A Comparative Study of YOLOv5 and YOLOv7 for Indian Sign Language Recognition

Mrs Nausheeda B.S[1], Eric Fernandes[2]

Asst. Professor, AIMIT St. Aloysius College Mangalore[1]

Student, AIMIT St. Aloysius College Mangalore[2]

ABSTRACT

Indian Sign Language (ISL) has seen a lot of development over the years to stand as a significant communication medium for the deaf community to reach out to the general hearing population. Technology has immense potential to further bridge the gap, especially through application of deep learning algorithms towards ISL recognition. Therefore, this study evaluates YOLOv5 and YOLOv7 for object detection, two powerful algorithms in object detection systems, in order to validate their efficacy in recognizing the gestures of ISL.
A dataset of 35 classes and 1748 images was used to train and test the models, and the metrics such as mAP, mean Average Precision were measured to assess their performance. From the results, though YOLOv7 exhibits more accuracy and hence more suitable for high-performance applications, the training time is relatively much higher, and more computing resources are consumed. On the other hand, while YOLOv5 is slightly less accurate, it is fast and efficient, hence being suitable for resource-constrained applications. The comparison helps guide the selection of a specific model to be applied depending on the requirements of an application for ISL recognition.

1. INTRODUCTION

Sign languages are the critical communication instruments which will seamlessly connect the deaf population with the hearing community. Number of dominant sign languages in India is Indian Sign Language ISL, and this is most important one empowering the community of deaf people. Nonetheless, there is a lack of interpreters and restricted availability of materials for study which creates major barriers towards smooth communication.
Recent technological advancements have catalyzed the development of novel strategies to address these issues. Significantly, deep learning methodologies using computer vision have emerged as effective tools for sign language recognition. However, gesture identification in Indian Sign Language faces unique challenges, such as variations in gestures, varying lighting conditions, and complex backgrounds.

The YOLO family of algorithms deeply influenced the development of deep learning and object detection technologies. Although Ultralytics published YOLOv5 in 2020, with very impressive improvements toward object detection and wide acceptance based on excellent documentation and ease of deployment [1], greater performance in terms of both speed and accuracy has been achieved by YOLOv7, published subsequently, at a 5-160 FPS with a greater need for computational power [5].

The following research targets the study and comparison in the identification performance of Indian Sign Language hand gestures of YOLOv5 and YOLOv7. Using these models through the training and the evaluation on the data set, this work investigates their performances with some evaluation metrics like mAP. Findings extracted from this analysis are likely to have beneficial applications towards enhancing real-time translation systems and assistive technology devices meant to fill numerous deaf persons' diverse demands.

2. LITERATURE REVIEW

New algorithms recently developed in object detection, including You Only Look Once (YOLO), can be applied to the detection of sign languages too. For example, translating gestures in real time for Indian Sign Language (ISL) using YOLOv5 and YOLOv7 appears pretty fast and accurate.

With efficiency and robust performance, the YOLOv5-based PyTorch application. It is based on the CSPDarknet backbone and enhanced data augmentation such as Mosaic, which improves the speed and accuracy. [6] YOLOv7 based on Darknet has E-ELAN and compound scaling which improves the accuracy and speed with minimal increase in computational costs. [7] While YOLOv5 and YOLOv7 do well in general object detection, their performance for ISL detection requires further study.

Dima and Ahmed (2021) studied the recognition of American Sign Language using YOLOv5 and reported high precision, recall, and mAP scores.[3] This might indicate that the YOLOv5 has a lot of potential but, on the other hand, the ISL uniqueness might create difficulties. Anu Rose Joy et al. (2023) examined YOLOv5 for ISL recognition, highlighting its potential for real-time text and audio translation.[11] They stress the necessity of strong ISL datasets for effective model training and evaluation.

Comparing YOLOv5 and YOLOv7 for object detection is worthwhile but not applicable to ISL. Olorunshola et al. (2023) found that YOLOv5 outperformed YOLOv7 in precision, mAP@0.5, and mAP@0.5:0.95, while YOLOv7 exhibited higher recall.[3] However, Yusof et al. (2024) reported YOLOv7 achieving the highest performance in road defect detection with an mAP@0.5 score of 79.0%.[9] These studies highlight the task-dependent nature of model performance, emphasizing the need for careful evaluation within the specific context of ISL recognition.

3 METHODOLOGY

3.1 Dataset Description

This is the data set of 35 classes and 1748 images. Where the train dataset constitutes 1212 images as train dataset, 341 ad valid dataset and 195 as test dataset without any augmentation performed and in preprocessing a stretch was done at 640x640, with auto orientation.

3.2 Model and Architectures

YOLOv5 is an object detection model proposed by Ultralytics in 2020 for real-time efficiency. The backbone used here is CSPDarknet for feature extraction and PANet for multi-scale feature aggregation. It is an anchor-based detection system with relatively lower computational requirements. This makes it run really fast on CPU and small GPUs, which makes it implement very quickly in the resource-limited setting and therefore enjoy great community support along with rich documentation.

Released in 2022, YOLOv7 is an improvement over object detection in terms of accuracy and real-time performance. It uses the E-ELAN, which is the extended efficient layer aggregation network for better gradient flow as well as RepConv or re-parameterized convolution layers for efficient training and inference. The YOLOv7 can provide a more excellent mAP compared to other variants in the YOLO series while attaining speed and accuracy balances at 5 and 160 FPS frame rates; therefore, it suits every application, such as the ones for autonomous driving, defect detection, and even medical imaging.

3.3 Experimental setup

This set of experiments was conducted in Google Colab, which provides free access to NVIDIA Tesla T4 16GB GPUs and 12GB RAM. The YOLOv5 and YOLOv7 models are cloned from GitHub. In addition, PyTorch, OpenCV, and matplotlib have been installed. These conditions permitted reliable and efficient training and evaluation.

The YOLOv5 and YOLOv7 used the same dataset that has divided into training, validation, and testing sets. It has trained YOLOv5 for 50 epochs. The YOLOv7 was trained for 70 epochs with a batch size of 16. Training these models started at an initial learning rate of 0.01 and using cosine scheduler during the course of training. Both models employed SGD optimizer with 0.937 momentum and weight decay of 0.0005 to warrant a fair comparison between the models.

3.4 Evaluation Metrics

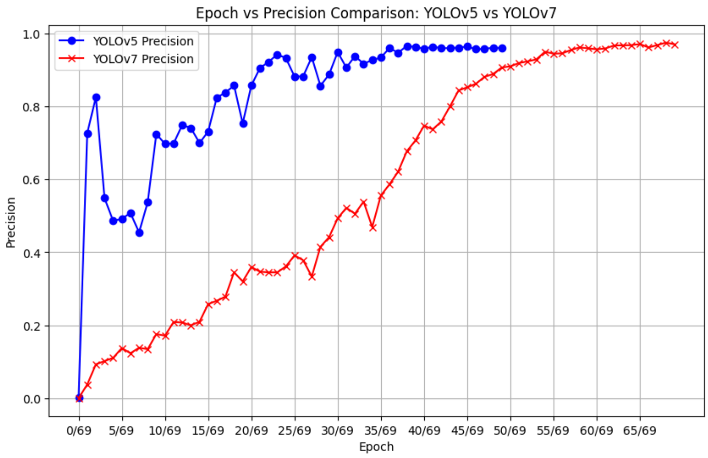

Figure 1. *comparison of precision between yolov5 and yolov7*

The training behavior of precision graph differs between YOLOv5 and YOLOv7. While YOLOv5 starts from 0 precision, peaks at 0.8 by the 3rd epoch and drops again to 0.4 and increases gradually to 0.9, YOLOv7 rises steadily from the start. Both achieved 0.9 precision at epoch 50; YOLOv5 presents early oscillations and, in contrast, YOLOv7 is steady throughout the training.

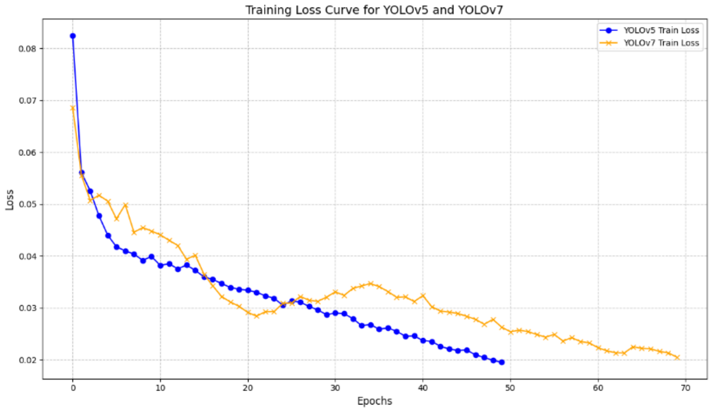

Figure 2. *comparison of Loss between yolov5 and yolov7*

The graph plots the training loss of Yolov5 and Yolov7 for 70 epochs, and both exhibit a declining trend. V5 starts at 0.08 and falls steeply until epoch 69 to settle at 0.02 after having spiked upward at around epoch 20. V7 starts at 0.07 and declines to 0.03 at epoch 30 before slightly oscillating up and down and settling at 0.02 by epoch 69.

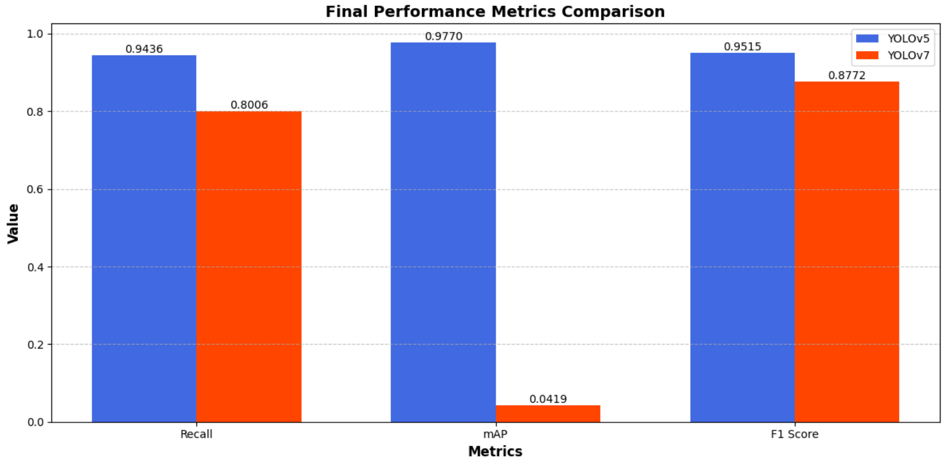

Figure 3. *final perfomace Metrics between the two models*

The image presents a comparative analysis of the final performance metrics between YOLOv5 and YOLOv7 object detection models. The metrics evaluated include Recall, mAP (mean Average Precision), and F1 Score. The graph clearly shows that YOLOv5 outperforms YOLOv7 in all three metrics, with YOLOv5 achieving higher values for Recall (0.9436 vs. 0.8006), mAP (0.9770 vs. 0.0419), and F1 Score (0.9515 vs. 0.8772).

Experiment Process:

Figure 4. test data

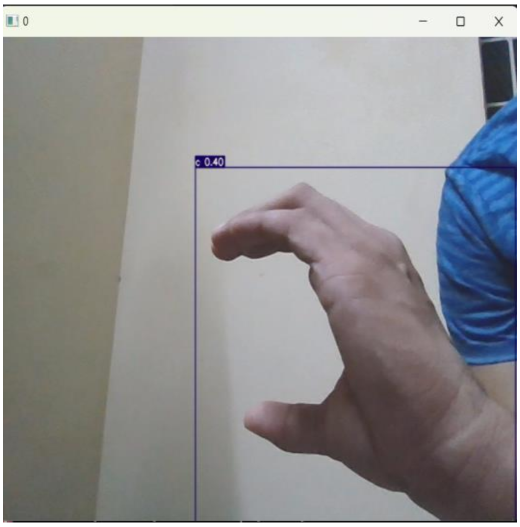

Figure 5. *live camera Feed of yolov7*

The experiment was done on the Indian Sign Language dataset with the use of static test images and a webcam input to gauge the performance of our trained models in YOLOv5 and YOLOv7. We used static test images chosen for accuracy to test the classification and identification of hand signs so that the test setting was controlled and repeatable. In addition, real-time inference was performed using a webcam feed to test the strength of the model in real-life scenarios with changes in illumination, background, and position of hands. The above-mentioned approach provided comprehensive evaluation for the model in laboratory as well as in natural conditions and thus holds great potential for implementation in real-world applications in the area of sign language recognition.

4. CONCLUSION

In conclusion, this study explored the performance of YOLOv5 and YOLOv7 models for Indian Sign Language recognition using a custom dataset. The experiments revealed that YOLOv5, with its faster training times, is well-suited for smaller datasets and rapid prototyping. However, its performance heavily relies on dataset quality and quantity, requiring extensive data augmentation to achieve accurate results. On the other hand, YOLOv7 demonstrated superior performance and robustness in detecting and classifying signs, even with minimal optimization. While YOLOv7 is resource-intensive and requires significantly longer training times, its ability to learn effectively from larger datasets makes it ideal for real-world applications where computational resources are available. These findings highlight a trade-off between speed and precision, offering valuable insights into model selection based on dataset size and resource constraints in sign language recognition tasks.

5. REFERENCES

[1] Jocher, G., et al. (2020). *YOLOv5*. GitHub. https://github.com/ultralytics/yolov5

[2] WongKinYiu. (2022). *YOLOv7: The Most Accurate Real-Time Object Detector*. GitHub. https://github.com/WongKinYiu/yolov7

[3] Olorunshola, O. E., Irhebhude, M. E., & Evwiekpaefe, A. E. (2023). A comparative study of YOLOv5 and YOLOv7 object detection algorithms. *Electrical and Electronics Engineering Department, Air Force Institute of Technology*. Kaduna: Nigerian Defence Academy.

[4] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). *You Only Look Once: Unified, Real-Time Object Detection*. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). https://doi.org/10.1109/CVPR.2016.91

[5] Wang, C. Y., Bochkovskiy, A., & Liao, H. Y. M. (2022). *YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors*. arXiv. https://arxiv.org/abs/2207.02696

[6] Hussain, M. (2024). *YOLOv5, YOLOv8, and YOLOv10: The go-to detectors for real-time vision*. arXiv:2407.02988v1 [cs.CV]. Retrieved from https://arxiv.org/abs/2407.02988

[7] Olorunshola, O. E., Irhebhude, M. E., & Evwiekpaefe, A. E. (2023). A comparative study of YOLOv5 and YOLOv7 object detection algorithms. *Journal of Computing and Social Informatics, 2*(1).

[8] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (CVPR), 779–788. https://doi.org/10.1109/CVPR.2016.91

[9] Yusof, N. I. M., Sophian, A., Zaki, H. F. M., Bawono, A. A., Embong, A. H., & A Comparative Study Ashraf, A. (2024). Assessing the performance of YOLOv5, YOLOv6, and YOLOv7 in road defect detection and classification: A comparative study. *Bulletin of Electrical Engineering and Informatics, 13*(1), 350–360. https://doi.org/10.11591/eei.v13i1.6317

[10] Liaqat Ali, M., & Zhang, Z. (2024). The YOLO framework: A comprehensive review of evolution, applications, and benchmarks in object detection. Preprint. https://doi.org/10.20944/preprints202410.1785.v1

[11] Joy, A. R., Titus, A., Fathima, P. S., Shibu, A. M., & Azeez, A. (2024). Indian sign language recognition using YOLOv5. *Amal Jyothi College of Engineering, Department of Computer Science & Engineering*.

[12] Roboflow. (2024). *Indian Sign Language Detection Dataset (Version 2)* [Data set]. Niladri Basu Roy Workspace. https://universe.roboflow.com/niladri-basu-roy-qnrm4/indian-sign-language-detection/dataset/2