J. Mantas et al. (Eds.)

doi:10.3233/SHTI240572

Detection of Pharmaceutical Personal Protective Equipment in Video Stream

Michael LEONTIEV[a,1] , Danil ZHILIKOV[a] , Lenar KLIMOV[a], Vyacheslav CHERTAN[a], Daniel BOBROV[a], Dmitry LOBANOV[a], Vasilii VOLOGDIN[a], Vladislav MASLOV[a] and Ksenia BALABAEVA[a,2]

[a] *BIOCAD, Saint Petersburg, Russia*

ORCiD ID: Danil Zhilikov https://orcid.org/0009-0002-2322-2872,Lenar Klimov https://orcid.org/0009-0006-4181-9583, Daniel Bobrov https://orcid.org/0009-0004-7120-9327, Michael Leontiev https://orcid.org/0009-0008-7739-7356, Dmitry Lobanov https://orcid.org/0009-0007-1265-3964 ,Vladislav Maslov https://orcid.org/0009-0007-4570-1272, Ksenia Balabaeva https://orcid.org/0000-0002-9138-0736

Abstract. Pharmaceutical manufacturing is a complex process, where each stage requires a high level of safety and sterility. Personal Protective Equipment (PPE) is used for this purpose. Despite all the measures of control, human factor (improper PPE wearing) causes numerous losses for human health and material property. This research proposes solid computer vision system for ensuring safety in pharmaceutical laboratories. For this we have tested wide range of state-of-the-art object detection methods. Composing previously obtained results in this sphere with our own approach to this problem, we have reached the high accuracy (mAP@0.5) ranging from 0.77 up to 0.98 in detecting all the elements of common set of PPE used in pharmaceutical laboratories. Our system is the step towards safe medicine producing.

Keywords. sterility and safety in pharmaceutical development, personal protective equipment, computer vision, object detection, monitoring in pharmaceutical development, PPE detection

1. Introduction

Pharmaceutical manufacturing involves a complex process from biochemical design to large-scale production, requiring sterile zones in labs where Personal Protective Equipment (PPE) like gloves, masks, and protective suits are mandatory to prevent contamination. However, human errors in adhering to these protocols can jeopardize both health and product integrity. Traditional manual monitoring of PPE compliance is becoming outdated, prompting a shift towards using convolutional neural networks (CNNs) for better enforcement.

Our study focuses on employing advanced CNNs, notably YOLOv8s [1], for real-time detection and proper usage monitoring of PPE in pharmaceutical environments,

1 Corresponding Author: Michael Leontiev, michlea@yandex.ru.
2 Corresponding Author: Ksenia Balabaeva, kyubalabaeva@gmail.com

alongside an alert system for protocol breaches. This research aims at enhancing safety by ensuring PPE compliance through AI, covering both visitor screening and sterile zone monitoring. In conclusion, the primary contributions of our project include: 1) the compilation of a processed dataset featuring 889 instances across 6 types of PPE: masks, suits, shoe covers, gloves, glasses and hoods; 2) the testing of diverse CV methods utilizing our dataset; and 3) the development of a ready-to-deploy open-source computer vision system, the code of experiments for real-time PPE detection is available in GIT.

2. Related Works

The Covid-19 pandemic has led to numerous studies on mask detection, with a lesser focus on the broader medical Personal Protective Equipment (PPE) detection. Most use the YOLO (You Only Look Once) algorithm for real-time detection due to its efficiency in predicting bounding boxes and class probabilities [2]. Wu et al. [1] enhanced YOLOv5 for PPE detection, reaching a mean Average Precision (mAP) of 0.97 at 53 frames per second. Other research focused on improving mask detection through modifications to the YOLO algorithm and introducing novel structures like mask-guided modules [3]. However, some methods, such as combining RetinaFace with ResNet, do not support real-time detection [4].

The CPPE-5 dataset, introduced by Dagli et al., enriches PPE detection research with over a thousand images and detailed annotations representing a variety of PPE types. While notable advancements in machine learning architectures for mask and PPE detection are evident, focusing predominantly on mask detection, limitations in comprehensive PPE compliance, feature extraction in complex environments, and performance metrics for time and spatial accuracy remain [6]. Other papers propose the use of CNN however they are primary focused only on mask detection [5].

However, there are notable limitations. Many models focus only on mask detection, neglecting comprehensive PPE compliance, and struggle with feature extraction in complex scenarios. The lack of performance metrics for time efficiency and spatial accuracy, alongside the challenges in computational speed and effectiveness at varying angles, point to areas needing further research.

In sum, while existing studies mark progress in using computer vision for enhancing public health safety, addressing their limitations is essential for broader applicability and improved performance. This includes expanding detection capabilities to cover a wider PPE range and improving feature extraction and performance metrics evaluation, which will enhance their utility in pandemic response efforts.

Future research should aim to expand PPE detection capabilities, improve feature extraction, and refine performance evaluation metrics to enhance the role of computer vision in public health safety during pandemics.

3. Methods

3.1. Dataset

In our study on enhancing PPE compliance in pharmaceutical environments, we developed a dataset comprising 889 images derived from video footage shot in simulated pharmaceutical settings. The simulation of the pharmaceutical process was conducted in educational facilities using artificial lighting. Actors, both with and without personal protective equipment (PPE), moved throughout the room to create a realistic training environment. This approach ensured varied examples of both correct and incorrect PPE usage. The PPE set includes suit with the hood, gloves, shoe covers, glasses and mask (Fig.1) The dataset is somewhat imbalanced, favoring gloves and shoe covers due to the anatomical characteristics of human beings. After extracting pivotal frames representing diverse scenarios of PPE wear, each image underwent detailed labeling to identify the PPE types and their wearing status — present or absent (including cases of incorrect wearing). This meticulous annotation was fundamental for training our computer vision system to discern between proper and improper PPE usage effectively. Since the frames from the video recording are temporally connected, the holdout method was used for splitting the data into training (75%), validation (12,5%) and test (12,5%) sets.

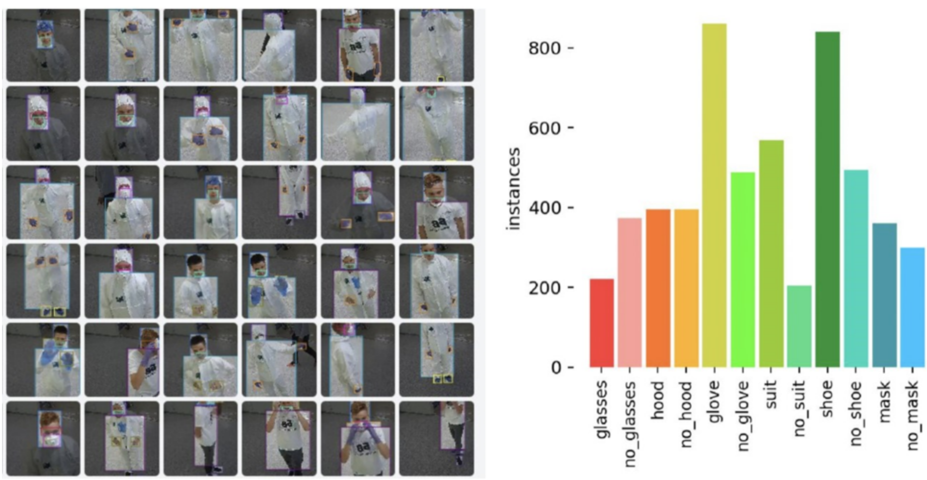

Figure 1. a) Dataset samples (right) and PPE classes distribution (left)

For those interested in exploring our dataset or conducting further research, the dataset is accessible via the following link: Download Dataset . While our primary objective has been to ameliorate safety within pharmaceutical manufacturing by precise PPE detection, we advocate for and encourage the dataset's application in broader PPE compliance research and development initiatives across various industrial domains.

3.2. Models

Convolutional neural networks (CNNs) have transformed digital image analysis, particularly in object detection, and are crucial for ensuring Personal Protective Equipment (PPE) compliance in pharmaceutical settings. Our research compares the efficiency of three advanced models: YOLOv8, Faster R-CNN, and Grounding DINO for this purpose.

YOLOv8, the latest in its series, is recognized for its fast processing speeds, enabling real-time PPE monitoring by predicting bounding boxes and class probabilities efficiently [1]. Faster R-CNN, by contrast, uses a two-step process for object detection, trading off speed for enhanced accuracy, which may benefit thorough PPE checks [8]. The Grounding DINO model, an innovative approach in open-set object detection, allows for the adaptation to a wide range of PPE equipment by using textual prompts instead of a predefined label set, offering potential for dynamic PPE compliance in pharmaceutical manufacturing [7]. Faster-RCNN and YOLOv8 were fine-tuned using transfer-learning of pre-trained weights on collected dataset. DINO-model was tuned using prompt-engineering procedure where we described the type of PPE to make it find them on the image. In summary, the selection of these models for experimental comparison acknowledges the diverse requirements of PPE detection in pharmaceutical settings—from the need for real-time monitoring and high accuracy to adaptability to new PPE types. This comparison aims to pinpoint the most effective AI-driven solutions to bolster safety measures, thereby mitigating risks associated with non-compliance.

3.3. Evaluation

In object detection, choosing the right evaluation metrics is key for assessing model performance. Two crucial metrics are Intersection over Union (IoU) and mean Average Precision (mAP). IoU measures the overlap between predicted and ground truth bounding boxes as a ratio, indicating localization accuracy. Precision evaluates the accuracy of detected objects, while recall measures the ability to detect all relevant objects. mAP averages the precision across all classes and IoU thresholds, providing a comprehensive view of a model's detection and classification performance. Thus, IoU and mAP are essential for evaluating and benchmarking object detection models.

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \quad (1)$$

Where N is the number of classes, and AP_i is the Average Precision for the i^{th}class.

4. Results, Discussion and Conclusions

Our study aimed to improve safety in pharmaceutical facilities by detecting improper PPE wearing. We evaluated various advanced computer vision (CV) models based on their mean Average Precision (mAP) at a 50% Intersection over Union (IoU) threshold.

Among the models tested, YOLOv demonstrated superior performance with a mAP of 88.9, significantly outperforming other models such as Faster R-CNN and GroundingDINO, which achieved mAP scores of 40.4 and 28.5, respectively. This disparity in performance highlights the effectiveness of YOLOv8 in accurately identifying instances of PPE misuse, making it a valuable tool for ensuring safety and compliance in the highly regulated environment of pharmaceutical manufacturing.

The comparative analysis of these models, with a focus on their applicability for PPE detection, has solidified the foundation of our proposed computer vision system designed to mitigate the risks associated with improper PPE usage, thereby contributing to the safer production of medicines 2. The developed computer vision system operates in stages: obtaining a video stream from the camera using the RTSP protocol and Python 3.11. The data then goes to the CV module for processing, where PPE and violation detection are performed using YOLOv8. Interaction with the model is based on FastAPI, and the results are stored in a PostgreSQL database. Additionally, a mechanism for real-time video streaming is implemented using WebSocket for demonstrating the module's operation. The application interface is implemented in Vue.js and is running in Docker. The code of experiments for real-time PPE detection is available in GitHub.

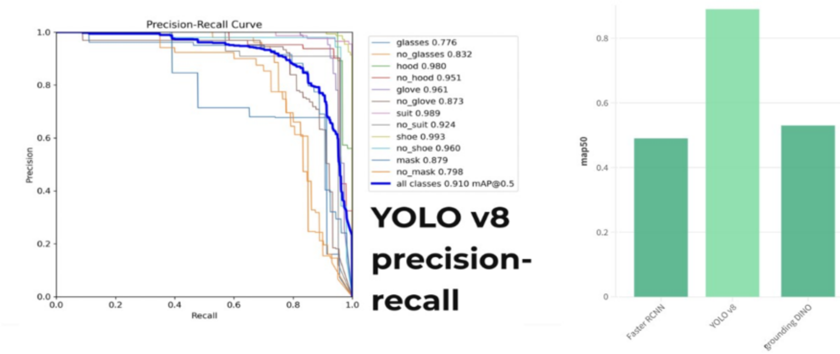

Figure 2. YOLOv8 Precision Recall Curve (left), Models comparison (mAP) (right)

Acknowledgement

The project was mostly implemented by school students during the educational program Big Challenges at the Sirius Education Center in collaboration with biotechnology company BIOCAD.

References

[1] Wu B, Pang C, Zeng X, Hu X. ME-YOLO: Improved YOLOv5 for Detecting Medical Personal Protective Equipment. Appl Sci. 2022;12(23):11978. doi:10.3390/app122311978.

[2] Redmon J. You Only Look Once: Unified, Real-Time Object Detection [Internet]. arXiv.org; 2015 Jun 08 [cited year month day]. Available from: https://arxiv.org/abs/1506.02640.

[3] Yu J, Zhang W. Face Mask Wearing Detection Algorithm Based on Improved YOLO-v4. Sensors. 2021;21(9):3263. doi:10.3390/s21093263.

[4] Cao Z, Shao M, Xu L, Mu S, Qu H. MaskHunter: real-time object detection of face masks during the COVID-19 pandemic. Iet Image Process. 2020;14(16):4359–4367. doi:10.1049/iet-ipr.2020.1119.

[5] Crespo F, Crespo A, et al. A computer vision˜ model to identify the incorrect use of face masks for COVID-19 awareness. Appl Sci. 2022;12(14):6924. doi:10.3390/app12146924.

[6] Dagli R, Shaikh AM. CPPE-5: Medical Personal Protective Equipment Dataset. SN Comput Sci. 2023;4(3). doi:10.1007/s42979-023-01748-7.

[7] Liu S. Grounding DINO: Marrying DINO with Grounded Pre-Training for Open-Set Object Detection [Internet]. arXiv.org; 2023 Mar 09 [cited year month day]. Available from: https://arxiv.org/abs/2303.05499.

[8] Ren S, He K, Girshick R, Sun J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Trans Pattern Anal Mach Intell. 2017;39(6):1137–1149. doi:10.1109/TPAMI.2016.2577031.