ISBN: 978-981-94-1156-6
2024 the 14th International Workshop on Computer Science and Engineering (WCSE 2024)
doi: 10.18178.wcse.2024.06.013

Evaluating YOLO Models for Enhanced Safety in Medication Dispensing

Anan Panphuech [1], Khwanchai Huailuk [1] and Sayan Kaennakham [2 +]

[1] Department of Interdisciplinary Science and Internationalization, Institute of Science, Suranaree University of Technology, Nakhon Ratchasima, 30000, Thailand;

[2] School of Mathematics and Geoinformatics, Institute of Science, Suranaree University of Technology, Nakhon Ratchasima, 30000, Thailand;

Abstract. This study assesses the effectiveness of YOLOv5 and YOLOv8 in enhancing safety in medication dispensing through precise identification of capsules and tablets. Using a dataset of 1,659 images, the models were evaluated across various metrics in training, validation, and testing phases. Results indicate that YOLOv8 outperforms YOLOv5 in most training and validation metrics, while YOLOv5 shows superior performance in testing. These findings highlight the potential of advanced object detection models to improve patient safety by reducing medication dispensing errors, offering valuable insights for the deployment of AI in healthcare environments.

Keywords: YOLOv5, YOLOv8, Pharmaceutical object detection, Deep learning in healthcare

1. Introduction

The accurate identification of capsules and tablets is critical for ensuring the safety and efficacy of pharmaceutical interventions [1]. Precise medication identification prevents adverse drug reactions and therapeutic failures, thereby safeguarding patient health. In healthcare settings where multiple medications are dispensed, such as hospitals and nursing homes, correct identification is essential for minimizing dispensing errors and enhancing treatment outcomes [2]. Consequently, the ability to distinguish pharmaceuticals accurately is a fundamental aspect of clinical safety and care effectiveness [3].

The identification of capsules and tablets has evolved from reliance on manual recognition of physical characteristics to the adoption of deep learning technologies [4]. Deep learning models, trained on extensive image datasets, enhance accuracy in detecting subtle discrepancies in medication appearances, including counterfeits[5]. This advancement not only minimizes human error but also streamlines pharmaceutical management, significantly improving patient safety and adherence to treatment protocols in healthcare environments. Several algorithms have been developed and deployed to address this challenge, including Faster R-CNN and RetinaNet, among others [6]. In particular, our attention is focused on those belonging to the YOLO family. YOLO, introduced by Joseph Redmon and colleagues [7], is an object detection algorithm that utilizes convolutional neural networks (CNN) to identify objects in real-time [8]. This single-stage method operates effectively on standard GPUs and is designed to split the image into a grid of cells, with each cell tasked with detecting objects in a specific area. This structure facilitates quicker object detection compared to traditional two-stage methods, making it especially advantageous for real-time applications. Over time, YOLO has undergone several iterations, each enhancing its speed, accuracy, and ability to detect objects of different sizes [9].

This study concentrates on applying two variants of YOLO algorithms, YOLO5 and YOLO8, to the task of identifying capsules and tablets. It seeks to comparatively assess the performance of these well-known algorithms, with the goal of providing deeper insights into their characteristics and capabilities for this specific task.

+ Corresponding author. Tel.: +0-4422-4383; fax: +0-4422-4185.
E-mail address: sayan_kk@g.sut.ac.th

2. Methodology

2.1. The Two YOLOs

The YOLO architecture consists of three main components: the backbone, neck, and head, each contributing to its speed and accuracy improvements [10].

Backbone: Extracts essential features from images using CNNs trained on datasets like ImageNet, with common backbones including VGG16, ResNet50, CSP-Darknet53, and EfficientRep.

Neck: Serves as a bridge between the backbone and head, integrating features from various layers using techniques like SPP and PAN.

Head: Manages final detection tasks such as bounding box predictions and classification, using single-stage, multi-scale, and anchor-based approaches.

YOLOv5: It focuses on providing a balanced performance between speed and accuracy. Fig 1 outlines the YOLOv5 architecture, consisting of three main parts: Backbone, Neck, and Head. The Backbone features layers like Focus, convolutional, C3 blocks, and Spatial Pyramid Pooling (SPP) for detailed feature extraction. The Neck, incorporating PANet, uses upsampling and concatenation to enhance these features. Finally, the Head employs multiple Conv2d layers to deliver precise object detection and classification across various scales.

YOLOv8: It introduces more sophisticated architectural enhancements and optimizations that improve both the model's efficiency and its detection capabilities across a wider range of object sizes and types. These enhancements often include more advanced attention mechanisms, improved backbone architectures for deeper and more effective feature extraction, and more complex neck designs that further enhance feature integration across scales. Additionally, YOLOv8 may incorporate newer training strategies and loss functions that refine the learning process, leading to higher precision and recall rates. As the full diagram of YOLOv8 consists of many components and occupies considerable space, interested readers are recommended to refer to [12].

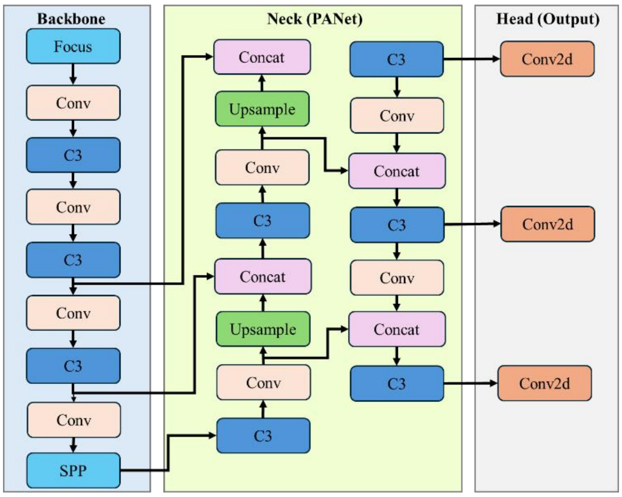

Fig. 1: A typical architecture of YOLOv5 [11].

Conducting a comparison between YOLOv5 and YOLOv8 is essential to identify advancements in object detection, assessing which model performs better in terms of accuracy, speed, and efficiency. This investigation helps in selecting the right model for specific applications, optimizes resource allocation, and guides future enhancements in AI technologies. Such comparative insights are crucial for developers, researchers, and industries aiming to deploy the most effective AI solutions in real-world scenarios.

2.2. The Dataset

For the comparative investigation between YOLOv5 and YOLOv8, the dataset (available from https://universe.roboflow.com/seblful/pills-detection-s9ywn/dataset/19) of 1,659 images is divided into training (87%), validation (8%), and test sets (5%), Fig 2. Images are uniformly resized to 640x640 pixels and simplified by dropping 117 classes. Augmentation techniques include generating three variations per training example through horizontal and vertical flips, cropping (0-20% zoom), rotating (-15° to +15°), and adjusting brightness (-

25% to +25%) and exposure (-8% to +8%). Additionally, 11% of images are converted to grayscale and slightly blurred (up to 1px), enriching the dataset's diversity to challenge and enhance model robustness.

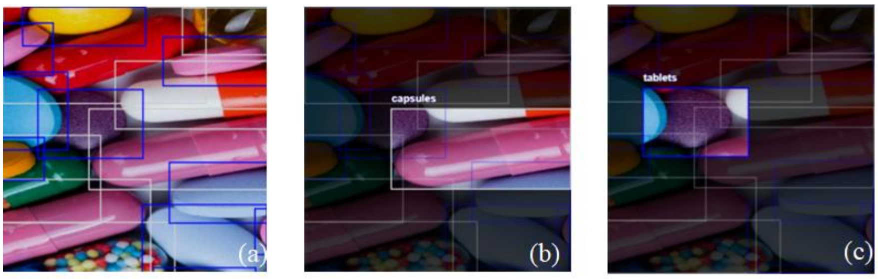

Fig. 2: Object detection using YOLO models: (a) Initial bounding box detections, (b) Refined detection highlighting capsules, (c) Refined detection highlighting tablets.

3. Main Results and General Discussions

Table 1 presents the comparative performance results of two versions of the You Only Look Once (YOLO) object detection model, specifically YOLOv5 and YOLOv8, as applied to a dataset consisting of capsules and tablets. The evaluation metrics used to assess the models' effectiveness include Precision, Recall, F1-Score, mean Average Precision at 50% intersection over union (mAP50), and mAP50-95, which is the mAP calculated over different IoU thresholds from 50% to 95%.

Table 1: Performance comparison of YOLOv5 and YOLOv8 across key metrics.

Metrics	Training		Validation		Testing	
	YOLOv5	YOLOv8	YOLOv5	YOLOv8	YOLOv5	YOLOv8
Precision	0.869	0.885	0.869	0.892	0.821	0.801
Recall	0.866	0.881	0.866	0.873	0.751	0.738
F1-Score	0.867	0.883	0.867	0.882	0.784	0.768
mAP50	0.914	0.935	0.914	0.934	0.795	0.789
mAP50-95	0.536	0.571	0.536	0.571	0.441	0.444

From the table, it is evident that both YOLOv5 and YOLOv8 show robust performance across all metrics in the training and validation phases. In the training phase, YOLOv8 outperforms YOLOv5 in terms of Precision (0.885 vs. 0.869), Recall (0.881 vs. 0.866), F1-Score (0.883 vs. 0.867), and mAP50 (0.935 vs. 0.914), suggesting that YOLOv8 may have a better overall fitting to the dataset. Similarly, in the validation phase, YOLOv8 continues to demonstrate slightly higher scores than YOLOv5 across all metrics, with notable improvements in Precision (0.892 vs. 0.869) and mAP50 (0.934 vs. 0.914).

However, the testing phase results reveal a different scenario where both models experience a drop in performance metrics compared to training and validation, indicating a possible overfitting to the training data. YOLOv5 and YOLOv8 score 0.821 and 0.801 in Precision, respectively, and similarly lower scores in other metrics like Recall, F1-Score, and both mAP50 and mAP50-95. This drop in performance in the real-world testing scenario underscores the challenges in generalizing the models beyond the training data and suggests the need for further tuning or training with a more diverse dataset to enhance their robustness and applicability in practical settings.

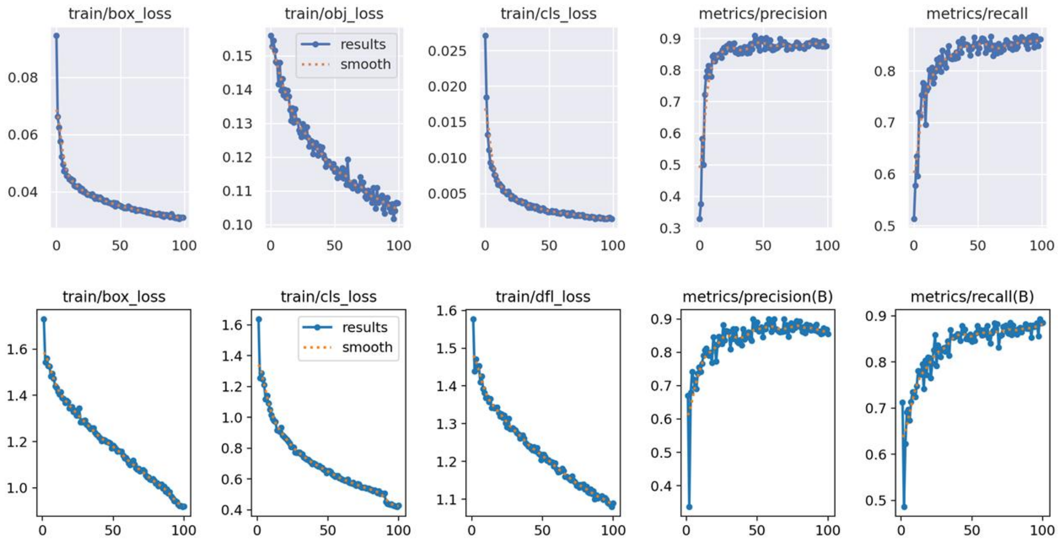

Fig. 3: Training loss and performance metrics of; Top) YOLOv5, and Bottom) YOLOv8 models over 100 epochs.

Fig 3 depicts the training performance of YOLOv5 and YOLOv8 models over 100 epochs, focusing on their ability to detect objects like capsules and tablets with the top row results are from YOLOv5. YOLOv8, shown in the bottom row, consistently shows superior performance compared to YOLOv5 in terms of loss reduction, with smoother and lower curves across all loss categories. This indicates that YOLOv8 has a more efficient learning process, potentially due to a more advanced or better-optimized architecture, leading to more accurate object localization and classification.

In terms of precision and recall, both models exhibit substantial improvement as training progresses, with values nearing 0.9 towards the end of the training cycle. However, YOLOv8 again edges out YOLOv5, achieving slightly higher scores in both metrics. This enhanced performance in precision and recall suggests that YOLOv8 is more effective at minimizing false positives and negatives, making it a more reliable choice for real-world object detection applications where accuracy is critical.

However, While YOLOv8 may demonstrate slight improvements over YOLOv5 in terms of lower loss values and higher precision and recall, the performance of YOLOv5 is still robust. It shows that it can reliably detect and classify objects with a high degree of accuracy, making it a competent model for practical applications where YOLOv8's enhancements might not be necessary or could be outweighed by other factors such as computational efficiency or resource constraints.

4. Conclusion

The comparative analysis of YOLOv5 and YOLOv8 for identifying capsules and tablets demonstrated that YOLOv8 generally excels in precision and learning efficiency during the training and validation phases. However, the observed performance decline in real-world testing for both models suggest an issue of overfitting to the training dataset. This highlights the necessity for further adjustments and the use of a more varied dataset to enhance the models' robustness and practical usability in medication dispensing systems. This research underscores the importance of selecting the appropriate AI model based on specific operational needs and conditions, aiming to optimize patient safety and medication efficacy in healthcare settings. There are several promising ideas for future work: using advanced data enhancement techniques, adding detailed comparative analyses, and analyzing different scenarios for YOLOv5 and YOLOv8 in feature extraction to enhance the article's comprehensiveness.

5. Acknowledgements

This work was supported by (i) Suranaree University of Technology (SUT, http://www.sut.ac.th), (ii) Thailand Science Research Innovation (TSRI, https://www.tsri.or.th), and (iii) National Science, Research and

Innovation Fund (NSRF) (NRIIS number 179269). The grant recipient is S. Kaennakham who would like to express his sincere gratitude for their support.

6. References

[1] D. M. Tenneti, D. A. Kittur, R. Hemavathy, T. Shankar, and R. Kumar. Advanced Object Detection for Capsules and Tablets Identification Through Deep Learning. *7th International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS)*. IEEE. 2023, pp. 1-6.

[2] K. L. James, D. Barlow, R. McArtney, S. Hiom, D. Roberts, and C. Whittlesea. Incidence, type and causes of dispensing errors: a review of the literature. *International journal of pharmacy practice.* 2009, **17**(1): 9-30.

[3] J. Zhou, J. He, G. Li, and Y. Liu. Identifying capsule defect based on an improved convolutional neural network. *Shock and Vibration.* 2020, **2020**.

[4] L. Tan, T. Huangfu, L. Wu, and W. Chen. Comparison of RetinaNet, SSD, and YOLO v3 for real-time pill identification. *BMC medical informatics and decision making.* 2021, **21**:1-11.

[5] S. Srivastava, A. V. Divekar, C. Anilkumar, I. Naik, V. Kulkarni, and V. Pattabiraman. Comparative analysis of deep learning image detection algorithms. *Journal of Big data.* 2021, **8** (1): 66.

[6] A. John and D. Meva. A comparative study of various object detection algorithms and performance analysis. *International Journal of Computer Sciences and Engineering.* 2020, **8** (10): 158-163.

[7] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi. You only look once: Unified, real-time object detection. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 779-788.

[8] J. Terven and D. Cordova-Esparza. A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS. *arXiv preprint arXiv:2304.00501,* 2023.

[9] R. Qureshi, M. G. RAGAB, S. J. ABDULKADER, A. ALQUSHAIB, E. H. SUMIEA, and H. Alhussian. A Comprehensive Systematic Review of YOLO for Medical Object Detection (2018 to 2023). *Authorea Preprints.* 2023.

[10] T. Diwan, G. Anirudh, and J. V. Tembhurne. Object detection using YOLO: Challenges, architectural successors, datasets and applications. *multimedia Tools and Applications.* 2023, **82** (6): 9243-9275.

[11] U. Nepal and H. Eslamiat. Comparing YOLOv3, YOLOv4 and YOLOv5 for autonomous landing spot detection in faulty UAVs. *Sensors.* 2022, **22** (2):464.

[12] F. M. Talaat and H. ZainEldin. An improved fire detection approach based on YOLO-v8 for smart cities. *Neural Computing and Applications.* 2023, **35** (28): 20939-20954.