASA TLX sub score differences between no stabilization (NS) and stabilization (S) conditions. An asterisk (*) indicates a statistically significant difference between conditions.

Workspace	Mental Demand	Physical Demand	Temporal Demand	Perfor-mance	Effort	Frustration
Workbench	0.000*	0.000*	0.001*	0.188	0.356	0.001*
Engine	0.005*	0.050*	0.000*	0.034*	0.002*	0.001*

Table 3 gives the number of SSQ scores of Workbench and Engine workspaces. S had a significant advantage over NS in terms of Total Severity score, for both workspaces. The S advantage was due to less nausea and oculomotor effort in the Workbench workspace (which was flatter but more cluttered), and due to disorientation in Engine workspace (which had more occlusions and disocclusions). Although the differences between conditions were significant, the Total Severity score did not increase enough to indicate the presence of simulator sickness in any workspace or condition (a difference of 70 between the pre- to post- exposure scores).

Table 3. p-values of SSQ Total Severity score differences between no stabilization (NS) and stabilization (S) conditions. An asterisk (*) indicates a statistically significant difference between conditions.

Workspace	Nausea	Oculomotor	Disorientation	Total Severity
Workbench	0.019*	0.001*	0.116	0.004*
Engine	0.053	0.060	0.019*	0.021*

While this suggests that our stabilization might not induce simulator sickness, and that discomfort levels are similar to those for a perfectly stabilized video, the absence of differences

between S and NS indicates that the exposure might have been too short and or that the workspaces were too simple.

Creation of a database of images and captions of surgical procedures

Telementoring surgeons as they perform surgery can be essential in the treatment of patients when in situ expertise is not available. Nonetheless, expert mentors are often unavailable to provide trainees with real-time medical guidance. When mentors are unavailable, a fallback autonomous mechanism should provide medical practitioners with the required guidance. However, AI/autonomous mentoring in medicine has been limited by the availability of generalizable prediction models, and surgical procedures datasets to train those models with. This subsection presents the initial steps towards the development of an intelligent artificial system for autonomous medical mentoring.

The methodology to create such an AI surrogate mentor includes: 1) the creation of a curated dataset of medical images and their respective step-by-step descriptions; and 2) the use of such a dataset by training a Deep Learning (DL) framework which generates medical instructions from images. This works presents a Database for AI Surgical Instruction (DAISI; https://engineering.purdue.edu/starproj/_daisi/). DAISI provides step-by-step demonstrations of how to perform medical procedures. This is done by including images and text descriptions of procedures from 20 medical disciplines. Each image-text pair describes how to complete a step in the procedure. The database was created via a mobile app containing input from 20 expert physicians from various medical centers, extracting data from academic medical textbooks related to the surgical technique, and acquiring imagery manually. Figure 12 showcases images from four procedures in DAISI.

We describe the process of obtaining medical images and caption from an existing mobile app created by expert physicians. The first step of the process involved searching for existing solutions that targeted part of the problem, such as online videos of medical courses. After exploring, our research team found a free app called Thumbroll. Thumbroll is an app created by the University of Washington School of Medicine, which allows medical students to learn critical techniques and procedures at their own pace through a scrolling functionality. Via simple touch interactions, the user can move between the different steps of a surgical procedure to receive a step-by-step demonstration of how to perform the procedure. Figure 13 showcases a screenshot of the app's main screen.

As showcased in Figure 13, the app is organized in a hierarchical manner: the main screen groups procedures into different categories. For example, under "Training Level", procedures are distributed into "Intern", "Junior Resident", "Clinical MD Trainee", among others. Under "Medical Specialty", procedures are distributed into "Internal Medicine", "Anatomy", "Radiology", among others. Once one of these categories is selected, the app presents the user with all the procedures that fall into this category, as depicted in Figure 14.

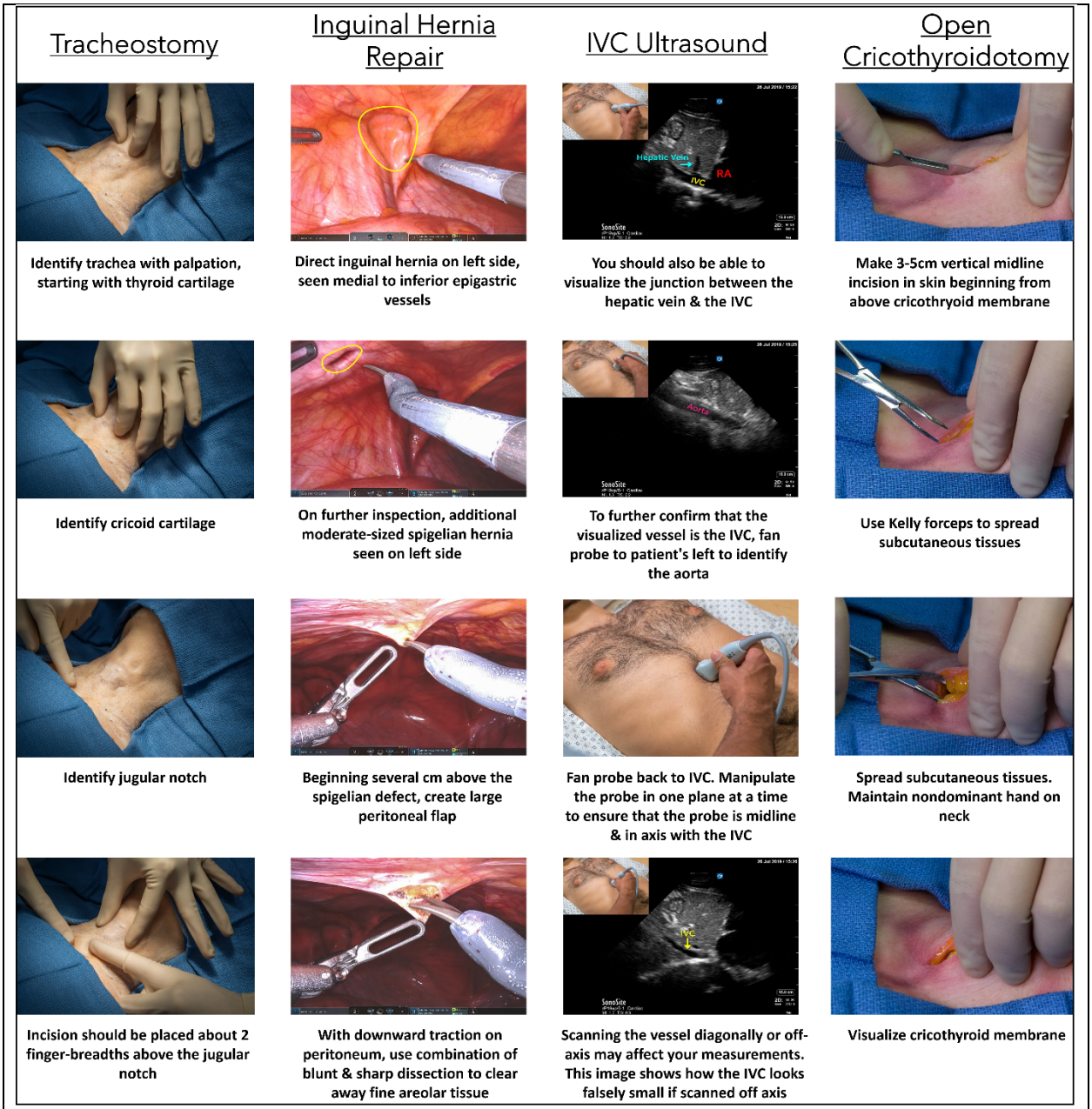


Figure 12. Example of four different images and their associated textual descriptions from four different procedures in the DAISI database. The database includes images from 20 disciplines such as emergency medicine, and ultrasound-guided diagnosis.

Once a procedure is selected, the app shows a step-by-step explanation of how to perform the procedure. The procedure can be navigated using a scrolling touch interaction. Each step is explained with an image and a caption of the procedure, as depicted in Figure 15.

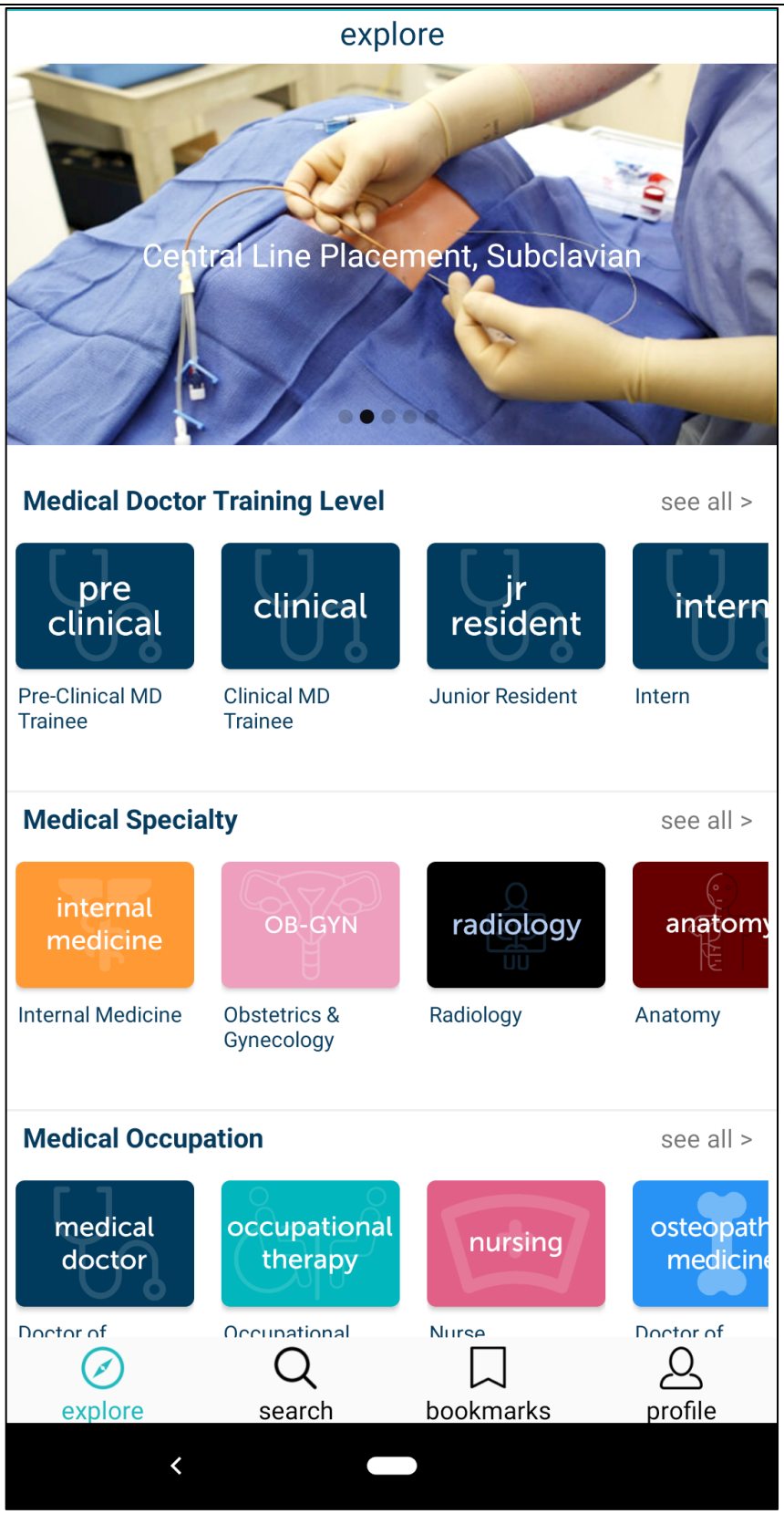


Figure 13. Thumbroll's main screen. Surgical procedures are grouped into different categories to facilitate browsing through the app.

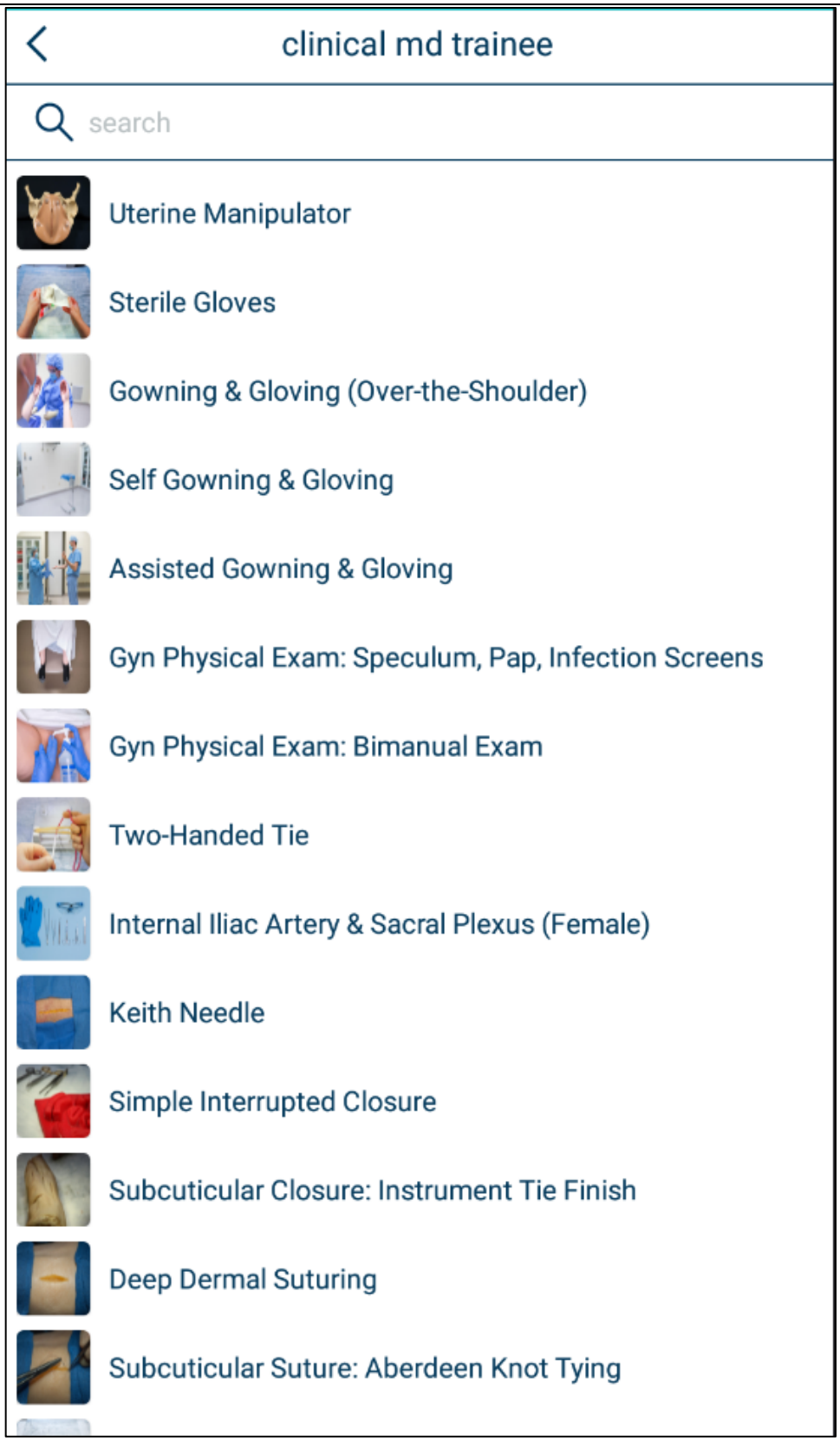


Figure 14. Thumbroll's "Clinical MD Trainee" category. All surgical procedures in this category are related to the training that Clinical MD Trainees need to perform as part of their training.

Figure 15. First step of the "Chest Tube Placement" procedure. The user can scroll through the steps of the procedure using touch commands. Each step is depicted by an image and a caption.

Our approach to create an initial version of the database surgical procedures' images and captions was to acquire the images and captions of all the procedures depicted in the Thumbroll app. In order to mine the data available from the app, we first explored where was the app obtaining the data from. For this purpose, we used Fiddler4, a software to monitor the network traffic inside a local area network. Figure 16 showcases part of the interface of the program. Each time a network request is made, the program acquires it and shows how was the HTTP request packet constructed.

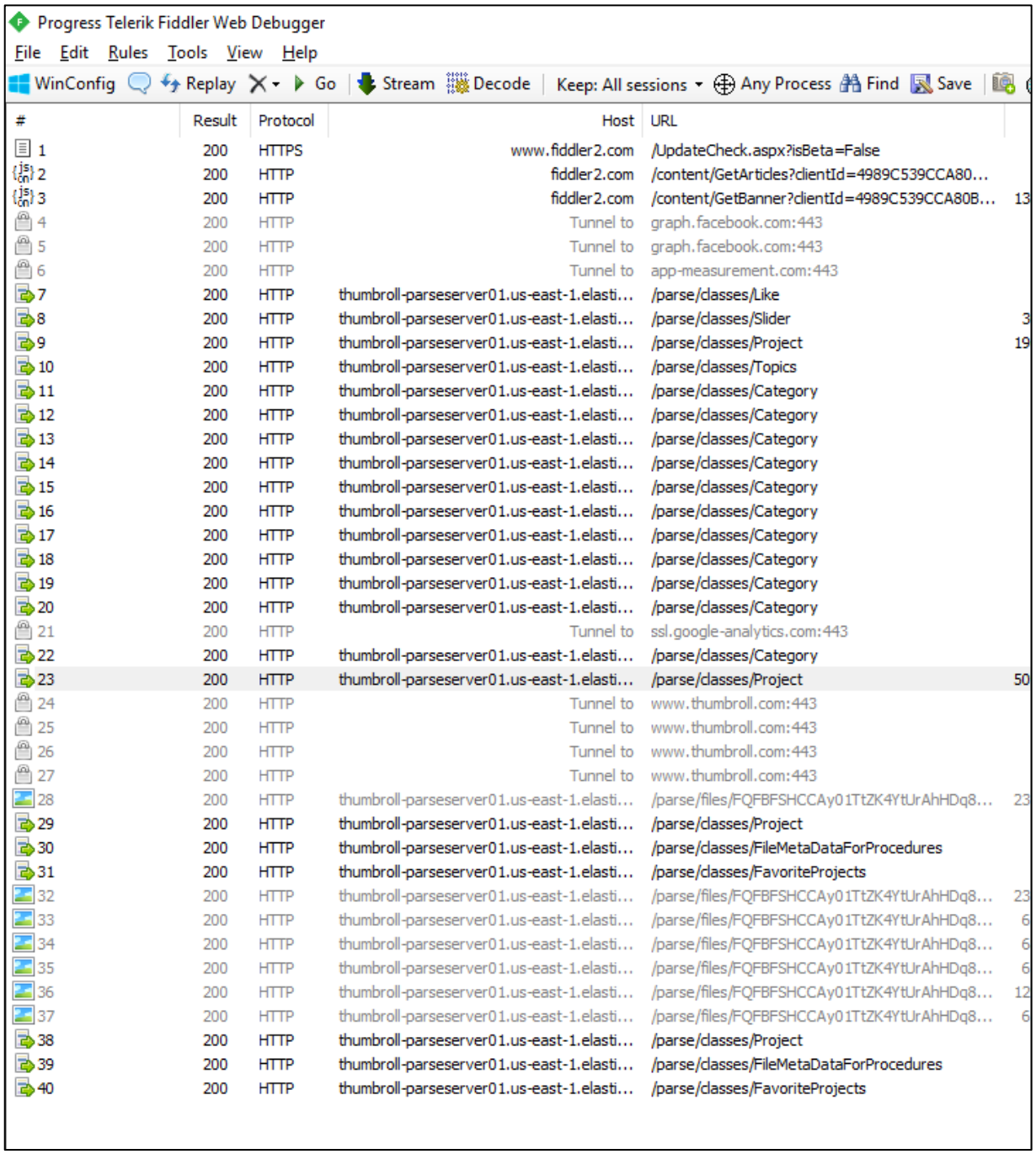


Figure 16. Fiddler4 network traffic software interface. Each line in the picture represents a HTTP request made inside the local computer network.

To successfully monitor the network traffic made by the app (which was running in a mobile cellphone), the software was configured to create a tunnel between the phone and the computer. First, the cellphone was connected to the internet using the computer's hotspot connection. This guaranteed that every HTTP request performed by the cellphone had to go through the computer first. Following this, the IP address of the computer was acquired, and Fiddler4 was configured to catch the requests coming from only a specific port (8888 in our case). Afterwards, the network's proxy was configured to specifically send the HTTP requests using the computer's IP address (and therefore make Fiddler4 think that the computer was sending the request instead of the cellphone), and through the port that Fiddler4 was configured to listen.

Through the setup previously described, each request performed by the Thumbroll app was captured and visualized by Fiddler4. Figure 17 depicts how both a HTTP request from the app and a response obtained from the app's server looked like. This specific request was happened when the "Clinical MD Trainee" category was selected. The result was a packet containing the general information (e.g. name, number of steps, thumbnail image) of all the surgical procedures under that category.

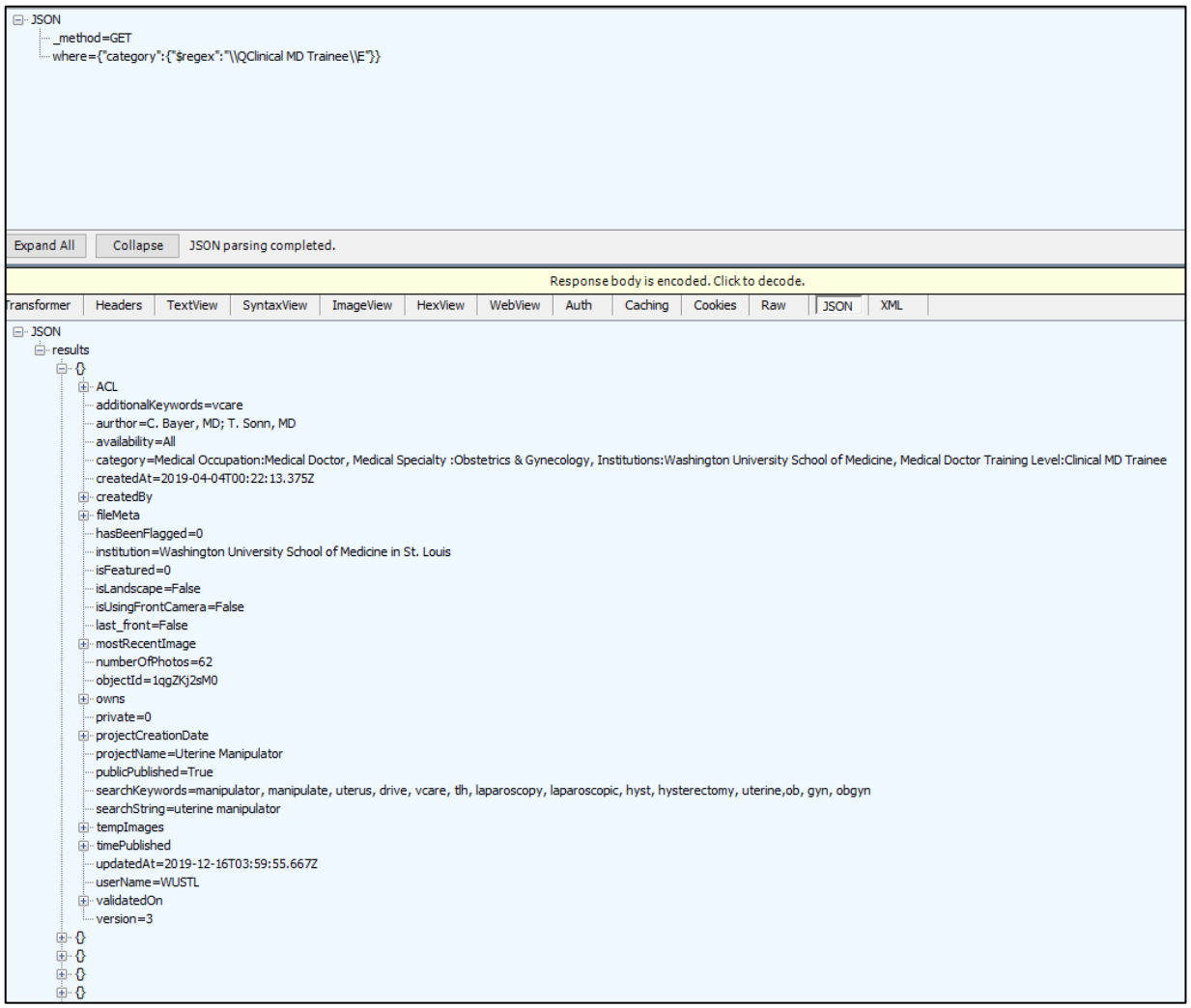


Figure 17. A HTTP request made by the Thumbroll that was captured by the Fiddler4 network traffic software. The app's server responded to the request giving information of all the surgical procedures inside the requested category.

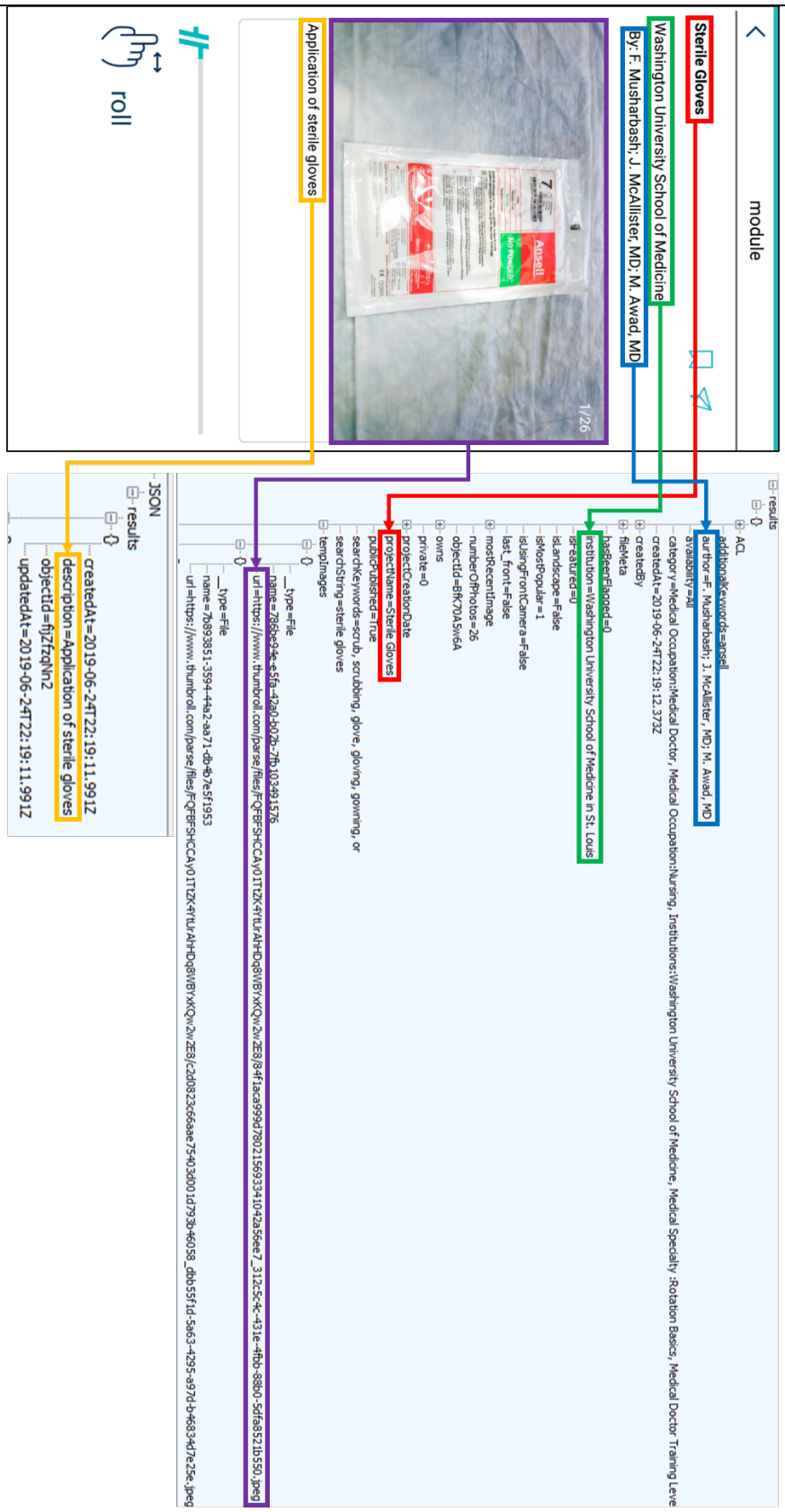


Figure 18. Example of how the information of each procedure can retrieved from the HTTP requests made by the Thumbroll app.

As depicted by Figure 18, a closer inspection of Fiddler4's captured requests revealed that the responses given to the HTTP requests generated by the Thumbroll app contained each procedure's general information (e.g. name, number of steps, images). This meant that the image and caption of each step of the procedure could be retrieved by constructing a HTTP request packet emulating those generated by the Thumbroll app.

A Python routine was therefore created to obtain the information from the Thumbroll dataset. First, a member of the research team had to manually click in the app all the different categories from Thumbroll's main screen (e.g. "Intern", "Junior Resident", "Clinical MD Trainee", "Internal Medicine", "Anatomy", "Radiology"). This allowed Fiddler4 to capture a response that contained the name of all the procedures in that specific category. The names of all these procedures were obtained from the HTTP responses and stored in a list. Afterwards, two HTTP requests were created for each procedure name that emulate those generate by the app. The first request was created to acquire the URL to all the images of each procedure. The second request was created to acquire the captions to all the images of each procedure. Figure 19 provides an example of how the responses to an emulated HTTP request generated with our Python script looked like.

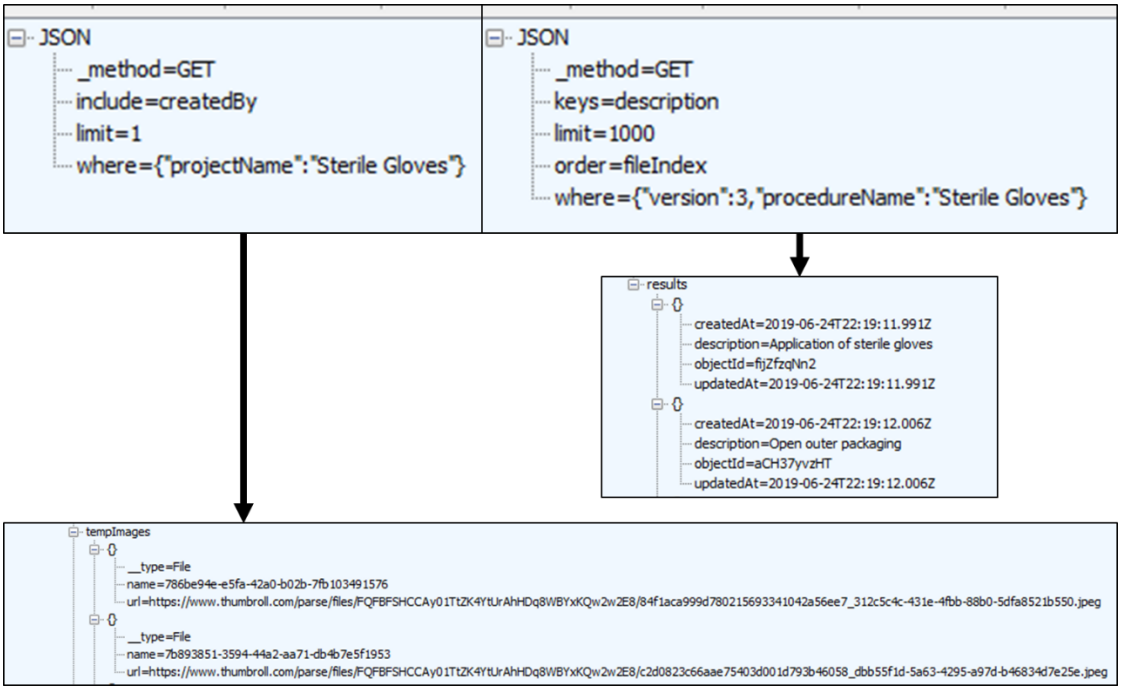


Figure 19. Example of responses to an emulated HTTP request. The images URL and the captions were contained in those responses.

The next step of the process was to obtain the images and captions from the captured HTTP responses. There responses were sent by the app server as a JSON message, which allowed the Python script to easily navigate and extract the desired information from the HTTP responses. The output of the step was a list of URLs and captions for each of the procedure names acquired in the previous step. Finally, a final HTTP request per obtained URL was performed to download each image. This was easily performed using Python's HTTPS routines. After all these steps, each procedure was stored as a folder that contained both the image and the caption of each step

depicted in the Thumbroll app. Figure 20 presents the section of the Python code that downloaded the images via HTTP requests.

```
def getImages():
    cwd = os.getcwd()

    for index in range(0,34):
        new_cwd = os.path.join(cwd,'NF'+str(index+1))

        with open(os.path.join(new_cwd,'Project.json')) as json_file:

            data = json.load(json_file)
            print(len(data['results']))

            # for x in range(len(data['results'])):
            #   fp = os.path.join(new_cwd,str(x+1).zfill(3))
            #   if not os.path.exists(fp):
            #       os.mkdir(fp)
            print("newProject")
            # file1 = open(os.path.join(new_cwd,'projectNames.txt'),'w')
            for procedure_index in range(0,len(data['results'])):

                print(data['results'][procedure_index]['projectName'])
            #     file1.write(data['results'][procedure_index]['projectName'])
            #     file1.write('\n')
            # file1.close()
                print("newFolder")
                images = data['results'][procedure_index]['tempImages']
                file_path = os.path.join(new_cwd,str(procedure_index+1).zfill(3))
                for image_index in range(len(images)):
                    final_file_path = os.path.join(file_path,str(image_index+1).zfill(3)+'.jpg')
                    #print(final_file_path)
                    #print(images[image_index]['url'])
                    with open(final_file_path,'wb') as handle:

                        response = requests.get(images[image_index]['url'], stream=True)

                        if not response.ok:
                            print(response)

                        for block in response.iter_content(1024):
                            if not block:
                                break

                            handle.write(block)
```

Figure 20. Python routine to perform a HTTP GET request. The routine will download the images given the URLs retrieved from the Thumboll app.

The result of this scrapping process were approximately 42500 images and captions. Nonetheless, several procedures were repeated within the different categories. A final cleaning routine was performed to generate a reduced database comprised of only those procedures that were not repeated between the categories. Figure 21 showcases how the images and matching captions are stored into one of the folders of the database.

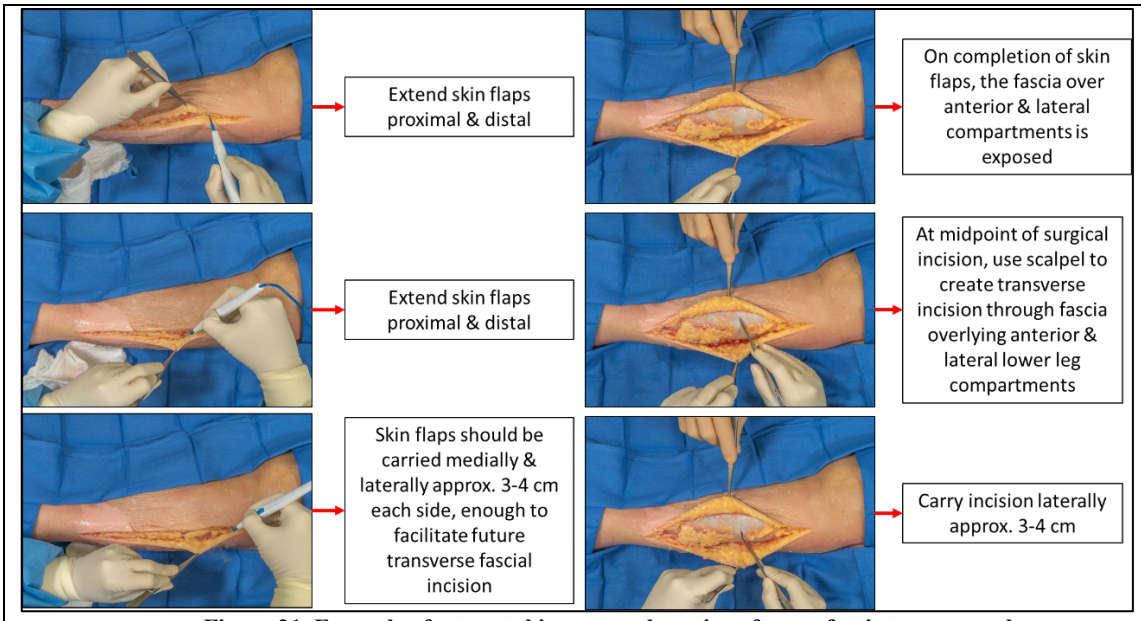


Figure 21. Example of extracted images and captions from a fasciotomy procedure.

The resulting DAISI database contains 14586 color images and text descriptions of instructions to perform surgical procedures. DAISI contains one example for each of the 198 medical procedures from 20 medical disciplines including ultrasound-guided diagnosis, trauma and gynecology. The DAISI dataset is divided into training and testing sets. The training set contains 13232 images from 173 medical procedures, and the testing set contains 1354 images from 25 different medical procedures than those used in training.

We used DAISI to train a Deep Learning model for autonomous mentoring. The algorithm receives images from medical procedures as input, and predicts an instruction associated with it. To generate text information from images, an encoder-decoder DL approach using a Convolutional Neural Network (CNN) and a Recursive Neural Network (RNN) was adopted. The CNN extracts and encodes visual features from the input images, and the RNN decodes these visual features into text descriptions, as depicted in Figure 22.

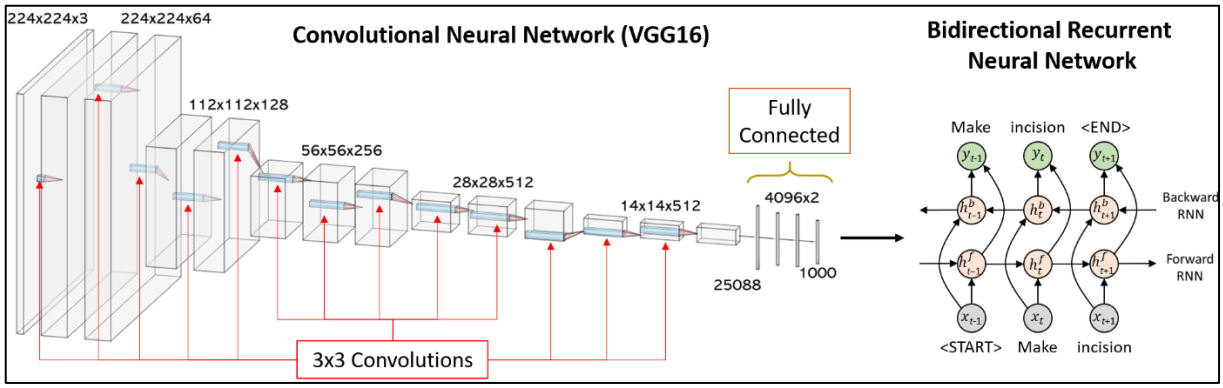


Figure 22. Schematic of our encoder-decoder architecture. The CNN obtains vectors representing input images. These vectors are then used in the training of a BRNN that learns to predict surgical instructions.

Captioning techniques require a vocabulary containing the words appearing in the dataset at least N times (defined by the *Word Count* parameter). This constrains the words used to

generate the instruction to a fix set. Our encoder-decoder architecture is based on NeuralTalk2. We use the VGG16 model as the encoder network. This model includes 13 convolutional layers with 5 pooling layers in-between. The convolutional layers use 3x3 convolutional filters to locate interest features in the images, and the pooling layers reduce the features' dimensionality. All hidden layers are equipped with Rectified Linear Units (ReLU). We performed cross validation using the Adam adaptive learning rate optimization to find individual learning rates for each parameter in the CNN. Finally, 4 fully connected layers are used to describe each image with a 1000-dimensional latent vector representation. We then use a Bidirectional Recurrent Neural Network (BRNN) as the decoder network to generate the text instructions. The BRNN predicts instructions not only by receiving the CNN's final latent vector, but also by leveraging context around the word. This context is determined via forward and backward hidden states (h_t^f and h_t^b, respectively) at each index t ($t = 1 \ldots T$), which denotes the position of a word in a sentence. The formulation of the BRNN follows:

$$c_v = W_{hi}\,[CNN_\theta(Img)]$$

$$h_t^f = \mathrm{ReLU}(W_{hx}x_t + W_{hf}h_{t-1}^f + b_f + \vec{1}(iter = 1) \odot b_v)$$

$$h_t^b = \mathrm{ReLU}(W_{hx}x_t + W_{hb}h_{t+1}^b + b_b + \vec{1}(iter = 1) \odot b_v)$$

$$y_t = Softmax(W_{ho}(h_t^f + h_t^b) + b_o)$$

W_{hi}, W_{hx}, W_{hf}, W_{hb}, W_{ho}; b_f, b_b, and b_o are the parameters and biases to be learnt by the model. $CNN_\theta(Img)$ is CNN's final latent vector of the image Img. Thus, the image context vector c_v provides the BRNN with information from the input image. This context vector c_v is provided only during the first iteration ($iter = 1$). The x_t and y_t vectors contain probabilities of each word in the vocabulary to be the word at the index \$t\$. The output vector y_t is used as x_{t+1} in the next iteration. In the first iteration, the output vector y_t depends only on the context vector c_v, as x_t takes a special initialization value (START) and h_t^f and h_t^b are initialized to 0. This formulation allows the model to predict more than one candidate instruction per image. The probability of each candidate being the correct instruction decreases for each additional prediction.

We evaluated our AI mentor using combinations of two parameters: *Image Resolution*: High (1260x840), Medium (315x210), and Low (63x42), and *Word Count*: 3, 5, and 7. Additionally, we conducted *Inter-procedure* and *Intra-procedure* evaluations. For the *Inter-procedure* setting, the model had no prior information regarding the procedures in the test set. For the *Intra-procedure* setting, a fraction of the images P in the same procedure were assigned to the training set, while the rest remained in the test set. The test set consisted of every $\frac{1}{P}$ images from each procedure. In our case, P was set to 0.5. While the *Intra-procedure* setting reduced generalizability among procedures, it enhanced performance for procedures in the test set. To evaluate the algorithm's performance, the BLEU metric was computed between the predicted and the ground truth instructions. This is a state-of-the-art metric to evaluate text production models. BLEU computes a 1-to-100 similarity score by comparing two sentences at the word n-gram level. We report cumulative BLEU scores for 1-grams to 4-grams for the model's top five candidate predictions, as they have reported correlations with human judgements.

Finally, expert physicians evaluated the algorithm's performance subjectively. We randomly selected 16 images from the testing set and their predicted instructions. Afterwards, we used a

survey to rate how related was each image to its predicted instruction. Each question in the survey included an image from a procedure, the name of the procedure; the instruction predicted, and a five scale ranking from: "Very Related" = 1, "Related" = 0.75, "Somewhat Related" = 0.5; "Not Related" = 0.25, and "Impossible to Tell" = 0. Figure 23 presents a screenshot of the web interface that was created for the experts to evaluate the captions predicted by our model.

Figure 23. Screenshot of the web interface that was created for the experts to evaluate the captions predicted by our model. 20 images were presented to the expert surgeons, each with its associated caption, procedure name and set of options to rate the predicted caption.

We validated our approach using four test folds. Figure 24 presents instructions predicted by our AI model. The predicted instruction is written inside the images, whereas the ground truth instruction is written below. The data followed three main trends: (1) high BLEU and subjective scores (e.g. Figure 24, example 1); (2) low BLEU scores but high subjective scores (e.g. Figure 24, examples 2, 3 and 4); and (3) low BLEU and subjective scores (e.g. Figure 24, examples 5 and 6). The first trend are descriptions considered as correct predictions: they were similar to

the ground truth and physicians considered them as adequate guidance. The second trend were descriptions that were not similar to the ground truth, but were consider as adequate guidance by the physicians. These descriptions included key elements from the image (e.g. tourniquet and gauze in Figure 24, examples 2 and 4, respectively), but did not use the phrasing of the ground truth. The third trend comprehends descriptions that were considered as incorrect predictions.

Figure 24. Examples of instructions predicted by our AI model. The predicted instruction is in white font, inside the images. The ground truth (GT) instruction is written below. The 1-gram BLEU score and the score given by the physicians is included. High, average, and low scores are in green, yellow and red font, respectively.

Figure 25 reports the cumulative BLEU scores for *Inter-procedure* and *Intra-procedure* testing. The captions predicted by our model obtained up to 67% BLEU 1-gram and 26% BLEU 4-gram scores. Our results surpassed those reported in state-of-the-art approaches for medical instructions prediction. Overall, the BLEU scores were slightly higher for higher *Word Count* values. A potential reason is that a reduced-size vocabulary increased the chance to learn meaningful relations between the images and the text descriptions. Likewise, the AI algorithm slightly favored smaller *Image Resolution* values. While the best results reported are not high, our algorithm tackles a challenging problem due to the interclass variance among different medical procedures, which in turns has an impact the prediction capability of the network. As a reference value, the BLEU 1-gram score when comparing the ground truth instructions with descriptions constructed using random words from the vocabulary is less than 0.1%. Our results show an improvement of over 4 folds over random guess.

Five expert physicians completed our subjective evaluation, for a total 80 responses. The physicians reported having 11.2 ± 3.3 years of medical expertise. On average, the physicians considered the predicted instructions to be "Somewhat Related" to the medical images (0.51 ± 0.32). While this is an encouraging result, a drastic improvement is still required for useful AI

mentoring for surgery. Therefore, the main value of this work is offering a baseline for future autonomous medical mentoring applications. The DAISI open dataset is a useful tool that allows the AI community to train machine learning models that learn clinical instructions. Future work includes adding more repetitions per procedure. While our *Intra-procedure* testing approach alleviates this limitation, more repetitions can improve the prediction results. Finally, data augmentation techniques can be used to increase the size of the dataset, for example by training a Generative Adversarial Neural network to create new images and descriptions.

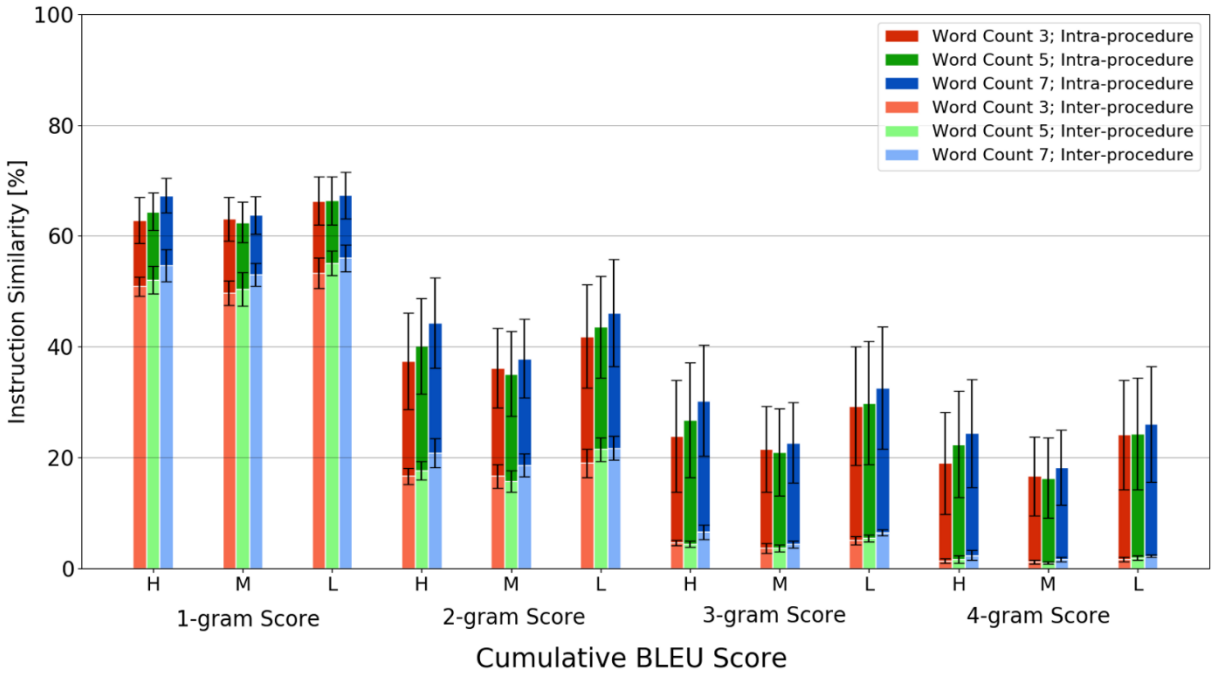


Figure 25. Cumulative *n*-gram BLEU scores. Our model was evaluated using three *Word Count* values (3, 5, 7), three *Image Resolution* values (H=1260x840, M=315x210, L=63x42) and two testing approaches (*Inter-procedure*, *Intra-procedure*). The model obtained up to 67% 1-gram BLEU scores, and up to 26% 4-gram BLEU scores.

Finally, to further pursue the commercialization of this approach, PI Dr. Juan Wachs and Mr. Edgar Rojas-Muñoz participated in the National Science Foundation Midwest Node Introduction to Customer Discovery program. This program gives participants a working knowledge of how to think about their ideas from a business and customer perspective, as well as how to properly conduct customer discovery interviews. As part of this course, we performed comprehensive interviews assessing the validity of mentorship approaches based on artificial intelligence. The interviews analyzed the current status of medical training curricula and discussed approaches to reinforce the issues found within these programs. Twenty medical practitioners from Tanzania, Costa Rica and United States were recruited for these interviews. The practitioners ranged from students to Attending-level practitioners from the military, urban and rural hospitals, non-governmental organizations, and first response teams. The interviews revealed that medical curricula are in the need of individualized training programs. An artificial mentor could address this gap by learning the patterns of the mentees and providing tailored learning opportunities based on the participant's expertise and performance. Additionally, the interviewees mentioned being interested in novel technologies capable of increasing their probability of success during emergency situations. This is particularly important in rural and austere settings, where the availability of training opportunities is limited. Nonetheless, such

training platforms need to be low-cost and easily maintainable to compete against mobile apps as means of reinforcing medical knowledge. In summary, the interviews confirmed the need of autonomous mentoring approaches, especially in rural settings and emergency situations. Such platforms should focus on individualizing the training, and on assisting the mentee's decisions, while also being cost-effective and easy to maintain.

Presentation at the 2019 Military Health System Research Symposium

A poster titled "Training Effectiveness for Point of Injury Medical Care - A Portable and Self-contained Approach for Surgical Telementoring: Towards Remote, Point of Injury Care" was presented on 22-Aug-2019 during the 2019 Military Health System Research Symposium. The poster elaborated on the latest user study evaluating the STAR platform's potential to coach US Navy corpsmen remotely. The study was performed in collaboration between Purdue University, Indiana University School of Medicine, and the Naval Medical Center Portsmouth. Figure 26 presents a photograph of PI Dr. Juan Wachs presenting the poster during the Symposium.

Figure 26. PI Dr. Juan Wachs presenting a poster at MHSRS 2019. The poster reported on user study that evaluated our platform in a simulated austere scenario.

Task 4.1- Specialize the system for fasciotomy on a cadaveric leg

Evaluating STAR as means to provide coaching and confidence

We wanted to evaluated STAR as a platform that could provide surgery residents and medical students with coaching and confidence as they performed medical procedures. To do this, we leveraged our study in the context of fasciotomies, reported in the previous annual report. In this study, 20 participants were guided by remote expert surgeons to perform leg fasciotomies on cadavers under one of two conditions: telementoring (STAR), or independently reviewing the procedure beforehand. To investigate the effect of the mentoring conditions with respect to the participant's expertise, participants were recruited from three different strata to encompass various expertise levels: only medical students (n = 6; 3 per condition), only residents (n = 14; 7 per condition), and a combination of medical students and first-year residents (n = 10; 5 per condition). The latter was considered the sub-group who would benefit the most from the coaching experience due to their relatively lower expertise.

Based on the participants' video footage of the STAR participants, acquired as they performed the fasciotomies, six additional measurements were obtained to objectively describe the confidence scores and the overall coaching quality. These measurements were: 1) the number of surgical AR annotations created by the remote mentor; 2) the number of times the mentor asked the mentee for confirmation (e.g. "*That structure looks like the nerve to me, do you think the same?*"); 3) the number of times the mentee asked the mentor for instruction (e.g. "*There is muscle at the posterior border of the tibia, what would you like me to do?*"); 4) the number of times the mentee asked the mentor for confirmation (e.g. "*Are you sure I can cut this?*"); 5) the number of corrections given by the mentor (e.g. "*No, use your scissors for that, not the knife*"); and 6) the percentage of the total completion time during which the mentee received guidance from the remote mentor. These measurements were obtained only for participants in the STAR condition, as participants in the Control condition did not receive telementoring. Table 4 reports the measurements of confidence and coaching for participants in the STAR condition, divided by the three population sub-groups. Medical students received significantly more (p = 0.04) corrections when compared to residents.

Table 4. Quantifications of coaching and confidence for the different expertise-based sub-groups.

Measurements of Coaching and Confidence	Med Students (*n* = 3)	Residents (*n* = 7)	Low-Expertise (*n* = 5)
Number of AR annotations created, mean (95% CI), count	18.00 (13.52-23.49)	19.29 (16.17-22.83)	20.00 (16.27-24.33)
Number of times the mentor asked for confirmation, mean (95% CI), count	8.67 (5.66-12.70)	6.14 (4.45-8.27)	8.00 (5.72-10.89)
Number of times the mentee asked for instruction, mean (95% CI), count	5.67 (3.30-9.07)	4.29 (2.89-6.12)	5.20 (3.40-7.62)
Number of times the mentee asked for confirmation, mean (95% CI), count	3.00 (1.37-5.69)	5.29 (3.72-7.29)	4.20 (2.60-6.42)
Number of corrections given by the mentor, mean (95% CI), count	7.00 (4.33-10.70)	3.86 (2.54-5.61)	6.40 (4.38-9.03)
Time during which the mentee received guidance, mean (95% CI), percentage	44.2 (39.5-48.9)	49.15 (42.78-55.52)	44.51 (42.33-46.69)

STAR participants reported significant improvements in all evaluated aspects of their confidence scores. These results demonstrate that an interactive telementoring experience with STAR's ARHMD had a positive impact in participants' confidence. Although studies have shown that health practitioners' confidence in their surgical skills is correlated to competence and self-assessment of their skill, surveys report that the health practitioners' confidence in their skills is not particularly high. Extrapolating our results, integrating STAR to current coaching programs could help reinforcing surgical knowledge and enhancing the self-confidence of health practitioners.

The measurements of confidence and coaching elaborated on the self-reported confidence scores. The remote mentors created 19 annotations per participant on average. The use of the annotations can be divided into four situations: 1) exemplifying which instrument to use (e.g. placing the icon of a scalpel after saying "*Cut here*"); 2) showing the location of anatomical structures (e.g. drawing a circle around the peroneal nerve; 3) showing the length and location of incisions (e.g. drawing a line along the leg to depict where to cut); and 4) acquiring a better awareness of the operating field (e.g. drawing a circle around the toe to determine the orientation of the leg). By creating these annotations, the mentor was able to convey more guidance, a possible reason of the increased performance and confidence scores of STAR participants.

Moreover, transmitting the real-time visual feedback of the operating field allowed the remote mentor to provide better coaching. The visual feedback allowed the mentor to ask for confirmation 7 times per participant on average, and to perform 5 corrections per participants on average. The following transcription exemplifies one of these situations:

Mentor: "*That looks the saphenous vein*."

Mentee: "*What should I do with it?*"

Mentor: "*Continue with your incision, just make sure to stay away from the vein.*"

Mentee continues with the incision, and gets dangerously close to the vein.

Mentor: "*Wow! Be careful there, you almost got the vein in the last movement you did. Try not to get your knife too close to this area* (the mentor draws a circle in the screen)."

In this example, the mentor corrected the mentee and provided more details about which area to avoid thanks to the visual feedback and the AR annotations. The visual feedback also allowed the mentee to ask for instructions and confirmations an average of 5 times per procedure, as depicted in the following example:

Mentee: "*There is still some muscle here* (the mentee points at a specific point in the leg). *I think I should cut there more."*

Mentor: "*No, it is okay if there is still some muscle there."*

The percentage of time during which mentees received guidance from the remote mentor can be associated with the increased performance and confidence scores. On average, the mentees received guidance for 10 minutes and 48 seconds, which represented 47% of the total task completion time. These results represent that STAR participants received remote guidance for almost half of their task completion time without incurring into statistically significant completion time increases.

Finally, to analyze the usefulness of our platform even further, we inspected the errors performed the participants as they performed the fasciotomies. Our metric, the Individual Performance Score (IPS), included a description of 11 common errors incurred by surgeons as they perform fasciotomies (E1 to E11). Our expert evaluators annotated whenever the participants of our experiment incurred into these errors. Table 5 presents the distribution of the errors with respect to the mentoring condition for the different expertise-based sub-groups. The

errors were divided into different levels of occurrence based on how many participants from each of the expertise-based sub-groups incurred in such error. The error was classified as low frequency if less than 20% of the participants in the sub-group incurred into the error. The error was classified as medium frequency if between 20% to 40% of the participants in the sub-group incurred into the error. Finally, the error was classified as high frequency if more than 40% of the participants in the sub-group incurred into the error. According to our low-medium-high frequency classification scheme, the sub-group with only medical students had 9 errors classified as low frequency, 1 classified as medium frequency, and 1 classified as high frequency; the sub-group with only resident had 8 errors classified as low frequency, 3 classified as medium frequency, and 0 classified as high frequency; and the sub-group with medical students and first-year residents combined had 6 errors classified as low frequency, 5 classified as medium frequency, and 0 classified as high frequency.

A 2-sample t-test was used to evaluate the hypothesis of whether receiving mentoring using STAR condition reduced the error occurrence was performed. The results revealed that participants in the STAR condition performed significantly less low frequency errors ($p = 0.05$), as well as significantly less medium frequency errors ($p < 0.001$). No statistical analyses were run for high frequency errors because not enough errors were classified into this category.

Breaking down the errors in this way allowed us to analyze which steps were more difficult, and whether receiving mentoring using STAR condition reduced the amount of times each specific error was performed. Based on our low-medium-high frequency classification scheme, errors E3, E5, E7, E9 and E11 were classified as low frequency for all the three expertise-based sub-groups. Most of these possible errors were related to incorrectly identifying and releasing the compartments (anterior, lateral and posterior). These results show that participants were able to release the outermost leg compartments without difficulties. Nonetheless, the process of releasing the deep posterior compartment (E10) was considered as medium frequency for both the only medical students and the medical students and first-year residents sub-groups. These results are not unexpected, as the process of releasing the deep posterior compartment can be more error prone. Therefore, considering this results, it is recommended paying extra attention to the instruction process of releasing this compartment while performing training for fasciotomies. Errors related to how participants used their tools (E1, E4, E6, and E7) were considered of medium frequency in most the cases (except E7). This reveals a deficiency in the way participants handled their surgical instruments. As a result, special emphasis should be placed in the correct use of surgical instruments during the residency years of future surgical personnel. Errors related to the identification and protection of anatomical landmarks (E2, E8, E7, and E11) were the most incurred during our experiment, to the point of being considered of high frequency for the only medical students sub-group. This reveal an aspect that should be reinforced in the current residency programs, as the ability of identifying and protecting internal anatomical structures is critical to proper surgical performance. These insights and the results from the statistical analyses reaffirm the validity of our platform as a novel method to increase medical confidence and provide accessible coaching.

Table 5. Distribution of the errors highlighted by IPS with respect to the mentoring condition for the different expertise-based sub-groups.

Error Code	Error Description	Med Students (n = 6)		Residents (n = 14)		Low-Expertise (n = 10)	
		STAR (n = 3)	CONTROL (n = 3)	STAR (n = 7)	CONTROL (n = 7)	STAR (n = 5)	CONTROL (n = 5)
E1	After initial anterolateral incision, uses scalpel or scissor instead of blunt instrument to avoid damaging the superficial peroneal nerve	0	0	1	2	0	1
		low frequency		medium frequency		low frequency	
E2	Does not protect superficial peroneal nerve while extending the intermuscular septum incision over the lateral and anterior compartments	0	1	1	3	0	2
		low frequency		medium frequency		medium frequency	
E3	Incorrectly identifies and releases the anterior compartment	0	0	0	0	0	0
		low frequency		low frequency		low frequency	
E4	The tip of the scissors was not directed away from the intermuscular septum while releasing the anterior compartment	0	1	0	1	0	2
		low frequency		low frequency		medium frequency	
E5	Incorrectly identifies and releases the lateral compartment	0	1	0	0	0	1
		low frequency		low frequency		low frequency	
E6	The tip of the scissors was not directed away from the intermuscular septum while releasing the lateral compartment	0	1	0	3	0	2
		low frequency		medium frequency		medium frequency	
E7	After initial posteromedial incision, uses scalpel or scissor instead of blunt instrument to avoid damaging the superficial peroneal nerve	0	0	0	0	0	0
		low frequency		low frequency		low frequency	
E8	Does not identify and retract the saphenous vein posteriorly	1	2	1	0	1	2
		high frequency		low frequency		medium frequency	
E9	Incorrectly identifies and releases the superficial posterior compartment	0	0	1	0	0	0
		low frequency		low frequency		low frequency	
E10	Incorrectly identifies and releases the deep posterior compartment	0	2	1	1	0	3
		medium frequency		low frequency		medium frequency	
E11	Fails to protect the neurovascular bundle while releasing the deep posterior compartment	0	0	0	0	0	0
		low frequency		low frequency		low frequency	

What opportunities for training and professional development has the project provided?
If the project was not intended to provide training and professional development opportunities or there is nothing significant to report during this reporting period, state "Nothing to Report."

Describe opportunities for training and professional development provided to anyone who worked on the project or anyone who was involved in the activities supported by the project. "Training" activities are those in which individuals with advanced professional skills and experience assist others in attaining greater proficiency. Training activities may include, for example, courses or one-on-one work with a mentor. "Professional development" activities result in increased knowledge or skill in one's area of expertise and may include workshops, conferences, seminars, study groups, and individual study. Include participation in conferences, workshops, and seminars not listed under major activities.

One of the final extra technical modules added to the system was an AI agent capable of predicting surgical instructions from medical images and videos. PI Dr. Juan Wachs and Mr. Edgar Rojas-Muñoz explored the possibilities of commercialization of such a module by completing the National Science Foundation Midwest Node Introduction to Customer Discover program. Through this program, both team members were able to gain exposure some of the entrepreneurship processes required to formulate a business idea. Thanks to this program and with the help of Purdue University, PI Dr. Juan Wachs and Mr. Edgar Rojas-Muñoz are in the process of applying for patent protection for the AI surgical mentoring agent.

How were the results disseminated to communities of interest?
If there is nothing significant to report during this reporting period, state "Nothing to Report."

Describe how the results were disseminated to communities of interest. Include any outreach activities that were undertaken to reach members of communities who are not usually aware of these project activities, for the purpose of enhancing public understanding and increasing interest in learning and careers in science, technology, and the humanities.

Several publications for academic journals and conferences have been submitted and accepted during this reporting period. In terms of journals, our submission to Nature Digital Medicine "Evaluation of an Augmented Reality Platform for Austere Surgical Telementoring: A Randomized Controlled Crossover Study in Cricothyroidotomies" was accepted for publication on April 2020. Moreover, our submission to Surgery "The System for Telementoring with Augmented Reality (STAR): A Head-Mounted Display to Improve Surgical Coaching and Confidence in Remote Areas" was accepted for publication on November 2019. Finally, our submission to Military Medicine "Telementoring in Leg Fasciotomies via Mixed-Reality: Clinical Evaluation of the STAR" was accepted for publication on July 2019. In terms of conferences publications, our submission "How About the Mentor? Effective Workspace Visualization in AR Telementoring" was accepted for publication in the IEEE Conference on Virtual Reality and 3D User Interfaces on March 2020; and our submission "DAISI: Database for AI Surgical Instruction" was submitted to the International Conference on Medical Image Computing and Computer Assisted Intervention and, in parallel, uploaded to ArXiv on April 2020.

Moreover, we presented a poster at the 2019 Military Health System Research Symposium: "Training Effectiveness for Point of Injury Medical Care - A Portable and Self-contained Approach for Surgical Telementoring: Towards Remote, Point of Injury Care". Finally, a demo of our system was given at the Augmented & Virtual Reality in Medicine Workshop, in the 2019 Scientific Assembly of the American College of Emergency Physicians.

What do you plan to do during the next reporting period to accomplish the goals?
If this is the final report, state "Nothing to Report."

Describe briefly what you plan to do during the next reporting period to accomplish the goals and objectives.

Nothing to report

4. **IMPACT:** Describe distinctive contributions, major accomplishments, innovations, successes, or any change in practice or behavior that has come about as a result of the project relative to:

What was the impact on the development of the principal discipline(s) of the project?
If there is nothing significant to report during this reporting period, state "Nothing to Report."

Describe how findings, results, techniques that were developed or extended, or other products from the project made an impact or are likely to make an impact on the base of knowledge, theory, and research in the principal disciplinary field(s) of the project. Summarize using language that an intelligent lay audience can understand (Scientific American style).

This technology will increase the sense of co-presence in the operating room between mentor and trainee. This is a fundamental step towards telexistence. Telexistence is a concept used to describe the framework that allows humans to have a real-time sensation of being and interacting with objects in places somewhere different from their actual location. The fundamental premise is that a higher sense of co-presence has an impact on the quality of mentorship. For example, by allowing the mentors to physically interact with the patient's anatomy though hand gestures (embodied interaction), the mentor's level of immersion and engagement will be significantly increased.

What was the impact on other disciplines?
If there is nothing significant to report during this reporting period, state "Nothing to Report."

Describe how the findings, results, or techniques that were developed or improved, or other products from the project made an impact or are likely to make an impact on other disciplines.

Our results are of benefit to the surgical community and the battlefield medicine communities by demonstrating that AR approaches to surgical telementoring are feasible in this environment which has been under-served by traditional approaches. In addition, the formal analysis of our first-person camera stabilization procedure makes a strong case that the plane-based stabilization we support is the proper direction for other VR and AR applications. This analysis will help the VR and AR research community define a good set of working principles for future applications, even outside the realm of surgical telementoring. Additionally, our DAISI database, source of information for our AI agent module, provides a baseline for AI algorithms to assist in autonomous medical mentoring.

What was the impact on technology transfer?
If there is nothing significant to report during this reporting period, state "Nothing to Report."

Describe ways in which the project made an impact, or is likely to make an impact, on commercial technology or public use, including:

- *transfer of results to entities in government or industry;*
- *instances where the research has led to the initiation of a start-up company; or*
- *adoption of new practices.*

As part of the dissemination efforts and code handoff, an installation guide was created. The installation guide includes a detailed, step-by-step explanation of the steps to: 1) download the code from the GitHub repositories; 2) install the code and dependencies in a PC and Microsoft HoloLens device; 3) run each of the subsystems of our telementoring platform; and 4) connect the subsystems between them. The document can be downloaded using the following link: https://engineering.purdue.edu/starproj/wp-content/uploads/STAR_Installation_Guide.pdf. Additionally, and with the help of Purdue University, PI Dr. Juan Wachs and Mr. Edgar Rojas-Muñoz are in the process of applying for patent protection for the AI surgical mentoring agent.

What was the impact on society beyond science and technology?
If there is nothing significant to report during this reporting period, state "Nothing to Report."

Describe how results from the project made an impact, or are likely to make an impact, beyond the bounds of science, engineering, and the academic world on areas such as:

- *improving public knowledge, attitudes, skills, and abilities;*
- *changing behavior, practices, decision making, policies (including regulatory policies), or social actions; or*
- *improving social, economic, civic, or environmental conditions.*

Our augmented-reality approach to surgical telementoring is well suited to any field that requires mentor assistance, not just surgical applications. The main principles of encouraging co-presence between trainee and mentor, of visual annotation of an operating field, of two-way audio communication, and of video stabilization for a mentor view are all important factors in any remote guidance system. We anticipate that our advances hold potential for manufacturing and inspection and in training skills in rural or developing regions.

5. **CHANGES/PROBLEMS:** The Project Director/Principal Investigator (PD/PI) is reminded that the recipient organization is required to obtain prior written approval from the awarding agency Grants Officer whenever there are significant changes in the project or its direction. If not previously reported in writing, provide the following additional information or state, "Nothing to Report," if applicable:

Changes in approach and reasons for change
Describe any changes in approach during the reporting period and reasons for these changes. Remember that significant changes in objectives and scope require prior approval of the agency.

There were no significant changes in our approach during this period.

Actual or anticipated problems or delays and actions or plans to resolve them
Describe problems or delays encountered during the reporting period and actions or plans to resolve them.

No significant problems were found.

Changes that had a significant impact on expenditures
Describe changes during the reporting period that may have had a significant impact on expenditures, for example, delays in hiring staff or favorable developments that enable meeting objectives at less cost than anticipated.

No changes

Significant changes in use or care of human subjects, vertebrate animals, biohazards, and/or select agents

Describe significant deviations, unexpected outcomes, or changes in approved protocols for the use or care of human subjects, vertebrate animals, biohazards, and/or select agents during the reporting period. If required, were these changes approved by the applicable institution committee (or equivalent) and reported to the agency? Also specify the applicable Institutional Review Board/Institutional Animal Care and Use Committee approval dates.

Significant changes in use or care of human subjects

No changes

Significant changes in use or care of vertebrate animals.

No changes

Significant changes in use of biohazards and/or select agents

No changes

6. PRODUCTS: List any products resulting from the project during the reporting period. If there is nothing to report under a particular item, state "Nothing to Report."

- **Publications, conference papers, and presentations**

Report only the major publication(s) resulting from the work under this award.

Journal publications. *List peer-reviewed articles or papers appearing in scientific, technical, or professional journals. Identify for each publication: Author(s); title; journal; volume: year; page numbers; status of publication (published; accepted, awaiting publication; submitted, under review; other); acknowledgement of federal support (yes/no).*

1.
Title: Surgical Telementoring with Augmented Reality: A Randomized Control Crossover Experiment to Provide Remote Assistance at the Point of Injury
Journal: Nature Digital Medicine
Authors: Edgar Rojas-Muñoz, Chengyuan Lin, Natalia Sanchez-Tamayo, Maria Eugenia Cabrera, Daniel Andersen, Voicu Popescu, Juan Barragan Noguera, Ben Zarzaur, Patrick Murphy, Kathryn Anderson, Thomas Douglas, Clare Griffis, Andrew W. Kirkpatrick, Jessica McKee, Juan Wachs
Status of Publication: Accepted, awaiting publication.
Acknowledgment of federal support: yes

2.
Title: The System for Telementoring with Augmented Reality (STAR): A Head-Mounted Display to Improve Surgical Coaching and Confidence in Remote Areas
Journal: Surgery
Authors: Edgar Rojas-Muñoz, Maria Eugenia Cabrera, Chengyuan Lin, Daniel Andersen, Voicu Popescu, Kathryn Anderson, Ben Zarzaur, Brian Mullis, Juan Wachs
Status of Publication: Published. E-pub ahead of print:
https://doi.org/10.1016/j.surg.2019.11.008
Acknowledgment of federal support: yes.

3.
Title: Telementoring in Leg Fasciotomies via Mixed-Reality: Clinical Evaluation of the STAR Platform
Journal: Journal of Military Medicine
Authors: Edgar Rojas-Muñoz, Maria Eugenia Cabrera, Chengyuan Lin, Natalia Sánchez-Tamayo, Dan Andersen, Voicu Popescu, Kathryn Anderson, Ben Zarzaur, Brian Mullis, Juan Wachs
Status of Publication: Published. Volume 184, no. Supplement_1 (2019): 513-520.
Acknowledgment of federal support: yes.

Books or other non-periodical, one-time publications. *Report any book, monograph, dissertation, abstract, or the like published as or in a separate publication, rather than a periodical or series. Include any significant publication in the proceedings of a one-time conference or in the report of a one-time study, commission, or the like. Identify for each one-time publication: Author(s); title; editor; title of collection, if applicable; bibliographic information; year; type of publication (e.g., book, thesis or dissertation);*

status of publication (published; accepted, awaiting publication; submitted, under review; other); acknowledgement of federal support (yes/no).

Other publications, conference papers, and presentations. *Identify any other publications, conference papers and/or presentations not reported above. Specify the status of the publication as noted above. List presentations made during the last year (international, national, local societies, military meetings, etc.). Use an asterisk (*) if presentation produced a manuscript.*

1.
Title: How About the Mentor? Effective Workspace Visualization in AR Telementoring
Conference: IEEE Conference on Virtual Reality and 3D User Interfaces (IEEE VR 2020)
Authors: Chengyuan Lin, Edgar Rojas-Muñoz, Maria Eugenia Cabrera, Natalia Sanchez-Tamayo, Daniel Andersen, Voicu Popescu, Juan Antonio Barragan Noguera, Ben Zarzaur, Pat Murphy, Kathryn Anderson, Thomas Douglas, Clare Griffis, and Juan Wachs
Status of Publication: Accepted and presented.
Acknowledgment of federal support: yes.

2.
Title: DAISI: Database for AI Surgical Instruction
Conference: International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI 2020)
Authors: Edgar Rojas-Muñoz, Kyle Couperus, Juan Wachs
Status of Publication: Submitted; available in the ArXiv online repository.
Acknowledgment of federal support: yes.

3.
Title: Training Effectiveness for Point of Injury Medical Care - A Portable and Self-contained Approach for Surgical Telementoring: Towards Remote, Point of Injury Care
Conference: Military Health System Research Symposium (MHSRS 2019)
Presenter: Juan Wachs; presented as poster.
Acknowledgment of federal support: yes.

4.
Demo: The System for Telementoring with Augmented Reality
Conference: Augmented & Virtual Reality in Medicine Workshop, in the 2019 Scientific Assembly of the American College of Emergency Physicians.

Presenter: Kyle Couperus
Acknowledgment of federal support: yes.

- **Website(s) or other Internet site(s)**
  List the URL for any Internet site(s) that disseminates the results of the research activities. A short description of each site should be provided. It is not necessary to include the publications already specified above in this section.

Official project website, with overview of research, links to publications, images, and videos.
https://engineering.purdue.edu/starproj

STAR Project Youtube page:
https://www.youtube.com/channel/UCSrh1dlsrvbGvE_hLenXwGQ

- **Technologies or techniques**
  Identify technologies or techniques that resulted from the research activities. In addition to a description of the technologies or techniques, describe how they will be shared.

A telementoring system based on an Augmented Reality Head-Mounted Display was developed and validated with US Navy corpsmen, surgery residents and medical students as mentees, and general and orthopaedic surgeons as mentors. This approach leverages a novel technology to design a surgical telementoring system that creates immersive experiences without introducing additional encumbrance in the surgeons' working space. The work on this technique was presented at international conferences and through journal publications, and the its code was made available.

- **Inventions, patent applications, and/or licenses**
  Identify inventions, patent applications with date, and/or licenses that have resulted from the research. State whether an application is provisional or non-provisional and indicate the application number. Submission of this information as part of an interim research performance progress report is not a substitute for any other invention reporting required under the terms and conditions of an award.

PI Dr. Juan Wachs and Mr. Edgar Rojas-Muñoz completed the National Science Foundation Midwest Node Introduction to Customer Discover program. Through this program, and with the help of Purdue University, PI Dr. Juan Wachs and Mr. Edgar Rojas-Muñoz are in the process of applying for patent protection for the AI surgical mentoring agent.

- **Other Products**

Identify any other reportable outcomes that were developed under this project. Reportable outcomes are defined as a research result that is or relates to a product, scientific advance, or research tool that makes a meaningful contribution toward the understanding, prevention, diagnosis, prognosis, treatment, and/or rehabilitation of a disease, injury or condition, or to improve the quality of life. Examples include:

- *data or databases;*
- *biospecimen collections;*
- *audio or video products;*
- *software;*
- *models;*
- *educational aids or curricula;*
- *instruments or equipment;*
- *research material (e.g., Germplasm; cell lines, DNA probes, animal models);*
- *clinical interventions;*
- *new business creation; and*
- *other.*

The projects' installation guide can be downloaded using the following link:
https://engineering.purdue.edu/starproj/wp-content/uploads/STAR_Installation_Guide.pdf

The following repositories can be used to acquire our platform's subsystems:

- Mentor System: https://github.com/edkazar/MentorSystemUWPWebRTC/
- Mentee System: https://github.com/practisebody/STAR/
- STAR Controller App:
  - PC Version: https://github.com/practisebody/STARController_UWP/
  - Phone Version: https://github.com/practisebody/STARController_Android/

The DAISI database of medical images and captions can be found in the following link:
https://engineering.purdue.edu/starproj/_daisi/

7. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS

What individuals have worked on the project?

Provide the following information for: (1) PDs/PIs; and (2) each person who has worked at least one person month per year on the project during the reporting period, regardless of the source of compensation (a person month equals approximately 160 hours of effort). If information is unchanged from a previous submission, provide the name only and indicate "no change."

Name:	*Juan P Wachs*
Project Role:	*Principal Investigator*
Researcher Identifier (e.g. ORCID ID):	*0000-0002-6425-5745*
Nearest person month worked:	*1 month*
Contribution to Project:	*Supervising the overall performance of the project. Coordinated visits to user study sites. Working with Edgar Rojas for the design of the large interaction table. Working with Natalia Sanchez in ultrasound device integration. Helping with journal publications.*
Name:	*Voicu Popescu*
Project Role:	*Co-Investigator*
Researcher Identifier (e.g. ORCID ID):	
Nearest person month worked:	*1 month*
Contribution to Project:	*Actively participated in and advised research assistant Chengyuan Lin in the research and development of the ARHMD system; in designing, conducting, and analyzing the results of user studies aimed at assessing STAR; in disseminating the project results in publications.*
Name:	*Edgar Rojas Muñoz*
Project Role:	*Research Assistant*
Researcher Identifier (e.g. ORCID ID):	*0000-0001-6909-375X*
Nearest person month worked:	*12 months*
Contribution to Project:	*Lead author on several publications related to the STAR platform. Developed and improved the Mentor System.*
Name:	*Chengyuan Lin*
Project Role:	*Research Assistant*
Researcher Identifier (e.g. ORCID ID):	
Nearest person month worked:	*12 months*
Contribution to Project:	*Lead author on several publications related to the STAR platform. Developed and integrated improvements to ARHMD-based trainee system. Researched and implemented pose stabilization for first-person mentee camera data for stable mentor view.*
Name:	*Natalia Sanchez*

Project Role:	*Research Assistant*
Researcher Identifier (e.g. ORCID ID):	
Nearest person month worked:	*5 months*
Contribution to Project:	*Integrated ultrasound device as video source to be visible to mentor. Designed and 3D printed tracker for use in determining pose of ultrasound device.*

Has there been a change in the active other support of the PD/PI(s) or senior/key personnel since the last reporting period?

If there is nothing significant to report during this reporting period, state "Nothing to Report."

If the active support has changed for the PD/PI(s) or senior/key personnel, then describe what the change has been. Changes may occur, for example, if a previously active grant has closed and/or if a previously pending grant is now active. Annotate this information so it is clear what has changed from the previous submission. Submission of other support information is not necessary for pending changes or for changes in the level of effort for active support reported previously. The awarding agency may require prior written approval if a change in active other support significantly impacts the effort on the project that is the subject of the project report.

Juan Wachs	09/01/2014 - 08/31/2017	0.23 SU	0.5 AY
University of Denver	$200,000		

NSF: MRI Development: Human Avatars: Enabling Research in Natural Communication with Virtual Tutors, Therapists, and Robotic Companions

Major Goals of the Project: The goal of the proposed MRI development project is to develop a life-like emotive software/hardware instrument in the form of robotic character heads that will support natural spoken dialogs between the robot and a human that closely models the face-to-face communication behaviors of a sensitive and effective human tutor, clinician or caregiver to a degree unachievable with current instrumentation.

Overlap: No overlap.

Juan Wachs	09/5/2014 - 08/31/2019	0 SU	0 AY
	$65,000		

NSF: Collaborative Research: I/UCRC for Robots and Sensors for the Human Wellbeing

Major Goals of the Project: The goal of the proposed center is to develop technology in the form of robots and sensors for assistive technologies to support therapies and rehabilitation of people with disabilities.

Overlap: No overlap.

Juan Wachs	04/1/2015 - 03/31/2016 $90,000	0.12 SU AY	0.5

THE NAVSUP FLEET LOGISTICS CENTER SAN DIEGO: An Efficient Real-Time Method for Detection and Characterization of UAVs

Major Goals of the Project: The research objective of this proposal is to develop a video-based methods for real-time detection of small, unmanned aerial vehicles (UAVs) leveraging on effective sense and avoid techniques. Such methods can be integrated into real-time on board processors. This, in turn, would lead to enhanced UAV's capabilities for detection of friendly and unfriendly airborne traffic and respond with appropriate alarms, maneuvers and notifications.
Overlap: No overlap.

What other organizations were involved as partners?

If there is nothing significant to report during this reporting period, state "Nothing to Report."

Describe partner organizations – academic institutions, other nonprofits, industrial or commercial firms, state or local governments, schools or school systems, or other organizations (foreign or domestic) – that were involved with the project. Partner organizations may have provided financial or in-kind support, supplied facilities or equipment, collaborated in the research, exchanged personnel, or otherwise contributed.

Provide the following information for each partnership:

Organization Name: Indiana University School of Medicine
Location of Organization: Indianapolis, Indiana, USA
Partner's contribution to the project (identify one or more)

- *Task 4.1- Specialize the system for fasciotomy on a cadaveric leg. The co-Investigators helped revising the manuscripts submitted to Military Medicine, Surgery, Nature Digital Medicine and IEEE VR.*
- *Collaboration: Dr. Zarzaur, Mrs. Anderson and Dr. Mullis collaborated with the project staff on the project.*

Organization Name: Naval Medical Center Portsmouth
Location of Organization: Portsmouth, Virginia, USA
Partner's contribution to the project (identify one or more)

- *Task 3.1- Specialize the system for a cric procedure on a patient simulator in an austere environment. The co-Investigators helped revising the manuscripts submitted to Nature Digital Medicine and IEEE VR.*
- *Collaboration: Dr. Douglas and Dr. Griffis collaborated with the project staff on the project.*

Organization Name: University of Calgary
Location of Organization: Calgary, Alberta, Canada
Partner's contribution to the project (identify one or more)

- *Task 3.1- Specialize the system for a cric procedure on a patient simulator in an austere environment. The collaborators helped drafting and revising the manuscript submitted to Nature Digital Medicine.*

- *Collaboration: Dr. McKee and Dr. Kirkpatrick collaborated with the project staff.*

Organization Name: Madigan Army Medical Center
Location of Organization: Joint Base Lewis-McChord, Washington, USA
Partner's contribution to the project (identify one or more)

- *Task 3.1- Specialize the system for a cric procedure on a patient simulator in an austere environment. The collaborator helped revising the manuscript submitted to MICCAI 2020. Additionally, he recruited 5 expert emergency physicians to validate the instructions predicted by our AI mentoring agent.*
- *Dissemination: The collaborator presented a demo of the platform in the Augmented & Virtual Reality in Medicine Workshop, in the 2019 Scientific Assembly of the ACEP.*
- *Collaboration: Dr. Couperus collaborated with the project staff.*

8. SPECIAL REPORTING REQUIREMENTS

COLLABORATIVE AWARDS: For collaborative awards, independent reports are required from BOTH the Initiating PI and the Collaborating/Partnering PI. A duplicative report is acceptable; however, tasks shall be clearly marked with the responsible PI and research site. A report shall be submitted to **https://ers.amedd.army.mil** for each unique award.

QUAD CHARTS: If applicable, the Quad Chart (available on https://www.usamraa.army.mil) should be updated and submitted with attachments.

9. **APPENDICES:** Attach all appendices that contain information that supplements, clarifies or supports the text. Examples include original copies of journal articles, reprints of manuscripts and abstracts, a curriculum vitae, patent applications, study questionnaires, and surveys, etc.