

Exploring Advanced Transfer Learning Models for Effective Irish Sign Language Recognition

Vaishaali Kondapalli(23200337), Minu Jose(23200724)

School of Mathematics and Statistics, University College Dublin



Abstract

This project focuses on the detection of Irish Sign Language (ISL) gestures using advanced transfer learning-based object detection models. The study utilized a custom dataset, three transfer learning models (YOLOv8, YOLO-NAS, and RT-DETR), and a standard sequential CNN for comparative analysis. Key aspects of this evaluation include data augmentation, hyperparameter tuning, and performance metrics like **mAP50**. Among these, YOLO-NAS demonstrated superior accuracy and consistency, marking a significant step forward in ISL recognition technology.

Methodology

- Data Collection and Preparation:** The dataset with seven phrases—Hello, Help, Please, Thank You, Water, Sorry, and Phone Number was constructed using **OpenCV** and manually annotated, using **LabelImg** program.
- A total of 445 images: 282 photos for training, 138 images for validation, and 25 images for testing. were collected.
- Data augmentation** approaches were used during training. These included random rotations, flips, scaling, and colour changes.



Introduction

- Sign languages are crucial for communication within the deaf and hard-of-hearing communities, with Irish Sign Language (ISL) playing a key role in cultural identity and social inclusion. This project investigates the application of transfer learning-based models—YOLO-NAS, YOLOv8, and RT-DETR—for detecting ISL gestures.
- By comparing these models to a standard sequential CNN and utilizing a custom dataset with image augmentation, the study aims to improve ISL recognition and contribute to its technological development.

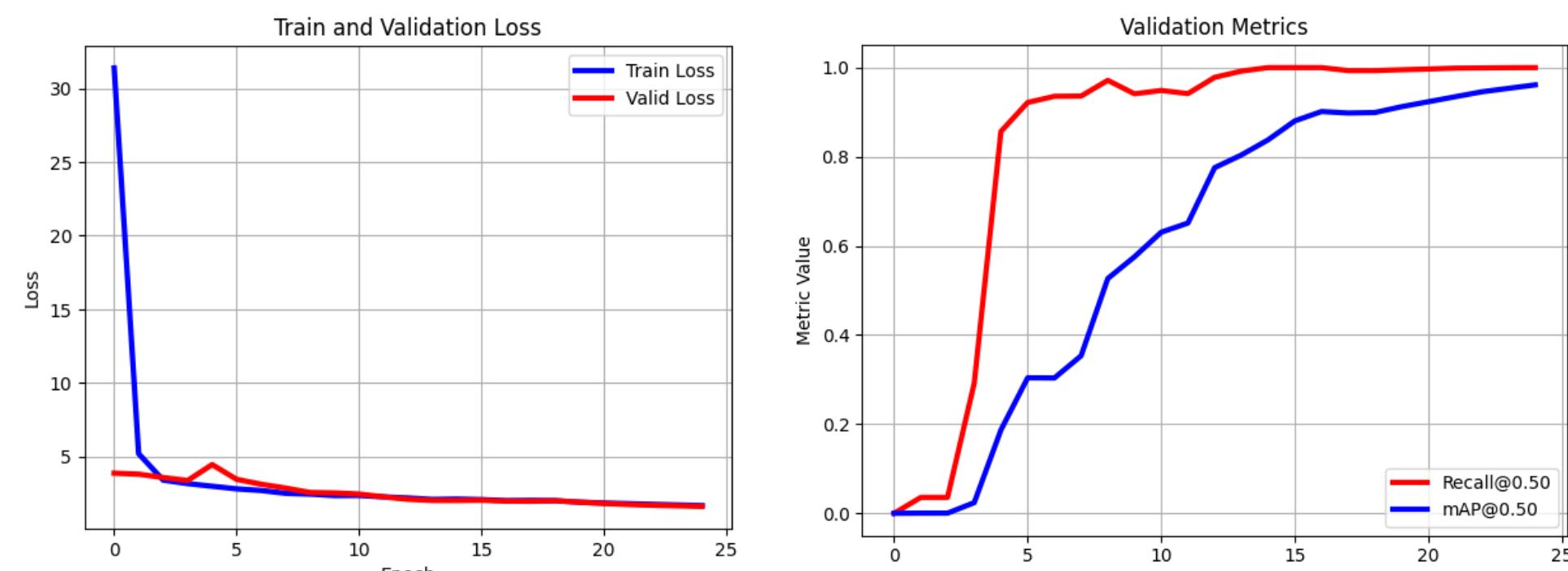
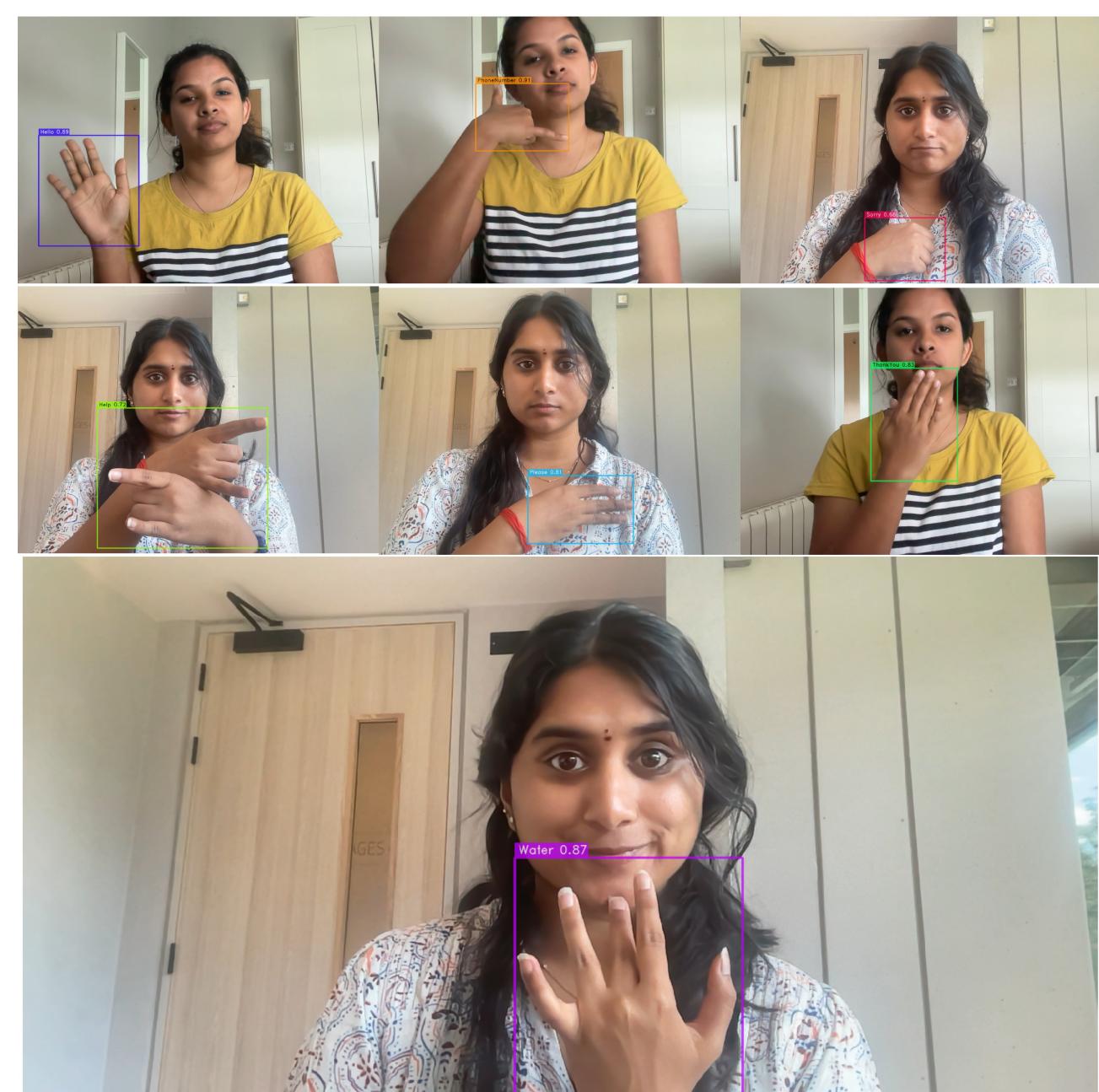
Model Selection

- YOLO-NAS (S sized):** The model is employed due to its high accuracy and remarkable inference speed. Its quantization-aware architecture ensures efficient performance even in resource-constrained environments.
- RT-DETR:** Since YOLO models' accuracy can be hindered by the use of **Non-Maximum Suppression (NMS)**, RT-DETR prompts to investigate whether DETRs can surpass the advanced YOLO detectors by eliminating the delay introduced by NMS.
- YOLOv8:** Its design incorporates the latest innovations in object detection, making it a strong candidate for real-time applications, which is why we chose to experiment with it alongside other state-of-the-art models.
- CNN:** Also experimented with a standard sequential CNN to evaluate its performance comparatively.

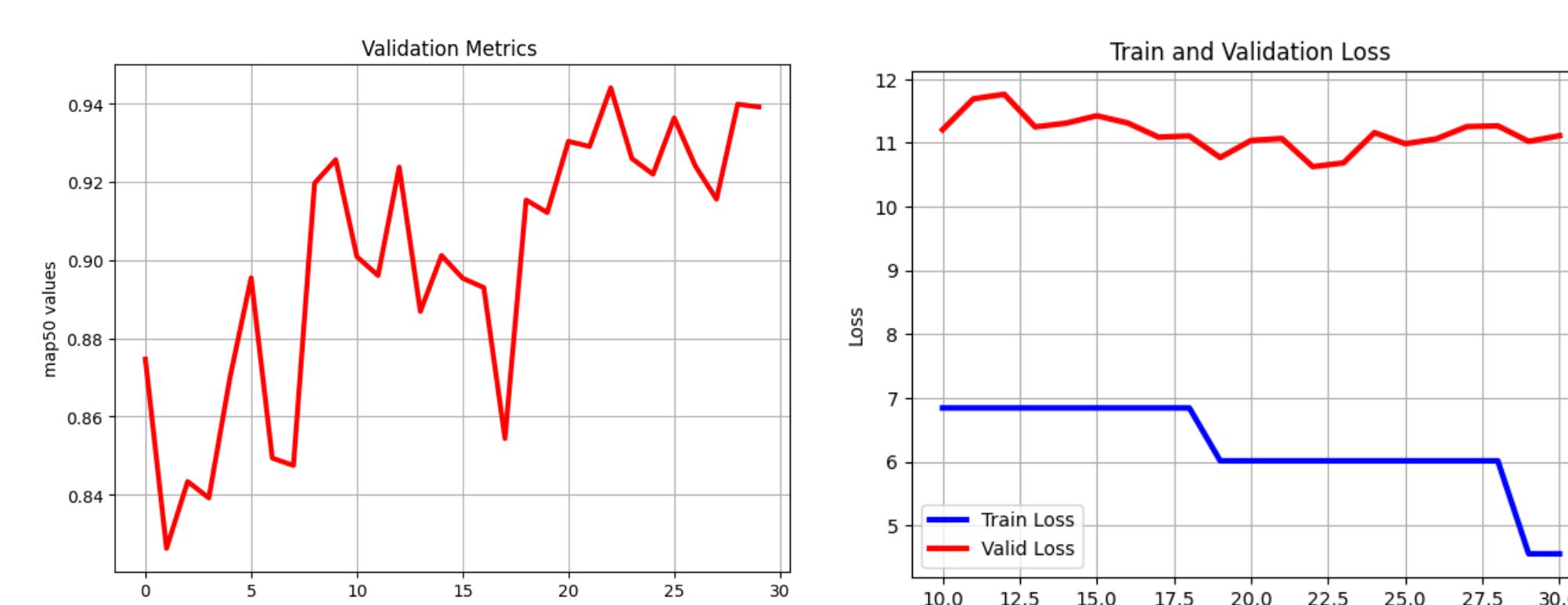
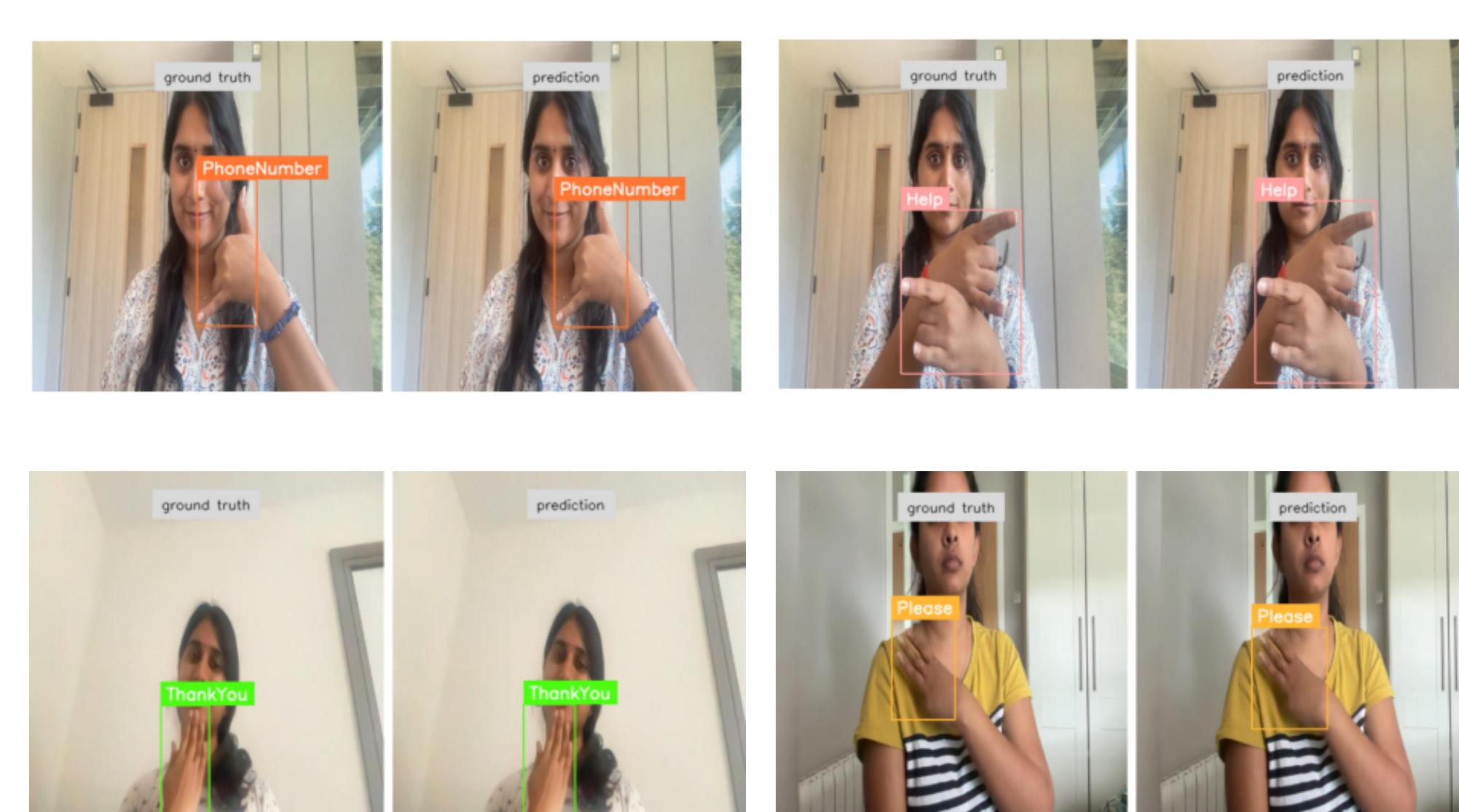
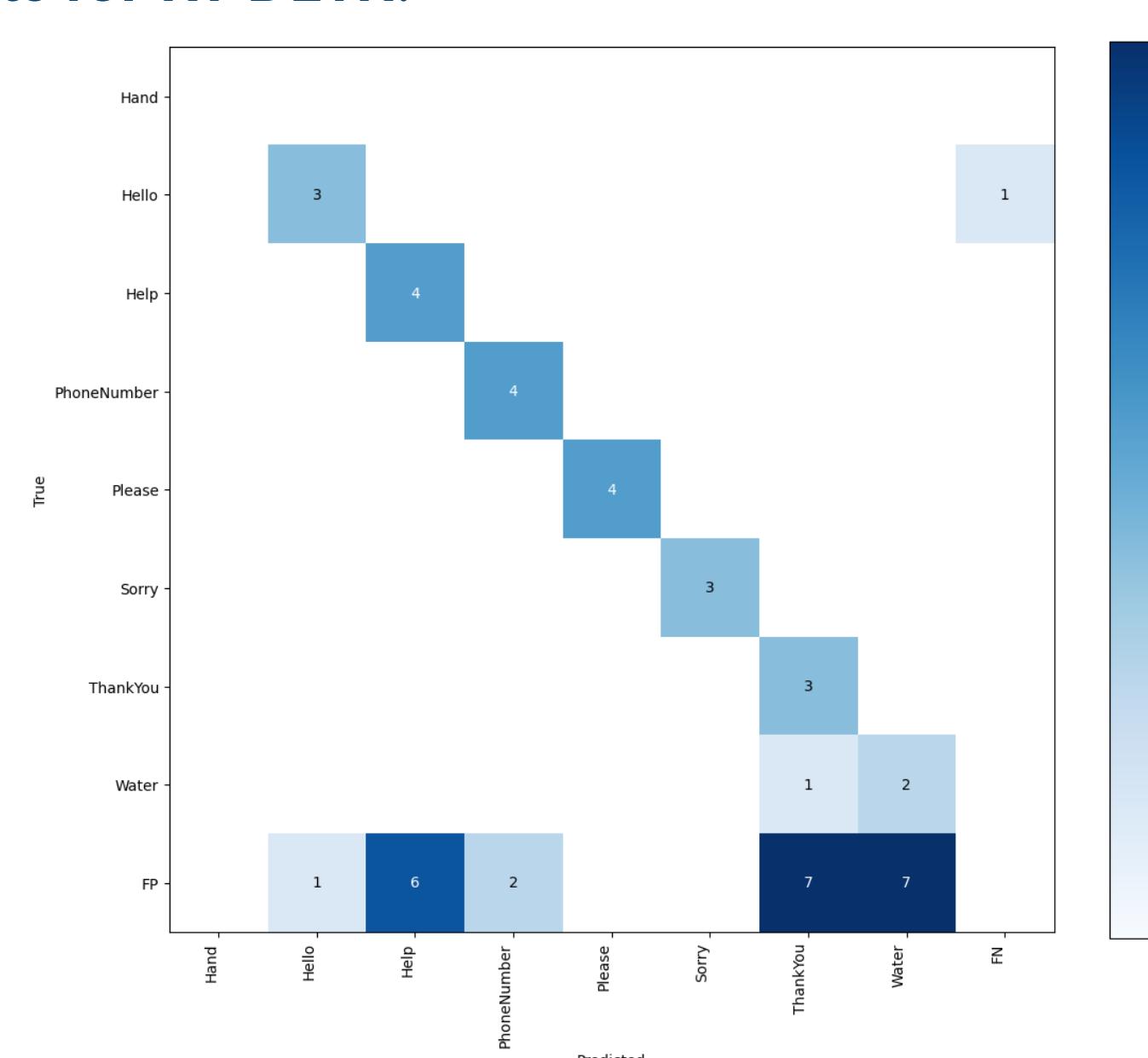
Results and Analysis

Results for YOLO-NAS S:

- High confidence scores across various gestures: high recall rate of 1.0 and mAP50 of 0.997. The minimal gap between training and validation loss portrays an ideal case scenario result.

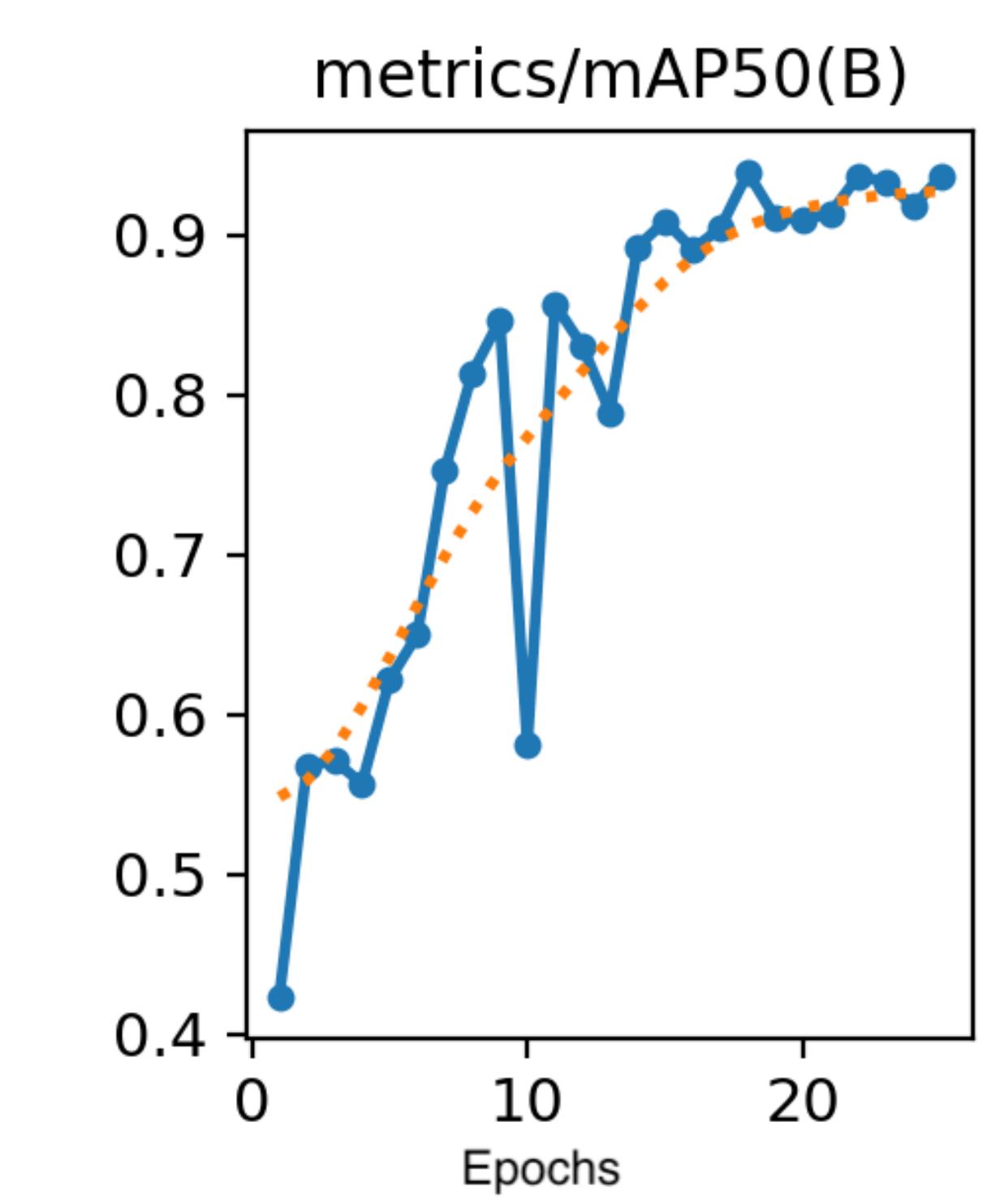
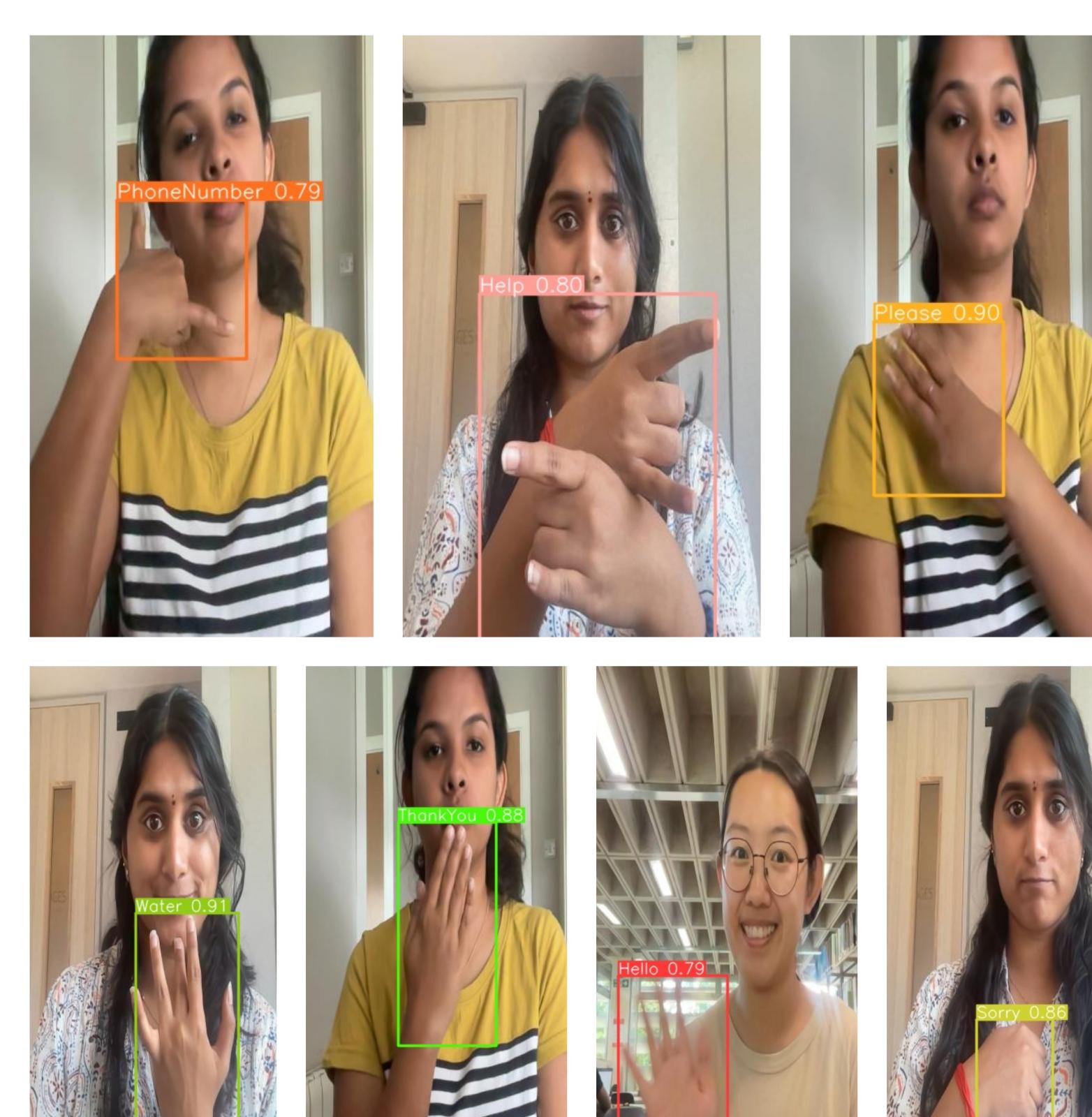


Results for RT-DETR:



- The mAP50 value for the RT-DETR model is 0.94. The confusion matrix points out that for "Water" and "ThankYou" classes the false positives are quite high compared to other classes.

Results for YOLOv8:



- Overall, YOLO-NAS outperforms the other models judging the consistent increasing mAP50 graph, in contrast to the more unstable performance of RT-DETR and YOLOv8.
- The sequential CNN model frequently misclassified test images with all five fingers as "Hello," indicating a need for further experimentation and fine-tuning to improve model performance.

Future Works

- Include more gestures and diverse examples, hence improving the model's generalisation and accuracy.
- Progress from sign language detection to a full packed sign language interpreter.
- Deployment strategies should be taken into account to make the system usable in real-world applications.

References

- Yian Zhao et al., "DETRs Beat YOLOs on Real-time Object Detection," arXiv preprint arXiv:2304.08049, 2023.
Anusha Puchakayala et al., "American Sign Language Recognition using Deep Learning," Proceedings of the 7th International Conference on Computing Methodologies and Communication (ICCMC-2023), IEEE Xplore, 2023, ISBN: 978-1-6654-6408-6, DOI: 10.1109/ICCMC56507.2023.1004049.
K. Abhinand et al., "Malayalam Sign Language Identification using Finetuned YOLOv8 and Computer Vision Techniques," arXiv:arXiv:2405.05702v1 [cs.CL], 08 May 2024. License: CC BY 4.0. A. Mishra et al., "ISL recognition of emergency words using mediapipe, CNN and LSTM," IEEE, 2023, pp. 322-325.