Real-time Object Detection in Sterile Compounding Areas

1st Andrew Waggoner
University of Tennessee
Online Master of Computer Science
Knoxville, TN, USA
awaggon2@vols.utk.edu

2nd Kyle Russell
University of Tennessee
Online Master of Computer Science
Huntsville, AL, USA
krusse15@vols.utk.edu

3rd Robert Owens
University of Tennessee
Online Master of Computer Science
Mount Pleasant, NC, USA
rowens24@vols.utk.edu

***Abstract*—Real-time object detection in sterile compounding areas is a complex task in a difficult environment that can yield impactful results. In this paper we outline a method using computer vision on a data set of images from sterile compounding areas and achieved and overall mean average precision (mAP) of 90%. Our method can be added to typical verification methods to improve or enhance the safety of sterile compounding by detecting objects such as IV bags, vials, and syringes.**

***Index Terms*—machine learning, deep learning, sterile compounding**

I. Introduction

Intravenous medications are prepared in hospital pharmacies in highly regulated, controlled, sterile compounding environments. Preserving the sterility and stability of admixtures in these rooms requires minimizing disturbances to airflow, which can be achieved by limiting the frequency of entering and exiting the room.

Acute care environments include all areas served within a hospital's scope of practice, including emergency rooms, intensive care units, general medicine services, and some ambulatory services provided within the hospital premises. Pharmacists operating in acute care environments are responsible for verifying all doses that leave the pharmacy, which includes drugs prepared in sterile compounding suites as well. Many technologies have been created to aid in verifying sterile compounded products without disturbing the relative sterility in the room, including gravimetric analysis and remote cameras.

The current standard evaluation methods are gravimetric analysis, barcode scanning, and remote viewing. Although these methods are accurate, their combination with current practices and workarounds can still result in compounding errors.

In this project we set out to answer two major questions.

1) Can computer vision and object detection provide an additional safety check without inserting an individual into the sterile compounding suite?
2) Can the model be trained not only to recognize IV bags, syringes, and vials but also the drug that was likely in the vial and the diluent in the bag?

We set out to answer these in a phased approach:

- Phase I: Can we classify the images as vial, bag, or syringe?
- Phase II: Can we additionally label the IV bags according to their diluent, i.e. dextrose, sodium chloride, lactacted ringers, etc.?
- Phase III: Can we also classify the vials as to the drug that it contains?

II. Related Work

A review of the current literature reveals several applications of computer vision in healthcare, including its use in sterile compounding environments. Some of those applications involve detecting object movements or monitoring sterile compounding techniques. [1], [2]

SterileAR, a computer vision-assisted training program, shows users flaws in their technique with a combination of augmented reality and computer vision. [2] It identifies and tracks IV bags, vials, and syringes and the users relative and absolute hand positions. The models used in this product were COCO SSD MobileNet v1 and Faster R-CNN. The purpose was not entirely for accurate identification, but the researchers commented that they abandoned the use of TensorFlow based models after several failed attempts to achieve the desired results. They did, however, achieve fairly good results at identifying some of the objects in a general sense after moving to a more traditional OpenCV based approach.

Qi and Lee conducted a study aiming to track objects in the sterile compounding process in a frame-by-frame fashion. [1] They preprocessed each frame with a Gaussian filter and then identified the object via thresholding. Each frame was compared to the previous frame to determine the borders of the object. They tracked the objects across the subsequent frames to "track" them across the workspace. They chose to use the median flow tracker and discriminate correlation-based filter tracking from OpenCV. They achieved 84 and 80 tracking accuracy for single and multi-object tracking respectively. Their method of object tracking improved the accuracy of object tracking, 81.66 versus 66.51, but it came with a higher processing time, approximately 1 second on average.

Eppel, et al. used computer vision to identify and detect liquid samples in hospitals and labs with a combination of two neural networks. [3] These networks were arranged in a

cascading manner, with one network responsible for identifying the vessel containing the sample, while the other focused on recognizing the sample itself. The researchers used a Mask RCNN object detection net for identifying the vessel and a FCNN for detecting the samples. They had a good accuracy rate in identifying the vessels, but struggled with non-liquid samples such as blood and solids.

III. Dataset

Our data set was images collected during the compounding process within several hospital sterile compounding suites. These are obtained by the verification process by a vendor, BD Pyxis IV Prep. The proper data governance and health system leadership were notified of this team's desire to investigate these images and gave their approval.

We had approximately 4,700 images for this project. The images contained common sterile compounding items drug vials, syringes (with and without needles attached), IV diluent bags, and randomly fluid tubing or baskets.

(a) Normal Saline Bag (b) Syringe and Vial

Fig. 1. Examples of images used.

It is important to note that the IV bags and drug vials have labels that are standard and branded according to FDA and manufacturer guidelines. This standardization could aid in the applicability of the data set to other similar images and applications. Of course there are many manufactures who could produce a certain chemical substance but with sufficient training it should be possible that the training from our data is somewhat transferable.

The images used were taken in different sterile compounding environments with different lighting and shadow effects on the images. This seemed to be a complication to developing an accurate model but also added some "noise" that made the model more robust once our training data got past a certain threshold.

IV. Methods

In the course of the literature it became apparent that there were very little object detection or classification algorithms used to aid in verification of items used in the sterile compounding process. In addition the publicly available image data sets did not contain items commonly used in the sterile compounding process. We decided that it was best to extend a currently available network to label previously unlabeled objects for use case that we could not find in the literature.

We used YOLO (You Only Look Once) version 8 to for object detection and identification. This would not only leverage publicly available data sets that YOLOv8 has been trained on, it will also allow us to custom label and train the model to detect the classes we are interested in. YOLOv8 is a fast accurate network that is relatively easy to use, especially when compared to R-CNN and F-CNN networks that was used in the related works we uncovered during our literature review. YOLO does not have the same flexibility that R-CNN or Faster R-CNN but makes up for that in fast detection and ease of use.

In our first attempt at object detection we used a publicly available image data set, Common Objects in Context (COCO). We wanted to investigate if the training weights could inform the model enough to detect syringe, vials, or bags for our purposes. [6] It did not identify many of the items in our data set. It did detect vials but improperly labeled them as "bottles". The COCO data set has 1.5 million object instances but IV bags, drug vials, and syringes are not common enough to be included. This indicated that our next steps would be to label our own images.

Fig. 2. Initial Prediction with Trained with COCO Data Set

We then turned to labeling images to train the model on a custom labeled data set. Even though we started with a small number of custom labeled images (twenty), the results were promising. The model was now identifying some of the classes a majority of the time. The accuracy was not as high as we would have liked with vials being the only reliable classification.

We then progressed to labeling 50 images between training, testing, and validation sets. The model was run for 100 epochs and accuracy and loss values looked much better. The images had bounding boxes in roughly the correct place and the labels were accurate most of the time. The problem we still were

trying to solve was labeling all items in the images with a bounding box and correct label text.

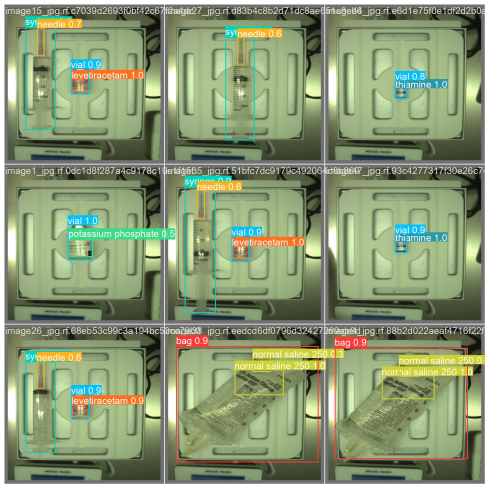

Fig. 3. Training with Custom Labeled Data Set with 100 Epochs

Our next step was too increase the labeled data to 200 or more images and train the model again. We augmentation techniques to increase the number of labeled images and improve the ability of the model to be more generalized. We took examples of correct classifications and put them back into the training data. We also wanted to leverage the model we trained and use it to assist in labeling more images. That proved to be very useful once we trained the model on more than 200 labeled images. Those additional model-assisted labeled images boosted accuracy and made the assisted labeling even better. We continued iterating this approach until we had a labeled data set of 1,000 images.

V. Results

A. Phase I Results

Phase I results were excellent. The mAP50-95 started increasing with only a few more labeled images for training. As the number of images increased to 700 and 1,000 the results continued to improve. As we increased our labeled images we also started creating more specific labels. For example bags were now, 0.9% Sodium Chloride Injection, 5% Dextrose Injection, Exactamix bag, etc. This affected the subsequent classifications but the final model ended up surpassing the original models results. This is summarized in Table I. For the purposes of a fair comparison over time we average the mAP50-95 over all the sub classes of bag and vial.

B. Phases II and III Results

We were able to tackle phases II and III concurrently. The solution for improving the results for each of these phases were the same steps. We ended up with 40 labled classes and the summary of the mAP50-95 for the final model are shown in Table II.

TABLE I
Evolution of Results (mAP50-95) for the Phase I Classes

Class	20 Images and 100 Epochs	700 Images at 50 Epochs	1,000 Images at 100 Epochs*
bag	0.796	0.681**	0.848**
vial	0.897	0.869**	0.979**
syringe	0.908	0.88	0.879
needle	0.654	0.772	0.805

* Model Assisted Labeling.
** Changed to more specific classes in future iterations. This represents the average of all classes of that broader category.

TABLE II
Final Model Results (mAP50-95) for the All Classes

Class	Final Model mAP50-95
0.9% Sodium Chloride Injection	0.766
5% Dextrose Injection	0.747
Arsenic Trioxide	0.895
Azacitidine	0.895
Cyclophosphamide	0.895
Folic Acid	0.995
Iron Sucrose	0.995
Leucovorin Calcium	0.881
Levemir	0.968
Levetiracetam	0.995
Ogivri	0.703
Ondansetron	0.995
Oxaliplatin	0.56
Paclitaxel	0.995
PhaSeal	0.852
Potassium Phosphates	0.995
Sterile Water	0.908
Thiamine HCl	0.955
Vancomycin HCl	0.995
carboplatin	0.995
exactamix bag	0.805
flourouracil	0.995
irinotecan hydrochloride	0.895
needle	0.805
small syringe	0.871
syringe	0.879
vedolizumab	0.995

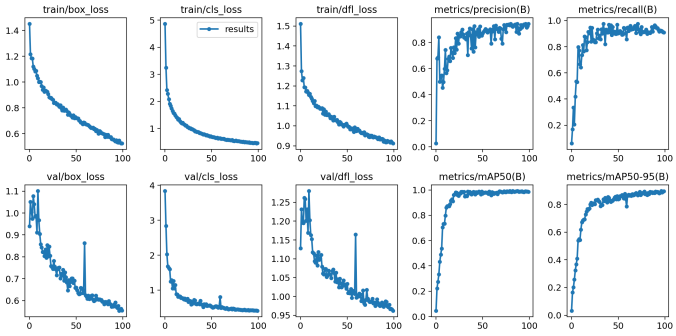

Fig. 4. Training Results

VI. Discussion

The results of the first phase of our project are better illustrated as an evolution of our approach. This learning that

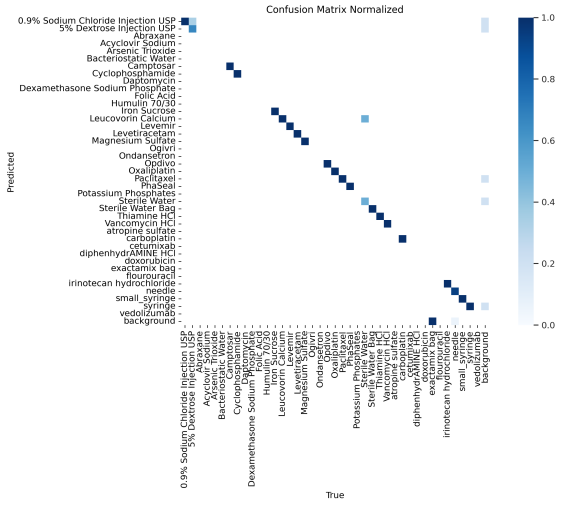

Fig. 5. Training Confusion Matrix

we gathered throughout the process helped inform our overall model creation. As we increased our number of labeled images the accuracy improved across the board for bag, vial, syringe, and needle. Further optimizations were employed to increase the accuracy for the other classes.

One of the first optimizations we employed was making sure the data labels were rotated if the items in the image were rotated. That provided a better match of the areas of interest in the image. This improved detection of that class and the overall accuracy of the model.

From this project the most common indication of the model accuracy for a given class seems to be directly related to the number of labeled images in the training set. We used data augmentation to add more correctly labeled images back into the training data set. Additionally the variations in lighting conditions and item position helped to provide a ”noise” and after training make the model more robust. This noise helped the model recognize new, unseen, items and helped with detection even if the items are positioned in an strange or unique way in the image.

Our dataset would be improved by having a small number of background images to train against, as our training set only included images with medical objects in the foreground. This would help our model detect objects in some of the more noisy environments.

Future directions of study would first start with a larger data set. Although our model worked well the data we had it would still need to be tested at scale. Hospital sterile compounding environments can produce thousands of sterile compounds per day. It is still unknown if this model would generalize enough for even more edge cases.

Another area of study that was postulated was deploying a more robust version of this model in such a way that you can use a smart phone to do real-time object detection. That would require formalizing this model into a production ready version and deploying as an API or mobile application. It could give end users the ability to use a common object to aid in the identification of large batches of sterile compounds.

Additional future areas of study would include combining the object detection with segmentation. The syringes had very distinct features in the images and when combined with segmentation of the syringe could help identify or calculate the percentage of the syringe that contains fluid. If you could add a verification of the product and the amounts as another check that would increase the utility of such an approach.

VII. Conclusion

The overall results were very good. We successfully trained an existing model on custom labeled data that had not previously been classified in the literature. The main indicator for generating a correct prediction was the number of annotations for a given class label. We would expect that as we correct for class imbalance in our dataset, our model’s predictions would become more robust. Slight variations in the lighting and item placement added an overall robustness to the model once we crossed a certain threshold of labeled images. The model-assisted labeling and increasing the number of correctly labeled images via data augmentation proved extremely useful in creating a high quality data set in a short period of time.

References

[1] S. Eppel, H. Xu, M. Bismuth, and A. Aspuru-Guzik, “Computer Vision for Recognition of Materials and Vessels in Chemistry Lab Settings and the Vector-LabPics Data Set,” ACS Cent. Sci., vol. 6, no. 10, pp. 1743–1752, Oct. 2020, doi: 10.1021/acscentsci.0c00460.

[2] Z. Qi and J. K. Lee, “Automated Object Tracking in Sterile Pharmacy Compounding,” in Computer Science Research Notes, Západočeská univerzita, 2019. doi: 10.24132/CSRN.2019.2901.1.16.

[3] “ISMP Survey Provides Insights into Pharmacy Sterile Compounding Systems and Practices,” Institute For Safe Medication Practices, Oct. 21, 2020. https://www.ismp.org/resources/ismp-survey-provides-insights-pharmacy-sterile-compounding-systems-and-practices (accessed Mar. 27, 2023).

[4] “Maximize Benefits of IV Workflow Management Systems by Addressing Workarounds and Errors,” Institute For Safe Medication Practices, Jan. 05, 2018. https://www.ismp.org/resources/maximize-benefits-iv-workflow-management-systems-addressing-workarounds-and-errors (accessed Mar. 27, 2023).

[5] D. Babichenko et al., “SterileAR: Exploration of Augmented Reality and Computer Vision Approaches for Real-Time Feedback in Sterile Compounding Training,” in 2020 6th International Conference of the Immersive Learning Research Network (iLRN), Jun. 2020, pp. 62–69. doi: 10.23919/iLRN47897.2020.9155164.

[6] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár and C.L. Zitnick, ”Microsoft COCO: Common Objects in Context,” in European Conference on Computer Vision (ECCV), 2014.