Proceedings of the 2023 Design of Medical Devices Conference
DMD2023
April 17-19, 2023, Minneapolis, MN, USA

DMD2023-3868

DEPLOYING COMPUTER VISION DETECTION METHOD IN MEDICAL SIMULATION TRAINING USING MACHINE LEARNING

Hang-Ling, Wu
The Pennsylvania State University
University Park, PA

Dailen Brown
The Pennsylvania State University
University Park, PA

Scarlett Miller
The Pennsylvania State University
University Park, PA

Jason Moore
The Pennsylvania State University
University Park, PA

ABSTRACT

A machine learning (ML) object detection algorithm was developed to replace the original color-based image detection algorithm for the Dynamic Haptic Robotic Trainer Plus (DHRT+). This image recognition system was used for medical training in Central Venous Catheterization (CVC). This image tracking allows for the training system to provide accurate performance feedback to the user during the training process. The ML object detection algorithm was developed and evaluated using training data. The results indicate that increasing the training data set improves the detection system's accuracy. The system was found to have an overall precision rate of 90.9% and a recall rate of 81.69%. This new ML model will be implemented into the DHRT+ system and used to train medical residents.

Keywords: Medical Simulation Training, Central Venous Catheterization, Computer Vision, Machine Learning

1. INTRODUCTION

The primary goal in medical training for residents is to provide a comprehensive skill set for their future career [1]. To deliver medication safely and efficiently, medical residents need to spend a significant amount of time to understand and master these skills. Thus, providing residents practice and feedback will help them better understand the procedures [2].

One of the common techniques to effectively train medical residents is by practicing with medical manikins. In this process a medical expert supervises the residents during the training process. This type of training provides a clear vision of procedure and accurate feedback from the expert to the residents. However, this method might not be as efficient as training through a simulation device [3]. Training with medical experts is less efficient both in cost and time. Difficulties arise in scheduling with experts due to other hospital needs. Additionally, Experts must be paid for their involvement in the training events. The time available for residents to practice is also limited based on the time of the scheduled sessions.

Simulation devices have the ability to provide automatic feedback without a medical expert, and also allows for flexibility in scheduling time for the residents to train. Such simulation systems would be beneficial for resident training regimes for Central Venous Catheterization (CVC), which is a common procedure used to provide medication to the heart by placing a catheter into a central vein of the patient [4]. The CVC surgery involves 15 major steps to complete [5] but the current training manikins for CVC only allow the insertion of the needle by the guidance of ultrasound, but does not allow the insertion of any other surgical tools. To address the above issues, recently a CVC simulation device has been developed to encounter these problems. The system, the Dynamic Haptic Robotic Trainer Plus (DHRT+) is a medical device targeted toward training medical residents to perform CVC [6].

The DHRT+ allows medical residents to perform CVC surgery from cutting and dilating the skin surface to inserting the catheter. In order to track if the resident is using the tools in the correct sequence, the DHRT+ uses Computer Vision (CV) methods. Previously, the CV system used color-based algorithms to identify the surgical tools used on the DHRT+, and it can identify the tools with high accuracy [6]. However, the CV system is less robust under certain circumstances such as extreme differential light exposure. Therefore, it is necessary to explore the use of Machine Learning (ML) algorithms. ML algorithms use mathematical models to predict the output from the given input data. It is used in various industries and provides significant improvement in object-based detection [7].

Currently, the implementation of ML in medical devices is focused on the accuracy of object detection in medical image analysis [8-9], providing statistical feedback for big data [10], and machine learning in virtual reality simulation for medical education [11]. The use of ML in medical simulation devices for tracking small tools is yet to be discussed. This paper presents the process, results, and conclusion for applying and evaluating the performance of a ML object detection algorithm for medical CVC training.

2. MATERIALS AND METHODS

The primary physical component used in the experiment is the DHRT+ shown in Figure 1. The system incorporates a single overhead camera which provides video data of the whole DHRT+ platform. With the assistance of an additional ring light, it provides more variety of operational environments. This platform separates into two major areas of interest. The top portion represents the medical tray, and the lower part represents the patient body, see Figure 2. The overhead camera tracks the position of the tools and the system provides feedback to the medical residents on whether they are performing the surgery with the correct tool order. The camera being used in DHRT+ is an Arducam 16MP camera module (4456x3496 max resolution).

FIGURE 1: DHRT+ CAMERA SYSTEM

FIGURE 2: DHRT+ PATIENT BODY AND MEDICAL TRAY

The algorithm which was deployed into the system is an open-source code name YOLOv5 [12]. The coding process is thus: First the environment of YOLOv5 is established, then the required modules are imported. Next, the training parameters are set, such as bench or epochs. Then the training data is imported into python. In order to run YOLOv5 to train the ML model, the labeling must be completed before the actual training occurs. Online ML tools by Roboflow (Des Moines, IA) were used to create the labels for the ML data set. Roboflow is a web-based service to create labeling or even training for machine learning models. Figure 3 shows the medical tools the algorithm is trained to detect, and Figure 4 gives an example of a labeled input image.

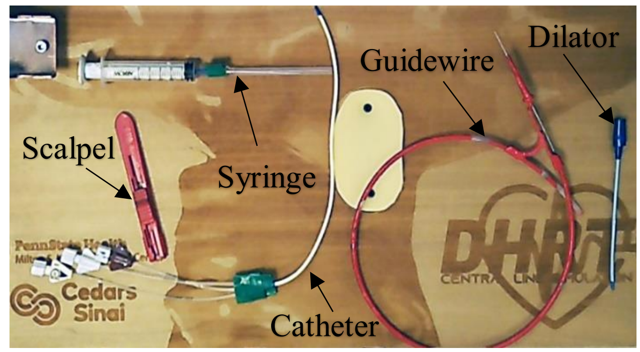

FIGURE 3: MEDICAL TOOLS THE ML IS TO IDENTIFY

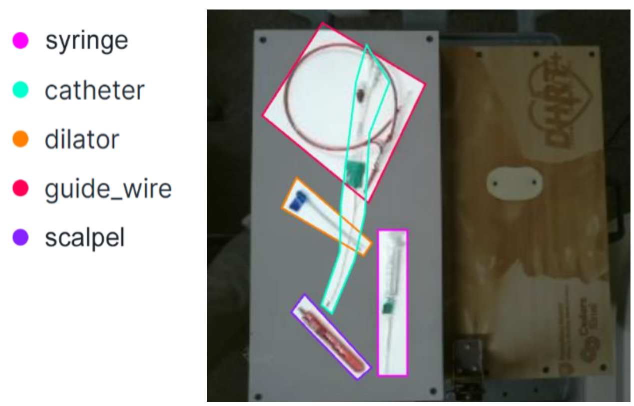

FIGURE 4: SAMPLE OF ANNOTATED IMAGE LABELS FOR ML TRAINING IN ROBOFLOW

To train the ML model, the YOLOv5 algorithm reads in the labeled data and, based on the characteristics of each image, builds the image detection algorithm. After the system finishes training the algorithm, it forms a file which can then be called as a function in python. During the development, the training data can be considered the most direct influence of the results toward the ML output.

Three different algorithms were developed based on the number of training images in the data sets, from 100, 300, and 800. The training data was captured through the overhead camera on DHRT+ and labeled with Roboflow. The first 100 training sets contain the medical tools distributed randomly on top of the DHRT+ surfaces. A similar method was used to expand the database from 100 to 300 for the second ML algorithm. The final set was created based on these 300 images, using an image augmentation method, provided in Roboflow, to expand the data into 800 images. This augmentation method includes image rotation, light exposure rate, or mirroring the image.

After the training process was complete, the validation data was passed through the system to test the accuracy and robustness of the system. Fifty validation images were taken in 4 different conditions to ensure the consistency of the algorithm. These validation images were collected by the same method as the training data, with the tools randomly placed on top of the tray. The only difference between the two data sets is the different environmental conditions, which can help validate the system under different circumstances.

Two metrics, the precision rate and the recall rate, were assessed to determine accuracy of the machine learning model. The precision rate equation in machine learning code is calculated in Equation 1:

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

where PR is the precision rate, TP is the count of true positives, and FP is the count of false positives. A true positive is defined as an object which was detected in the image and was actually there. A false positive is an object that was detected in the image but was not actually there. The recall rate is calculated in Equation 2:

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

where RR is the Recall Rate, TP is the count of true positives, and FN is the count of false negatives. A false negative is defined as an object which was in the image but was not detected by the algorithm.

3. RESULTS AND DISCUSSION

The precision rate, shown in Figure 5, represents the number of tools correctly identified out of the tools detected on the tray, computed over 50 validation images. To ensure the machine learning system is more robust than the previous image recognition algorithm, it is essential to expose the system to various circumstances, such as indoor or natural light exposure. The hypothesis was that as the size of the training data set increases, the precision rate of the system will also increase. This statement holds throughout all environment settings except location A with lighting support. The precision rate at location A without light starts at 93.11% for 100 images, then moves to 94.33% with 800 training images. Location B without light and with light also present a similar trend. But for location A with light, the precision recorded by the 100 and 300 data sets increased from 83.71% to 87.5% but the precision rate drops back to 86.61% when increased to the 800 data set. During the experiment for location A with light, the ML algorithm had a consistent error of detecting the syringe when it was not present, a false positive.

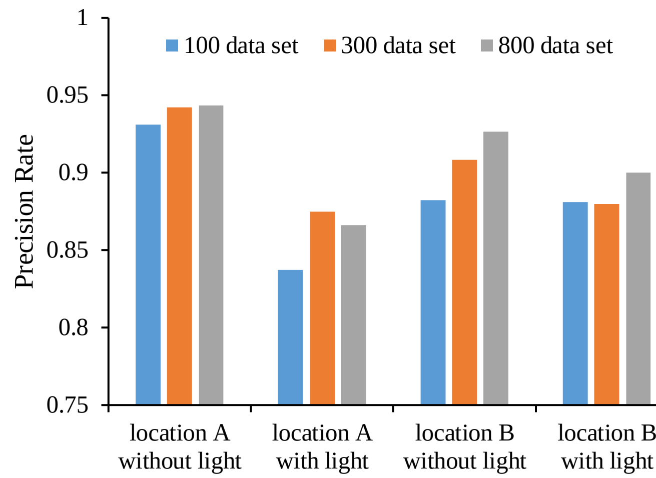

FIGURE 5: OVERALL PRECISION RATE OF 100, 300, AND 800 DATA SET

The recall rate, shown in figure 6, represents how many of the tools actually on the tray were correctly identified by the algorithm. One significant result is that the results do not match the hypothesis that with a higher number of images in the data set, the recall rate would be better. However, only location A without light and location B with light provided this result. Location A with light generated a negative trend from 87.22% to 77.83%—location B without light yielded from 75% to 69.1%, then back to 73.88.

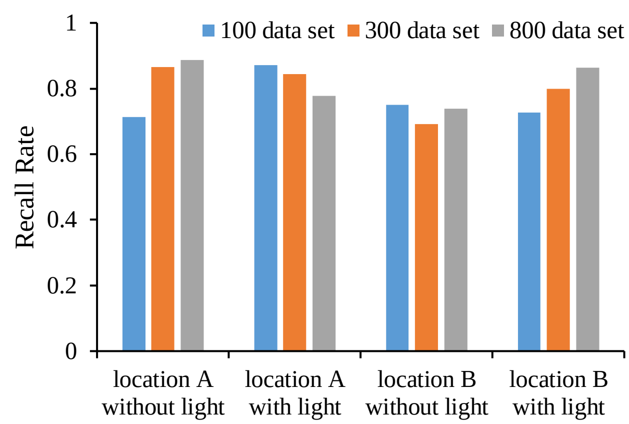

FIGURE 6: OVERALL RECALL RATE OF 100, 300, AND 800 DATA SET

Demonstrating the robustness of the ML algorithm, the output from the 800 data set resulted in a precision rate of 86% and above across the different environments, while also providing an average recall rate of 81.69%, see Figure 7.

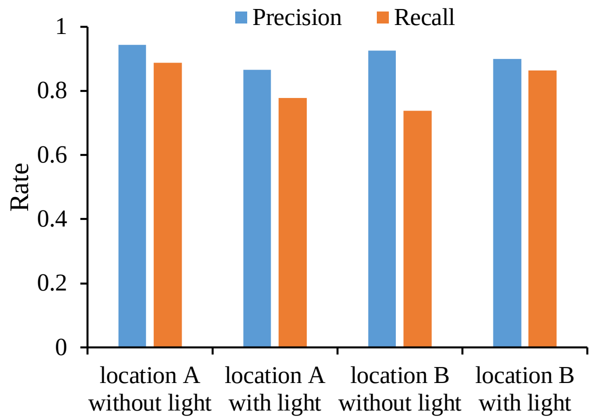

FIGURE 7: OVERALL PERFORMANCE OF 800 DATA SET ML TRAINING MODEL

Figure 8 shows the average precision and recall rates for each data set across all testing environments. On average, both precision and recall improved as the size of the data set increased, with the precision rate increasing from 88.28% to 90.9% and the recall rate increasing from 76.57% to 81.69%.

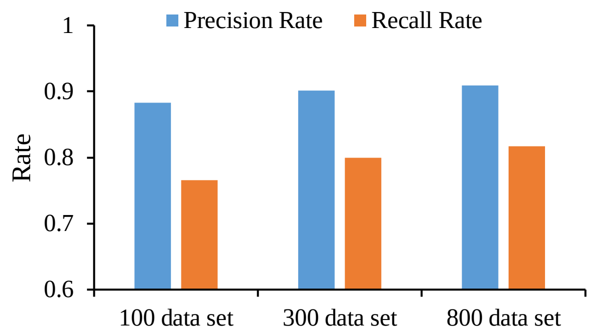

FIGURE 8: AVERAGE PRECISION AND RECALL RATE FROM 100, 300, AND 800 DATA SET

Table 1 shows the relationship between the ring lighting and the precision rate. The precision rate was always lower with the ring light than without the ring light. Location A had an average drop of 7.95% and location B had an average drop of 1.87%, and the overall precision rate dropped by 4.91%. The reason is likely due to the training data sets. Most of the training data were taken under the condition without ring light on, and this might have caused the model to favor the environmental condition without the ring light. Thus, by adding more variety in the data set, such as ring lighting, there could be a raise in the robustness of the ML model.

The current ML system provides an overall precision rate of 90.9% with the recall rate of 81.69%. One way to increase the system accuracy is to average out the response from the system. Those 50 sets of validation data are based on individual images instead of a live recording. When the system is running in real time, it would be possible to average the results across video frames. This will help the system automatically eliminate the outliers, thus increasing the accuracy. Furthermore, accuracy could be improved by further increasing the number of images in the set. Future work will look to increase the accuracy of the system by developing a new custom training algorithm designated for the DHRT+ system rather than using an open source algorithm such as YOLOv5. This ML algorithm could fit the requirements of the DHRT+ system, such as an increase in the sensitivity toward recognizing the shape instead of the color of the tools.

Given the methodology employed in the data collection and experimentation, there is a potential for the model to be overfit to the system. However, the system is specifically designed for use as a central venous catheterization training device and is not intended for deployment on other medical devices or surgical tools. The objective of the machine learning algorithm is to ensure that the DHRT+ device can operate under a variety of circumstances and environments. The results demonstrated across different locations and lighting conditions validate the algorithm. Therefore, while overfitting of the machine learning model may be a concern in general, it is not necessarily a negative aspect for this specialized system and may even provide valuable feedback for the DHRT+ device.

TABLE 1. PRECISION DROP RATE FROM RING LIGHT

	Location A	Location B
100 Data Set	9.40%	0.11%
300 Data Set	6.72%	2.83%
800 Data Set	7.72%	2.67%
Average Rate	7.95%	1.87%
Overall Average	4.91%	

4. CONCLUSION

Deploying an ML system into medical simulation devices can help further advancement for training methods. Combining the ML model into DHRT+ helped to increase the accuracy compared with the original image recognition algorithm. Throughout the training process of the ML model, expanding the training sample from 100, 300, and 800 can help raise both precision and recall rates. The ML model was tested under different environmental settings to ensure system consistency, such as light exposure. It found that the precision rate gradually increased with the amount of data set, from 88% to 91%. The recall rate increased from 76% to 81%. The additional assistance from the ring light on DHRT+ did not provide a positive contribution, instead causing a decrease in precision rate at an average of 4.9%. The system still needs accuracy improvement and validation testing under different environments before deploying to an actual training machine in the medical center. Thus, there will be further work to improve the DHRT+ and ML systems. Nevertheless, these experiments show that escalating the training data set can increase the overall performance of the ML system.

ACKNOWLEDGEMENTS

Research reported in this publication was supported by the National Heart, Lung, and Blood Institute of the National Institutes of Health under Award Number R01HL127316. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Coauthors Dr. Moore and Miller own equity in Medulate, which may have a future interest in this project. Company ownership has been reviewed by the University's Individual Conflict of Interest Committee and is currently being managed by the University.

REFERENCES

[1] Monrouxe, L. V., 2010, "Identity, Identification and Medical Education: Why Should We Care?," Med. Educ., **44**(1), pp. 40–49.

[2] Datta, R., Upadhyay, K. K., and Jaideep, C. N., 2012, "Simulation and Its Role in Medical Education," Med. J. Armed Forces India, **68**(2), pp. 167–172.

[3] Okuda, Y., Bryson, E. O., DeMaria, S., Jacobson, L., Quinones, J., Shen, B., and Levine, A. I., 2009, "The Utility of Simulation in Medical Education: What Is the Evidence?," Mt. Sinai J. Med., **76**(4), pp. 330–343.

[4] McGee, D. C., and Gould, M. K., 2003, "Preventing Complications of Central Venous Catheterization," N. Engl. J. Med., **348**(12), pp. 1123–1133.

[5] Chen, H.-E., Yovanoff, M. A., Pepley, D. F., Sonntag, C. C., Resident Physician, M., Mirkin, K. A., Han, D. C., Professor of Surgery, M., Moore, J. Z., Professor Mechanical Engineering, A., Miller, S. R., and Professor, A., "Can Haptic Simulators Distinguish Expert Performance? A Case Study in Central Venous Catheterization in Surgical Education."

[6] Brown, D., Wu, H., Satpathy, Y., Gonzalez-Vargas, J. M., Tzamaras, H., Miller, S., and Moore, J., 2022, "Computer Vision Enabled Smart Tray For Central Venous Catherization Training," *Frontiers in Biomedical Devices, BIOMED - 2022 Design of Medical Devices Conference, DMD 2022*.

[7] Kumar, Y., Kaur, K., and Singh, G., 2020, "Machine Learning Aspects and Its Applications towards Different Research Areas," *Proceedings of International Conference on Computation, Automation and Knowledge Management, ICCAKM 2020*, pp. 150–156.

[8] Erickson, B. J., Korfiatis, P., Akkus, Z., and Kline, T. L., 2017, "Machine Learning for Medical Imaging," Radiographics, **37**(2), pp. 505–515.

[9] Tajbakhsh, N., Hu, Y., Cao, J., Yan, X., Xiao, Y., Lu, Y., ... & Ding, X. (2019, April). "Surrogate supervision for medical image analysis: Effective deep learning from limited quantities of labeled data," *2019 IEEE 16th International Symposium on Biomedical Imaging,* pp. 1251-1255.

[10] Xu, S., and Chan, H. K., 2019, "Forecasting Medical Device Demand with Online Search Queries: A Big Data and Machine Learning Approach," *Procedia Manufacturing*, Elsevier, pp. 32–39.

[11] Winkler-Schwartz, A., Bissonnette, V., Mirchi, N., Ponnudurai, N., Yilmaz, R., Ledwos, N., ... & Del Maestro, R. F. (2019). "Artificial intelligence in medical education: best practices using machine learning to assess surgical expertise in virtual reality simulation, " *Journal of surgical education*, 76(6), 1681-1690

[12] Sadiq, M., Masood, S., and Pal, O., 2022, "FD-YOLOv5: A Fuzzy Image Enhancement Based Robust Object Detection Model for Safety Helmet Detection," Int. J. Fuzzy Syst., **24**(5), pp. 2600–2616.