NTU Journal of Engineering and Technology (ICSDT2025) special issue
DOI: https://doi.org/10.56286/ntujet.icsdt2025

IRAQI Academic Scientific Journals

P-ISSN: 2788-9971 E-ISSN: 2788-998X

NTU Journal of Engineering and Technology

Available online at: https://journals.ntu.edu.iq/index.php/NTU-JET/index

Deep Learning YOLO Models in Hand Gesture Recognition

Amina TH. Mohealdeen [1], Emad A. Mohammed [1]

[1] Department Computer Engineering Technology, Technical Engineering University, Northern Technical University, Mosul, Iraq
aminathaker512@ntu.edu.iq , e.a.mohammed@ntu.edu.iq

Article Informations

Received: 15-03- 2025,
Accepted: 28-04-2025,
Published online: 01-03-2026

Corresponding author:
Name: Amina TH. Mohealdeen
Affiliation : Northern Technical University
Email:aminathaker512@ntu.e

Key Words:
YOLO models,
Hand gesture,
YOLO11,
YOLO10,
AI.

ABSTRACT

Hand gesture detection is essential for improving human-computer interaction, considerably advancing the creation of more intuitive and effective interfaces. This study examines the efficacy of three sophisticated object identification models in identifying hand motions: YOLOv8, YOLOv10, and YOLOv11. A dataset was created for this purpose, enhanced via Roboflow, and utilized for training and evaluating the performance of these models based on metrics including mAP, Precision, Recall, and F1-score. Performance was enhanced by adjusting the optimizers (Adam, SGD, and AdamW) and hyperparameters, including learning rate and epochs. The findings indicate outstanding performance, with YOLOv8 and YOLOv10 attaining mean Average Precisions of 99.2% and 99.1%, respectively, alongside precisions of 98.1% and 98.6%, recalls of 97.7% and 97.2%, YOLOv11 had a mean Average Precision of 99.2%, a precision of 98.6%, and a recall of 98.4%, These findings demonstrate that YOLOv11 adeptly manages more complex datasets while preserving an ideal equilibrium between speed and precision.

1. Introduction

Human-computer interaction has become an essential aspect of modern technol-ogy, driven by rapid advances in digitalization.[1] Gestures play a key role in this development, providing an intuitive means of controlling a device. Gesture recogni-tion, especially in applications such as sign language, can enhance communication for people with hearing disabilities [2].

This technology uses hand gestures to convey specific meanings or commands, based on machine learning techniques and advanced feature extraction methods common in computer vision tasks, such as image classification.[3] Despite the pro-gress, challenges remain, particularly in accurately detecting hand positions or movements, as variations in hand shapes and gestures across images complicate localization. Although artificial intelligence and machine learning have made signifi-cant contributions in many fields, there is an urgent need for further development in gesture detection. This need has prompted researchers to explore advanced methods and intensive learning models based on object detection and image recognition tech-niques [4], [5].

YOLO (You Only Look Once) models have shown exceptional performance on detection and classification tasks, including the recently released YOLOv8, YOLOv10, and YOLOv11. This study evaluates the effectiveness of these models in recognizing hand gestures representing numbers zero through nine in American Sign Language. The models were evaluated on a dataset of 6,000 images for training and testing, based on accuracy metrics such as recall and precision. This study discusses the potential application of these models in robotics simulation for gesture recogni-tion and automation, with the expectation that these developments will contribute to improvements in gesture control, robotics, assistive technologies, and human-computer interaction systems.

2. Relevant Literature

The YOLO (You Only Look Once) algorithm plays a crucial role for hand gesture detection through different data types such as films, photos, and wearable devices according to recent research findings. Such algorithms demonstrate their capability for improving both accuracy levels and response times which leads to successful task execution in dynamic environments. The YOLOv2 and YOLOv3 models using DarkNet-19 and DarkNet-53 achieved detection accuracies of 99.10% and 99.18% according to Rubin Bose and V. Sathiesh Kumar. The prediction times reached 20 milliseconds for YOLOv3 and 25 milliseconds for YOLOv3 while executing sign language communication tasks [6] because of their efficient processing capabilities.

The research by Hongchao Zhuang et al. showed YOLOv4 successfully operated a music box while reaching 97.8% precision along with improved speed. The practical usefulness of YOLO emerges from using it to recognize gestures in interactive systems [7].

Soukaina Chraa Mesbahi et al. improved YOLOv7 for hearing-impaired communication with 98.20% accuracy and F1 scores exceeding 98.4% using 50,000 images. Research revealed that this advancement demonstrated how the model excels at detecting gestures in real time and enabling sign language communication [8].

Renxiang Chen and Xia Tian modified YOLOv5 through the integration of CBAM and an effective layer aggregation network for detecting hand signals across various backgrounds. The researchers obtained mAP0.5:0.95 scores that measured 75.6% from the EgoHands dataset and 66.8% from TinyHGR at frame rates exceeding 64 frames per second. Relevant research demonstrated that YOLOv5 could achieve optimal performance for hand gesture recognition purposes in complex environmental conditions [9].

YOLOv8 model proved its efficiency for processing datasets of lower volume. The YOLOv8 demonstrated excellent performance reaching 97.21% accuracy while detecting 51 signs within the Malayalam Sign Language dataset after using preprocessing methods [10].

Vidhyasagar et al. made a real-time recognition system for American Sign Language (ASL) through their utilization of YOLO. The reported system reached 95.7% average accuracy which indicates the significance of using diverse datasets coupled with suitable preprocessing methods for optimal gesture detection performance [11].

Mujahid et al., develops a streamlined hand gesture recognition model using YOLOv3 and DarkNet-53 deep learning models which ensures real-time operation. The training process of the model utilized hand gesture data from a custom-made dataset which led to remarkable performance metrics in all measurement categories including accuracy and precision and recall and F-1 score (97.68%, 94.88%, 98.66%, 96.70% respectively). The accuracy rates of YOLOv3 surpassed VGG16 and SSD models since they reached between 82% and 85% accuracy. This model proved reliable in low-resolution conditions thus making it appropriate for real-time systems that involve human-computer interaction and communication between impaired people. The proposed system displayed efficient video-based gesture recognition capabilities that make it suitable for real-time GPU-based implementation [12].

Moysiadis et al., focuses on improving human-robot interaction (HRI) in agriculture through hand gesture recognition. Six machine learning classifiers were evaluated for real-time detection of five hand gestures by researchers. A depth camera recorded gestures that ROS translated into commands which operated a UGV. The best results emerged from K-nearest neighbor (KNN) in combination with Linear Discriminant Analysis (LDA) and Support Vector Machine (SVM) among the analyzed classifiers. The research proved that the framework operated effectively in actual field settings allowing the UGV to execute tasks through hand gestures that directed it to follow people and execute stationary moves or location changes. The analysis demonstrates the benefits of combining gesture-based communication systems for enhancing collaborative robotics in agriculture [13].

Herbaz et al. created real-time hand gesture recognition systems through the implementation of YOLO (You Only Look Once) deep learning model. Project work explanation includes details about how data processing matters throughout model training and testing phases according to the researchers. The researchers evaluated YOLO versions YOLOv5, YOLOv6, and YOLOv8 using more than 5000 images obtained from Roboflow. The hand gesture recognition operations served best through YOLOv8 because this version delivered accurate and fast performance. The work advances assistive technology development because it enables real-time gesture recognition which enhances human-computer interaction. YOLO proves its versatility to support immediate gesture recognition requirements in specific environments because of its adaptable design [14].

The research work of Soukaina et al. demonstrates how YOLO (You Only Look Once) deep learning models detect hand gestures performed in sign language. The research concluded that YOLOv5 emerged as the top model because it achieved 99.50% mean Average Precision (mAP) for hand gesture detection with superior speed and accuracy and precision than YOLOv3, YOLOv4, YOLOv4-tiny. YOLOv5 shows better performance in real-time hand gesture identification under changing lighting and various background scenarios thus making it the most suitable model selection for assisting hard of hearing individuals through sign language communication. The research provides crucial information to improve human-computer interaction system procedures [15].

3. Proposed Methodology

The implementation process for deploying object detection models with YOLOv8, YOLOv10 and YOLOv11 to a custom dataset requires multiple important operational steps shown in Figure (1). The first requirement for YOLO models involves collecting and preparing images with their related labels for proper functioning. The dataset annotation process becomes easier through Roboflow which helps users create proper training structure. Data augmentation through image rotation and flipping and scaling techniques are implemented on prepared datasets to improve their diversity which ensures good generalization from the model. The model starts by selecting YOLOv8, YOLOv10 or YOLOv11 depending on the task specifications.

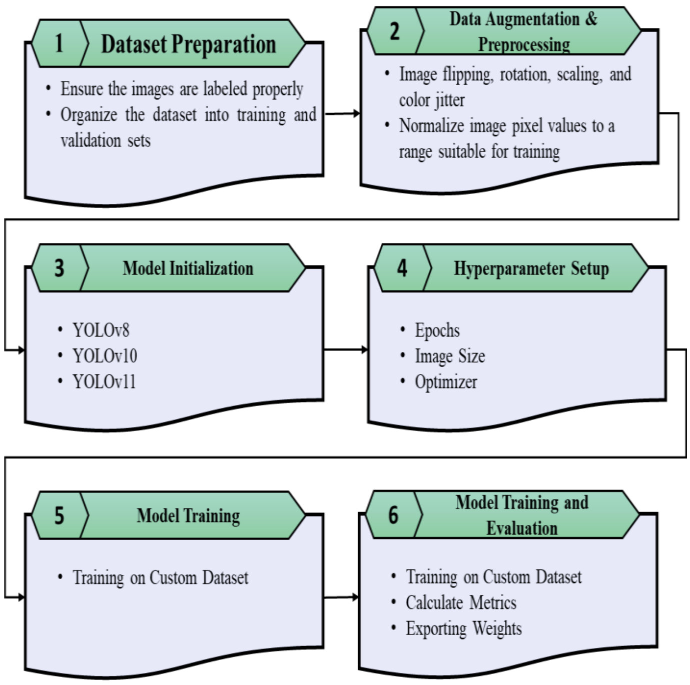

Fig. 1. The diagram of the proposed hand gesture recognition system

The integration of pretrained weights enables quick adaptation between the custom dataset and these models following transfer learning principles. The model performance becomes optimized by applying specific hyperparameter values which include epochs number along with image size and optimizer type and learning rate determination. The model receives training from the designated custom dataset that teaches it to identify and detect items from labeled photographs. The training process allows the model to adjust its weight values as a means to reduce loss and enhance its accuracy levels. The trained model completes evaluation using validation data that generates precision, recall and mAP (mean Average Precision) performance metrics to measure its object detection accuracy.

The deployed weights function as the best-performing versions which get saved before they get exported as either PyTorch or ONNX formats. A test of new data images using inference determines model performance before deployment occurs either through Roboflow cloud services or local hardware. The deployment process requires developers to develop either a REST API for remote access or to implement the NVIDIA Jetson devices for edge computing applications. YOLO models become effective for object detection applications through a structured development approach that enables successful deployment to various scenarios with high performance levels.

3.1 Dataset collection

The hand gesture identification dataset, depicting numeric values from 0 to 9, was sourced from an open-access resource. It amalgamates two RGB picture datasets to augment diversity and quality. The initial dataset [16] comprised over 5000 photos and was refined and condensed to roughly 2000 high-quality photographs. An extra 1000 photos from a different dataset [17] were integrated to enhance diversity in visual and environmental situations. The dataset was annotated in YOLO format utilizing the Roboflow platform, which enables rapid training and classification by storing annotations in text files that commence with the class identification. The resulting dataset comprises 10 classes, each denoting a numeral from 0 to 9. Fig.2 depicts the hand movements present in the dataset.

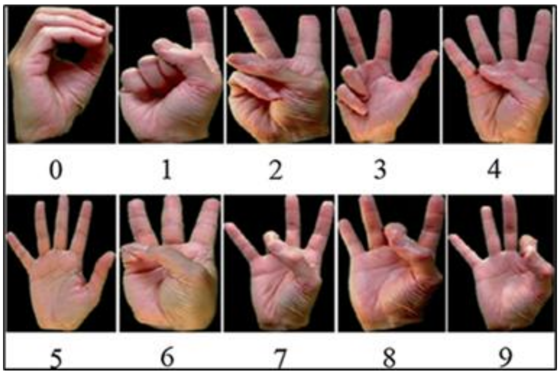

Fig. 2. Hand gesture set to be detected.

3.2 Data processing

Data Preprocessing is crucial for improving model performance. The initial collection of 3,000 photos was expanded to 6,000 using augmentation techniques. Images were resized to 640×640 pixels, horizontally flipped to generalize hand gestures, then randomly rotated to accommodate various orientations. Adjustments to brightness and contrast, noise, enhanced resilience to variations in illumination, and noisy data. The final dataset was divided into 70% for training, 20% for validation, and 10% for testing, providing adequate data for model enhancement and generalization.

3.3 The proposed algorithms

You Only Look Once (YOLO) is a powerful object detection system capable of identifying multiple objects within a single image or frame. We divide the image into a grid, where each cell identifies objects, predicts the coordinates of the selected square, and calculates class probabilities. We apply the YOLO algorithm to YOLO v8, YOLOv10, and YOLOv11 models after manual annotation and preprocessing. YOLO v8: Optimized for speed and accuracy, it is excellent at finding small objects with great accuracy, making it suitable for real-time applications that require fast and accurate data processing [18].

YOLO v10: Enhances detection accuracy and computational efficiency and skillfully manage challenging conditions such as low light, occlusions, and crowded surroundings, thus increasing its adaptability to practical applications.[19] YOLO v11 enhances object detection by improving flexibility and generalization, achieving high accuracy while reducing computational resource requirements, and thus making it suitable for large-scale applications [20].

3.4 System training

YOLO models necessitate critical factors that influence accuracy, speed, and memory consumption the paramount parameters encompass:

- Larger input sizes enhance accuracy but necessitate increased processing power and memory.
- Training Parameters: The learning rate, weight decay, and optimizer govern weight adjustments and affect the model's convergence efficiency. An elevated learning rate accelerates training, while weight decay mitigates overfitting.
- Early Stopping: Mitigates overfitting by ceasing training when performance plateaus, optimizing the utilization of computational resources.

4. Experiments and Results

4.1 Evaluation metrics

This study evaluates the effectiveness of YOLO models in hand gesture recognition using different critical metrics:

Average Precision (AP). The area under the precision-recall curve, which encapsulates the effectiveness of the model across varying confidence thresholds.

$$AP = \int_0^1 Precision\,(r)dr \qquad (1)$$

Where Precision(r) is the precision at recall level r. This is the area under the precision-recall curve. It summarizes the model's ability to maintain high precision across different recall levels, providing an aggregate measure of performance.

Precision. The proportion of accurate positive predictions, indicating the model's ability to minimize false positives. Recall evaluates the model's ability to identify all relevant events.

$$Precision = \frac{TP}{TP+FP} \qquad (2)$$

Where: TP = True Positives (correctly identified positive samples), FP = False Positives (incorrectly identified as positive)

F1 Score. A balanced statistic that integrates precision and recall, useful when these values are disproportionate.

$$F1-Score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \qquad (3)$$

4.2 Detection result of the YOLO model

The models efficiently classify all classes of hand gestures, with boundary boxes accurately defining the essential hand regions for accurate gesture recognition. The models show strong performance on zoomed-in and close-up hand images, ensuring accurate recognition and appropriate class identifier assignment. We evaluated multiple iterations of YOLO to determine the most effective model for hand gesture recognition. Training used 640 × 640-pixel images, with learning rates of 0.001 and 0.01. We used three optimization algorithms: SGD (stochastic gradient descent), Adam (which combines SGD with RMSProp to improve stability), and AdamW (which combines weight decomposition to increase model generalization).

YOLOv8: The YOLOv8 model training showed excellent performance across different configurations, where changing optimizers, learning rates, epochs, and training periods significantly affected its accuracy and efficiency, Table 1 shows the training values of YOLOv8, the most important training results can be summarized as follows:

- Adam optimizer achieved the highest mAP (99.2%) with balanced precision, recall, and F1 scores (97.9%), providing efficient training times.
- Training for 50 epochs proved to be efficient for the dataset (6000 images), while early stopping (26-43 epochs) prevented overfitting.
- A learning rate of 0.001 consistently provided the best mAP (99.0%-99.2%), while a learning rate of 0.01 boosted recall (97.8%) but slightly reduced mAP.

Adam was identified as the most effective optimizer for YOLOv8, successfully reconciling accuracy and training efficiency. The confusion matrix Figure 3 indicates good classifications of the data as evidenced by the high values along the diagonal (e.g., 110, 176, 117), indicating accurate classifications of images. Off-diagonal misclassifications, such as those between classes 2 and 8, were rare and thus had little impact on the overall performance of the model. Figure 4 shows the Training and validation metrics plot shows a converging trend with decreasing validation losses, while mAP50 and mAP50-95 values remain stable, indicating the robustness of the model and its ability to generalize to detect and classify objects on diverse datasets.

Table 1. The performance of YOLOV8 training values

Model	Optimizer	Epochs	mAP (%)	Precision (Pr %)	Recall (Rc%)	F1-Score (%)
	SGD		98.9	97.1	98	97.3
	Adam	50	99.2	98	98.4	97.9
	AdamW		99	98.8	97.4	98.2
	SGD		99	97	97.6	97.5
	Adam	100	99	98.1	97.7	99.1
YOLO v8	AdamW		99.1	97.7	97.6	99.1
	SGD		98.5	97.5	97.8	97.1
	Adam	50	95.6	93.8	95.7	96.8
	AdamW		99	98.1	97.8	97.7
	SGD		99	98.1	97.3	97.9
		100				
	AdamW		98.9	98.2	97.4	97.9

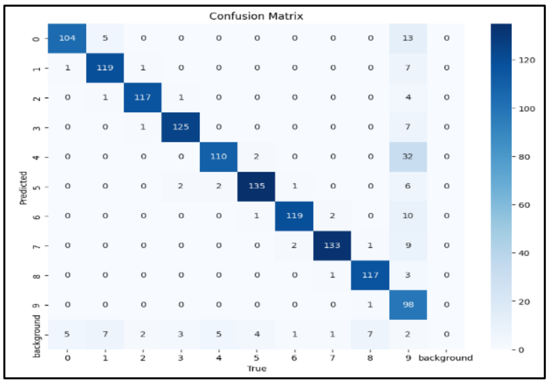

Fig. 3. Confusion matrix for YOLOv8 trained with Adam optimizer.

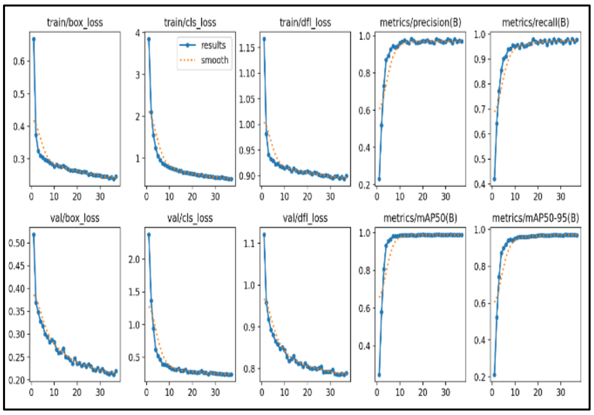

Fig. 4. training and validation performance metrics of YOLOv8

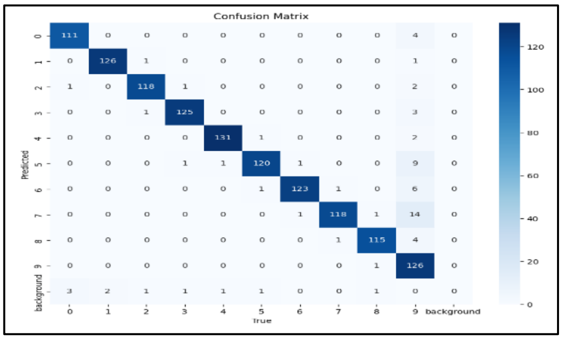

Fig. 5. Confusion matrix for YOLOv10 trained with AdamW optimizer

YOLOV10: YOLOv10 has exceptional processing speed and accuracy. Table 2 provides more facts about this, based on performance metrics for YOLOv10. the model training results are good and can be summarized as follows:

- The AdamW optimizer achieved the highest mAP value (99.1%) and accuracy (98.6%), excelling in complex datasets but requiring longer training times (about 3 hours).
- Performing training for 50 epochs improved performance, resulting in marginal improvements thereafter, and implementing early stopping (about 40 epochs) mitigated overfitting. Low learning rates (0.001) improved mAP and F1-score while increasing rates (0.01) accelerated convergence but reduced recall.
- The confusion matrix shown in Fig. 5 shows negligible misclassifications with only 5 samples misclassified to class 0., with a high diagonal value indicating good class detection. Figure 6 plots of training and validation metrics show steady convergence with decreasing validation losses and continuous improvement in metrics such as recall, precision, and mAP, reflecting the effectiveness of YOLOv10 in handling complex datasets.

Table 2. The Performance Metrics for Yolov10

Model	Optimizer	Epochs	Learning Rate	mAP (%)	Precision (Pr %)	Recall (Rc%)	F1-Score (%)
YOLO v10	SGD	50	0.001	98.8	97.3	94.8	96
	Adam			99	96.8	97.4	97.1
	AdamW			99.1	98.5	97.1	97.8
	SGD	100		98.3	94.7	94.1	94.4
	Adam			98.4	94.4	94.3	94.4
	AdamW			98.9	98.2	97.2	98.2
	SGD	50	0.01	98.8	96	94.9	95.5
	Adam			98.8	97.9	94.3	96
	AdamW			99	98.6	96.4	97.8
	SGD	100		98.2	95.5	92.2	93.8

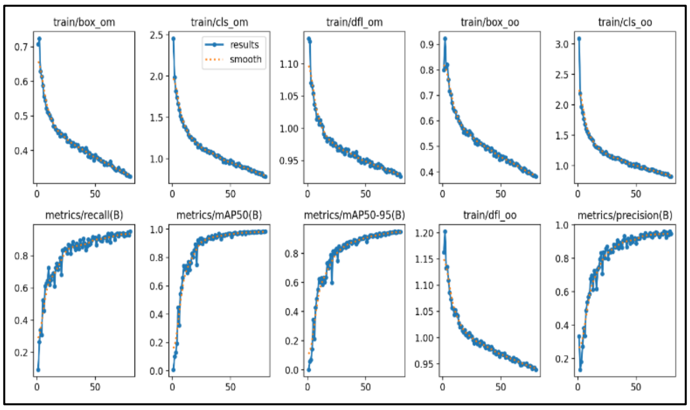

Fig. 6. Training and Validation Performance Metrics for YOLOv10

YOLOV11: The training results of YOLOv11 are presented in Table 3, highlighting the metrics and configurations during the training process, the YOLOv11 model, under the influence of optimizers, learning rates, epochs, and training periods, did remarkably well in object detection. The results can be summarized as follows:

- The optimal configuration, in terms of learning rate of 0.001, 50 epochs, and early stopping, AdamW is the best among the compared optimizers, as it attains the highest mAP at 99.3%, balanced accuracy of 98.9%, a recall of 97.4%, and an F1 score of 98.0%.
- Low learning rates stabilized training, while higher rates improved speed but slightly reduced recall. The confusion matrix Figure 7 reveals high classification accuracy together with very low misclassifications, indicating the strong generalization abilities of the model. As an example, Class 0 made 109 correct predictions with only 2 errors, while Class 1 made 127 correct predictions with only 1 error.

The loss curves in Figure 8 demonstrate the stability of learning by steady decreases in both box loss and classification loss; also, very low validation losses hint at excellent learning. To sum up, YOLOv11 has very good accuracy, efficiency, and reliability about object detection tasks for low-light and low-light environments.

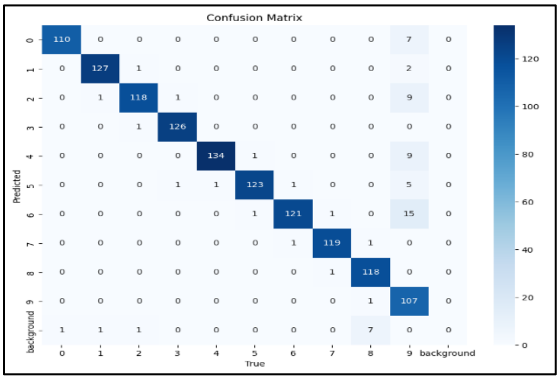

Fig. 7. Confusion matrix for YOLOv11 trained with AdamW optimizer (learning rate = 0.001)

Table 3. The Performance Metrics for Yolov11.

Model	Optimizer	Epochs	Learning Rate	mAP (%)	Precision (Pr %)	Recall (Rc%)	F1-Score (%)
YOLO v11	SGD	50	0.001	99.1	97.5	96.1	97.4
	Adam			99.1	97.8	97.9	97.8
	AdamW			99.2	98.9	98.4	98.6
	SGD	100		98.8	96.4	95.7	96.8
	Adam			99.2	98.6	98.4	98.5
	AdamW			99.2	98.2	98.2	98.4
	SGD	50	0.01	99.2	98.3	98.9	98.5
	Adam			99.2	98.2	97.3	97.6
	AdamW			99.2	98.2	97.3	97.6
	SGD	100		98.4	97.6	95.7	96.6

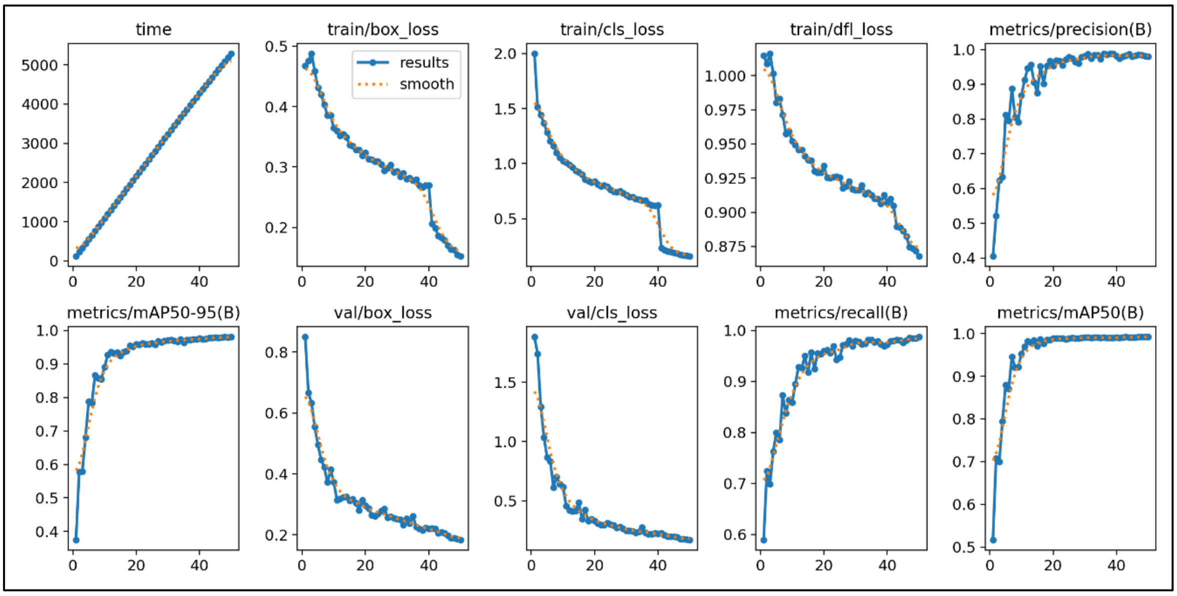

Fig. 8. Training and Validation Performance Metrics for YOLOv11.

YOLO models established reliable performance in detecting hand gestures as they operated effectively through various system configurations. All hand gesture classes received proper classification from these models through precise hand boundary boxes which enabled effective gesture identification. The system demonstrated excellent performance when processing hand images both close-up and zoomed-in which resulted in accurate recognition and identification regardless of the image conditions.

The YOLOv8 model achieved remarkable performance levels when deployed through different optimizer setups. Adam optimizer produced the top training outcome according to Table 1 where it reached 99.2% mean average precision (mAP). The training process delivered efficient results when precision reached 98% and recall achieved 98.4% and F1 score maintained 97.9% marking a suitable balance in performance metrics. The choice of learning rate at 0.001 produced the best mAP scores within the range of 99.0% to 99.2% yet using 0.01 resulted in better recall performance at the cost of slightly diminished mAP. The best results emerged from using Adam optimizer with 50 epochs of training which included early stopping (between 26 to 43 epochs) for preventing overfitting leading to trustworthy outcomes. YOLOv8 demonstrated excellent classification precision as shown in Fig. 3 through its well-distributed high values on the matrix diagonal which minimized wrong predictions especially between classes 2 and 8. YOLOv8 achieved stable convergence through the training and validation metrics that showed decreasing validation losses which demonstrated its general and training effectiveness.

YOLOv10 achieved fast and accurate performance in its operation. A combination of AdamW optimizer led to the best performance results which delivered 99.1% mAP with 98.5% precision and a 97.1% recall during 3 hours of training time according to Table 2. The AdamW configuration delivered superior performance metrics among all options when processing intricate datasets. Similar to YOLOv8 the use of a 0.001 learning rate improved mAP and F1 score while higher learning rate 0.01 brought better recall but produced slightly lower mAP. The best performance came from executing 50 epochs followed by an implementation of early stopping after 40 epochs to avoid overfitting. The confusion matrix presented in Fig. 5 showed minimal misclassification errors since just five samples mistakenly fell into class 0. The model showed excellent generalization across datasets through continuous improvements in precision recall and mAP according to performance metrics Fig. 6.

YOLOv11 achieved excellent object detection performance in low-light conditions according to Table 3 results. The utilization of AdamW optimizer with 0.001 learning rate and 50 epochs reached 99.2% mAP while ensuring optimal precision (98.9%) and recall (98.4%) and F1 score (98.6%). Early stopping as part of this optimizer configuration achieved the best stability while being efficient at preventing model overfitting. Training speed increased with greater learning rate values which resulted in marginal deterioration of detection accuracy. The confusion matrix reveals excellent classification precision because it shows only minimal wrong predictions (Class 0 had 110 correct predictions together with 7 mistakes). The loss curves Fig. 8 showed stable learning because they displayed a steady reduction in both box and classification losses and very low validation losses which demonstrated effective model learning. YOLOv11 proved to be a reliable solution for detecting objects at high speed with strong accuracy within difficult conditions.

4.3 Present study contribution

The research improves hand gesture recognition science by establishing YOLOv11 as an outstanding detector for gestures under diverse backgrounds and environmen-tal changes. The application of AdamW optimization led our model to achieve out-standing performance results including 99.2% mAP accuracy and 98.9% precision supported by 98.4% recall and 98.6% F1 score thus establishing it as an effective classification system for hand gestures. Early stopping with other optimization meth-ods achieved efficient training while solving real-time performance requirements by diminishing overfitting problems. This study evaluates the performance effects of different learning rates and epochs throughout training as it shows practical applica-tions of YOLO for sign language communication systems which help hearing-impaired persons.

5. Conclusion

In general, the YOLOv11 model combined with AdamW optimizer was continuously giving the best balance in accuracy, efficiency, and generalization—the best choice for the task of critical object detection and gesture recognition. The presented model has great performances over all configurations, thus showing the robustness in dealing with complex datasets and unstructured environments. Although YOLOv8 and YOLOv10 are viable alternatives demonstrating competitive performance, they displayed increased sensitivity to variations in hyperparameters and necessitated more meticulous fine-tuning to achieve the same level of efficiency and precision as YOLOv11.

The key importance of fine-tuning the parameters—learning rate, batch sizes, and training durations—was shown in the research for getting optimal results. A smaller learning rate helped to stabilize the training process and improved the accuracy, while the application of early stopping prevented overfitting and reduced the consumption of computational resources.

The results show the great potential of YOLO models for a really large field of applications in robotics, human-computer interaction, and real-time control systems. This could now lead to future research in the integration of YOLOv11 into robotic simulations and gesture-based interfaces for more intuitive and efficient human-robot collaboration.

References

[1] Q. Zhang, X. Guo, H. Xu, and G. Guo, "Design of Writing Robot Experiment System Based on Machine Vision," Research and Exploration in Laboratory, vol. 39, pp. 85–89, 2020.

[2] E. A. Mohammed and G. H. Mohammed, "Robotic vision based automatic pesticide sprayer for infected citrus leaves using machine learning," Przegląd Elektrotechniczny, vol. 99, no. 8, pp. 98–101, 2023, doi: 10.15199/48.2023.08.16.

[3] M. B. S. Bakti and Y. M. Pranoto, "Pengenalan Angka Sistem Isyarat Bahasa Indonesia Dengan Menggunakan Metode Convolutional Neural Network," Inotek, vol. 3, pp. 11–16, 2021, doi: 10.29407/inotek.v3i1.504.

[4] C. Xu, W. Cai, Y. Li, J. Zhou, and L. Wei, "Accurate Hand Detection from Single-Color Images by Reconstructing Hand Appearances," Sensors, vol. 20, no. 1, p. 192, 2020, doi: 10.3390/s20010192.

[5] C. Dewi, R.-C. Chen, and H. Yu, "Weight analysis for various prohibitory sign detection and recognition using deep learning," Multimedia Tools and Applications, vol. 79, pp. 32897–32915, 2020, doi: 10.1007/s11042-020-09509-x.

[6] S. R. Bose and V. S. Kumar, "Precise Hand Gesture Recognition under Hard Visual Environments using Deep Architecture," SN Computer Science, vol. 5, p. 214, 2024, doi: 10.1007/s42979-023-02474-w.

[7] Z. Zheng, "Gesture recognition real-time control system based on YOLOV4," Journal of Physics: Conference Series, vol. 2196, p. 012026, 2022, doi: 10.1088/1742-6596/2196/1/012026.

[8] S. C. Mesbahi, M. A. Mahraz, J. Riffi, and H. Tairi, "Hand Gesture Recognition Based on Various Deep Learning YOLO Models," International Journal of Advanced Computer Science and Applications, vol. 14, no. 4, pp. 307–319, 2023, doi: 10.14569/IJACSA.2023.0140435.

[9] R. Chen and X. Tian, "Gesture Detection and Recognition Based on Object Detection in Complex Background," Applied Sciences, vol. 13, p. 4480, 2023, doi: 10.3390/app13074480.

[10] E. Daniel, V. Kathiresan, C. Priyadarshini, R. G. Nancy, and P. Sindhu, "Real Time Sign Recognition using YOLOv8 Object Detection Algorithm for Malayalam Sign Language," Fusion: Practice and Applications, vol. 17, pp. 135–145, 2025, doi: 10.54216/FPA.170110.

[11] B. S. Vidhyasagar, A. S. Lakshmanan, M. K. Abishek, and S. Kalimuthu, "Video Captioning Based on Sign Language Using YOLOV8 Model," in Internet of Things, D. Puthal, S. Mohanty, and B.-Y. Choi, Eds., Advances in Information and Communication Technology, vol. 683. Cham, Switzerland: Springer, 2023, pp. 306–315, doi: 10.1007/978-3-031-45878-1_21.

[12] A. Mujahid et al., "Real-time hand gesture recognition based on deep learning YOLOv3 model," Applied Sciences, vol. 11, no. 9, 2021.

[13] V. Moysiadis et al., "An Integrated Real-Time Hand Gesture Recognition Framework for Human–Robot Interaction in Agriculture," Applied Sciences, vol. 12, no. 16, 2022.

[14] N. Herbaz, H. El Idrissi, and A. Badri, "Deep Learning Empowered Hand Gesture Recognition: Using YOLO Techniques," in Proc. 14th Int. Conf. Intelligent Systems: Theories and Applications (SITA), 2023.

[15] S. C. Mesbahi, M. A. Mahraz, J. Riffi, and H. Tairi, "Hand Gesture Recognition Based on Various Deep Learning YOLO Models," International Journal of Advanced Computer Science and Applications, vol. 14, no. 4, pp. 307–319, 2023.

[16] S. M. Rayeed, "American Sign Language Digit Dataset," Kaggle, [Online]. Available: https://www.kaggle.com/datasets/rayeed045/american-sign-language-digit-dataset Accessed: Dec. 20, 2024].

[17] K. Abdelhamed, "Sign Language Digits Images Dataset," Kaggle, [Online]. Available: https://www.kaggle.com/datasets/kareemabdelhamed/sign-language-digits-images-dataset [Accessed: Dec. 20, 2024].

[18] Ultralytics, "YOLOv8 Documentation," 2020. [Online]. Available: https://docs.ultralytics.com/

[19] Ultralytics, "YOLOv10: Real-Time End-to-End Object Detection," 2024. [Online]. Available: https://ultralytics.com/

[20] Ultralytics, "YOLO11," 2024. [Online]. Available: https://ultralytics.com/