Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 7, 665-675
2025
Publisher: Learning Gate
DOI: 10.55214/25768484.v9i7.8707
© 2025 by the authors; licensee Learning Gate

MPH-YOLO method for isolating the controller hand amidst multiple hands in digital interaction environments

Agustinus Rudatyo Himamunanto[1,2], Supriadi Rustad[1], Mochammad Arief Soeleman[1], Guruh Fajar Shidik[1]
[1]Faculty of Computer Science, Dian Nuswantoro University, Semarang, Indonesia; rudatyo@ukrim.ac.id (A.R.H.).
[2]Department of Informatics, Immanuel Christian University, Yogyakarta, Indonesia.

Abstract: Hands play a vital role in human-computer interaction. However, ambiguity arises when multiple hand gestures appear within the same frame, leading to misinterpretation in gesture-controlled systems. This study proposes a hybrid method combining MediaPipe Hands (MPH) and a modified YOLO framework to isolate a single control hand using a visual marker. MPH detects hand landmarks, and YOLO identifies whether the hand contains a marker. Experiments involving 800 test videos showed that the method achieved 97.5% accuracy in correctly identifying the controller hand across various visual conditions. The proposed approach contributes to the robustness and precision of real-time gesture-based control systems.

Keywords: *Gesture recognition, Hand isolation, Hand marker, MediaPipe Hands, YOLO.*

1. Introduction

The human hand is one of the most frequently utilized body parts in everyday activities [1, 2]. Among its many functions, the hand serves as an effective medium for nonverbal communication, facilitating the expression of meaning through specific gesture patterns [3].

With the advancement of digital and computing technologies, the role of the hand has expanded significantly, particularly as a direct operational interface for various digital systems. Hands have been employed in a broad range of control applications, including computer command execution [4, 5] television set control [6] light level control of lamps [7] robot motion control [8] speaker sound scale control [9] sign language recognition [10] area access security [11] and so on.

Typically, a hand intended for control is visually captured through a camera device. The captured hand image is then processed to interpret the intended gesture, which is subsequently translated into an executable command corresponding to that gesture [12]. However, a significant challenge arises when more than one hand appears within the same visual frame. This condition introduces potential ambiguity, which can adversely affect recognition accuracy and system reliability.

MediaPipe Hands (MPH) is a popular framework that works to find the hand to wrist area [13]. YOLO is a popular framework that has the working characteristic of marking similar objects or uniform objects that appear in one frame [14]. YOLO needs special treatment when it comes to marking or locking only one object among a number of similar objects or a number of uniform objects. In accordance with its role and training, YOLO will work and try to find all objects that have similarities with its training data [15].

This research proposes a method for isolating a specific target hand from among multiple visually similar hand objects by combining *MediaPipe Hands* (MPH) for hand localization and YOLO for marker-based identification. The objective is to enable selective hand recognition, ensuring that only the designated controller hand is processed while disregarding others. This approach is expected to enhance the reliability of gesture-based control systems and support applications such as sign language

© 2025 by the authors; licensee Learning Gate
History: Received: 23 April 2025; Revised: 6 June 2025; Accepted: 9 June 2025; Published: 9 July 2025
* Correspondence: rudatyo@ukrim.ac.id

recognition for individuals with special needs or elderly users, as well as gesture-driven access authentication.

2. Literature Review

Based on the literature we reviewed, there appears to be a lack of studies that directly address the challenge of isolating a single target hand from among multiple visually similar hand objects. In particular, we found no prior work that explores this problem through the integration of Media Pipe for landmark detection and YOLO for marker-based classification within the same framework.

In the domain of computer vision applied to hand gesture recognition, gesture detection [16] and segmentation [17] are fundamental operations for characterizing hand presence in visual input. Some methods achieve high detection accuracy, exceeding 95% [18] and demonstrate strong segmentation performance [8]. However, these approaches often suffer from delays, particularly during object locking or visual tagging [19] which are further worsened under dynamic conditions such as hand movement [20].

YOLO-based models have shown strong performance in object detection tasks, including hand-related applications [21]. Nonetheless, YOLO's default behavior is to detect all similar or uniform objects within a frame indiscriminately [22] limiting its use in selective hand recognition scenarios.

Several studies have integrated YOLO with hand gesture recognition. Ye, et al. [23] proposed a YOLOv3-based model combined with PCA and Bayes classifiers to improve classification accuracy using single-hand input. Mujahid, et al. [24] implemented a YOLOv3 system with Pascal VOC for hand gesture recognition, reporting accuracy, precision, recall, and F1-scores of 97.68%, 94.88%, 98.66%, and 96.70%, respectively.

Chang, et al. [25] deployed a YOLO-based gesture recognition model on edge devices using Kubernetes, focusing on dynamic hand gestures from single-hand inputs. Similarly, Wibowo and Sugiarto [26] utilized YOLOv5 for recognizing hand symbols in human-computer interaction, also limited to single-hand data.

Based on the literature we reviewed, there appears to be a lack of studies that directly address the challenge of isolating a single target hand from among multiple visually similar hand objects. In particular, we found no prior work that explores this problem through the integration of MediaPipe Hands for landmark detection and YOLO for marker-based classification within the same framework.

3. Method

MediaPipe Hands (MPH) is a hand-tracking framework developed by Google and introduced by Zhang, et al. [13] at the Fourth Workshop on Computer Vision for AR/VR. The framework is designed to detect and localize hand and wrist regions in real-time by constructing a topological model of hand joints using 21 landmark points across the hand surface as illustrated in Figure 01.

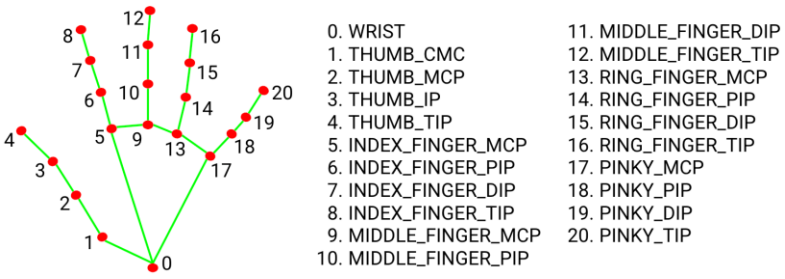


Figure 1.
21-point hand construction marker (Taken from Google MediaPipe Hands).

YOLO (You Only Look Once) was first introduced by Redmon [14] at the IEEE Conference on Computer Vision and Pattern Recognition [25]. It is a real-time object detection algorithm that

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 7: 665-675, 2025
DOI: 10.55214/25768484.v9i7.8707
© 2025 by the authors; licensee Learning Gate

leverages a Convolutional Neural Network (CNN) to perform a single forward pass over the entire image, enabling simultaneous prediction of object classes and bounding boxes. As implied by its name, YOLO processes the input image in one pass, making it highly efficient for real-time applications. An overview of the YOLO workflow is illustrated in Figure 02.

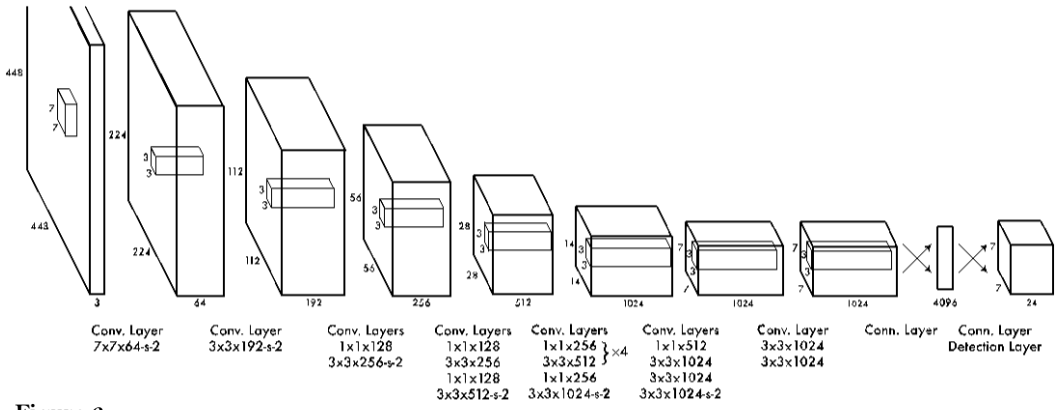


Figure 2.
Flowchart of object detection process in YOLO Ge, et al. [21].

The YOLO algorithm applies a convolutional neural network (CNN) to divide an input image into a grid structure, predicting bounding boxes and class probabilities for each grid cell. Each bounding box is defined by five parameters: the x and y coordinates (center point), width, height, and a confidence score indicating the likelihood of object presence. The output dimension is typically expressed as $S \times S \times (B \times 5 + C)$, where S is the number of grid divisions, B is the number of bounding boxes per grid cell, and C is the number of object classes. All bounding box attributes are normalized to values between 0 and 1, with coordinate normalization based on the grid cell's top-left corner and size normalization relative to the full image dimensions.

Given the characteristics of MediaPipe Hands and YOLO, the proposed approach aims to rapidly localize hand regions and selectively identify the marked hand. This study explores the integration of the MPH–YOLO framework to support real-time applications involving hand-based control. Modifications are introduced to enable YOLO to focus on detecting only one designated object among multiple visually similar instances.

To achieve object-level locking, the YOLO framework is extended to mark only a single hand object within a frame. A schematic of the proposed enhancement is presented in Figure 03.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 7: 665-675, 2025
DOI: 10.55214/25768484.v9i7.8707
© 2025 by the authors; licensee Learning Gate

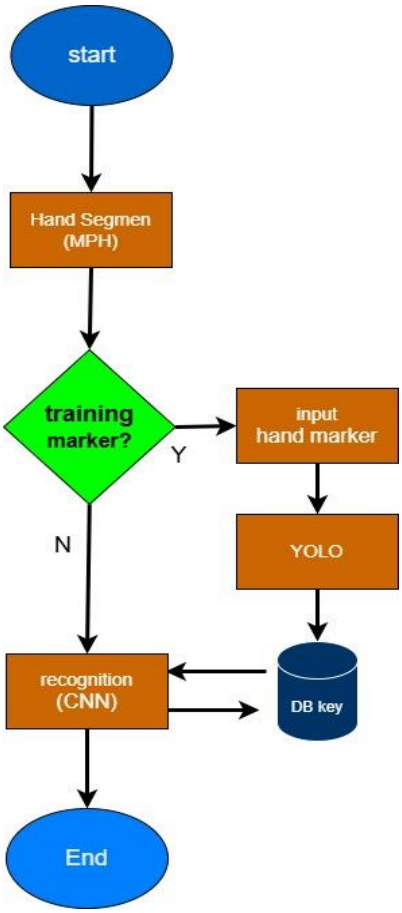


Figure 3.
Flowchart of *MPH-YOLO* approach.

The input data in this study consists of short-duration hand gesture videos, each ranging from 3 to 5 seconds and recorded at 25–30 frames per second (fps). These videos were specifically selected to represent diverse lighting conditions, including normal room lighting and low-light environments, in order to evaluate the robustness of the proposed modified YOLO method. The dataset comprises both single-hand gesture scenes and multi-hand configurations within the same frame. This variation is intended to simulate real-world interaction scenarios and test the method's ability to isolate the target hand accurately. The inclusion of multiple hand instances also serves to assess the selectivity of the system in distinguishing and locking onto only the designated hand gesture.

The input videos in this study are categorized into two distinct groups based on the number of visible hands in each frame. The first group consists of videos containing only a single hand performing specific gesture movements. The second group includes videos where multiple hands appear simultaneously, allowing evaluation of the system's ability to isolate the intended target hand amidst visual clutter. This classification is designed to simulate both ideal and complex real-world scenarios. Representative samples from each video group are illustrated in Figure 04.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 7: 665-675, 2025
DOI: 10.55214/25768484.v9i7.8707
© 2025 by the authors; licensee Learning Gate



Figure 4.
Example of 1-hand and multi-hand visual data input.

The total amount of hand gesture video input data that will be provided is 3000 data in the form of both primary datasets and secondary datasets. The secondary dataset that will be used is taken from the HaGRID hand gesture dataset [27] the 20bn Jester Hand dataset and IPN Hand Dataset [28].

HaGRID dataset stands for HAnd Gesture Recognition Image Dataset. The size of HaGRID is 716GB and the dataset contains 552,992 FullHD (1920×1080) RGB images divided into 18 gesture classes. In addition, some images have a no gesture class if there is a free second hand in the frame. These additional classes contain 123,589 samples. The data is divided into 92% training, and 8% testing assigned based on the subject's user id, with 509,323 images for training and 43,669 images for testing. The 20bn-Jester Hand database has a total of 71724 training data and 9036 validation sample data. The image size is 100x100 with 36 frames. The IPN Hand Dataset contains over 4,000 example gestures and 800,000 frames from 50 subjects, 13 classes of static and dynamic gestures.

The secondary dataset comprises hand gesture video inputs that were independently collected and constructed during the course of this research. These videos were developed to complement existing datasets and to simulate varied gesture patterns under controlled conditions. In addition to the secondary dataset, a primary dataset was also prepared, consisting of up to 500 video samples of hand gesture movements. This primary dataset serves as the core training and testing material for validating the proposed method's performance.

This study adopts a 3-fold cross-validation approach to evaluate the performance of the proposed method. The dataset is divided into three equal parts, with each part used once as validation data while the remaining two parts serve as training data. Cross-validation, also known as k-fold validation, is commonly employed to improve evaluation reliability and minimize overfitting [29]. The analysis of test data is conducted only after all testing procedures have been completed and the corresponding output has been generated.

4. Initial Preparation and Initiation

The proposed hand isolation method integrates a sequential detection and classification pipeline. MediaPipe Hands (MPH) is first employed to localize the wrist region based on landmark detection. The extracted region is then passed to YOLOv8, which classifies the presence of a visual marker to determine whether the hand should be isolated for further processing.

Media Pipe Hands (MPH) is utilized to detect hand landmarks in each video frame. Once the landmarks are extracted, the minimum and maximum positions across all landmark points are computed to define a bounding box encompassing the entire hand and wrist region. This bounding box is represented by the coordinates (x_1, y_1) and (x_2, y_2), corresponding to the top-left and bottom-right corners of the region of interest.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 7: 665-675, 2025
DOI: 10.55214/25768484.v9i7.8707
© 2025 by the authors; licensee Learning Gate

```
x1_hand = int(min(hand_landmarks.landmark[i].x * frame.shape[1] for i in range(21)))
y1_hand = int(min(hand_landmarks.landmark[i].y * frame.shape[0] for i in range(21)))
x2_hand = int(max(hand_landmarks.landmark[i].x * frame.shape[1] for i in range(21)))
y2_hand = int(max(hand_landmarks.landmark[i].y * frame.shape[0] for i in range(21)))
```

Figure 5.
Python snippet for computes the minimum and maximum x and y.

As shown in Figure 05, this code snippet computes the minimum and maximum x and y coordinates from 21 hand landmarks to form a bounding box that encapsulates the entire hand region. The coordinates are normalized based on frame dimensions to ensure accurate spatial localization.

After obtaining the bounding box coordinates, the segmentation process is carried out to extract the hand region of interest (ROI) for classification. This is achieved using OpenCV's built-in cropping function, which takes the coordinates (x_1, y_1, x_2, y_2) to isolate the relevant region from the original frame. A script for resulting segmentation output are provided in the Figure 06. below.

```
ROI = frame[y1_hand:y2_hand, x1_hand:x2_hand]
```

Figure 6.
Python snippet for ROI cropping using bounding box coordinates.

Figure 7 below shows an example of segmented hand data used during training. The image illustrates how the region of interest (ROI) was extracted from the original frame to isolate the hand. This segmentation process was essential for preparing consistent input for the classification model.



Figure 7.
Example of segmentation results: white marker, black marker and no marker.

There are 1000 data without markers (normal hands) and 1000 data with markers (hands with markers on the wrist). The data is then manually labelled with two classes that are band, which is hand with marker and no band, for hand without marker. All this data was then divided into 70% training data, 20% validation data, 10% test data.

YOLO is used to train a classification model to distinguish between hands with and without markers on the wrist. The labelled training data was used to train the model. The training data used 65 epochs and resulted in an accuracy above 90%, as shown in Figure 08. The following graph, Figure 08., shows the results of checking the accuracy.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 7: 665-675, 2025
DOI: 10.55214/25768484.v9i7.8707
© 2025 by the authors; licensee Learning Gate

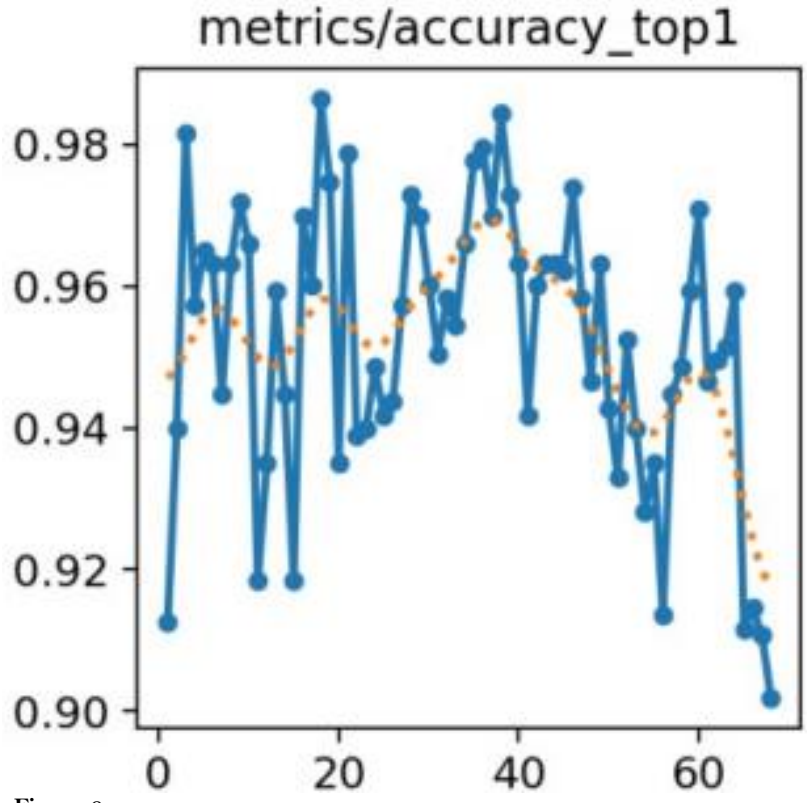


Figure 8.
Post-training accuracy check graph.

After the YOLO model is complete, all scripts are put together where the video is taken, each frame is processed by detecting the wrist body part using *MPH*, the wrist result is cropped, then becomes input to *YOLOv8* to be classified whether it has a marker or not. Only hands with markers will be retrieved and processed for further needs.

5. Result and Discussion

The testing process began with training the system to recognize palm patterns marked with visual indicators, which were stored as reference data. It was followed by training with unmarked hand gestures to evaluate the system's response to non-target inputs. Both primary and secondary datasets were used during testing. The first scenario involved a single unmarked hand to check whether the system would ignore non-relevant inputs. The next test involved a single hand with a marker to confirm whether the system could correctly detect and classify the palm. A final test used multiple hands, with only one bearing a marker, to examine whether the system could focus on and isolate the correct palm object. An example of this testing scenario is presented in Figure 09.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 7: 665-675, 2025
DOI: 10.55214/25768484.v9i7.8707
© 2025 by the authors; licensee Learning Gate



Figure 9.
Input examples for different test scenarios: one hand without marker, one hand with marker, multiple hands without marker, multiple hands with one marked hand.

The proposed model integrates *MediaPipe Hands* for detecting the hand area and YOLO for classifying the presence of a visual marker. This combination shows strong performance in isolating the target hand, achieving accuracy above 90% under challenging conditions such as low lighting and rapid hand motion. Optimal results are obtained when the hand is positioned at a distance of approximately 1.0 to 1.1 meters from the camera.

The average processing time per frame was 36.8 milliseconds, which enables the system to operate at approximately 25–26 frames per second. This latency is within the acceptable threshold for real-time gesture control applications. Additionally, the marker must remain fully visible and unobstructed by other objects to ensure successful isolation.

The following below, Table 01., is a table of complete test results conducted 800 times involving 800 data of various conditions.

Table 1.
Testing Results.

Testing	Total	Correct result
One nonmarker hand	200	200
One Hand Marker	200	200
Many nonmarker hands	200	200
Many hands + One marked hand	200	180

Based on the test results involving 800 test data in the form of video data with visual images of one hand or more than one hand image, it is known that the YOLO modification results show success with 97.5% accuracy.

Several recent studies have demonstrated strong performance in hand gesture recognition and tracking using various deep learning approaches. Mujahid, et al. [24] utilized YOLOv3 for single-hand gesture classification and reported an accuracy of 97.68%, while Chang, et al. [25] implemented a YOLO-based model optimized for edge deployment, achieving 96.5% in real-time gesture recognition. Wibowo and Sugiarto [26] applied YOLOv5 for human–computer interaction using hand symbol classification, with performance ranging from 95% to 96%.

In a different domain, Budzinski, et al. [30] employed MediaPipe's 3D hand tracking to evaluate simulated hand movement accuracy in medical training environments. Their model achieved an average error of 3.1 mm over a 500 mm range, corresponding to an estimated tracking precision of

© 2025 by the authors; licensee Learning Gate

approximately 99.4%. However, their work did not involve gesture recognition or object isolation in visually cluttered scenes.

More recently, Alsharif, et al. [31] combined YOLOv8 and MediaPipe for static ASL alphabet recognition. Their model achieved high precision (98%), recall (98%), and F1-score (99%), with a mean average precision (mAP) of 98% and mAP50–95 of 93%. Despite the high metrics, their study was conducted on single-hand static gesture images, without the challenge of isolating one hand among visually similar distractors.

Table 2.
Final Comparative Table of Hand Recognition Studies.

Study	Method	Task Scenario	Performance Metric
Mujahid, et al. [24]	YOLOv3 (Single hand)	Single-hand gesture classification	Accuracy: 97.68%
Chang, et al. [25]	YOLO + Edge deployment	One-hand real-time gesture recognition	Accuracy: 96.5%
Wibowo and Sugiarto [26]	YOLOv5 for gesture classification	Single-hand HCI-based symbol classification	Accuracy: 95–96%
Budzinski, et al. [30]	MediaPipe 3D tracking (medical training)	3D motion analysis for simulated hand movement	3.1 mm avg error (~99.4%)
Alsharif, et al. [31]	YOLOv8 + MediaPipe (ASL static images)	ASL alphabet recognition from static gesture images	Precision: 98%, Recall: 98%, F1: 99%, mAP: 98%, mAP50–95: 93%
This Study (MPH–YOLO)	MediaPipe + YOLOv8 (Multi-hand)	Isolating one hand among multiple similar hands	Accuracy: 97.5%

Compared to these studies, as shown in Table 02., the present research introduces a novel contribution by addressing the challenge of isolating a single target hand from among multiple visually similar hand objects. Using a modified integration of MediaPipe and YOLOv8, our model achieved an accuracy of 97.5% under test conditions that included low lighting, hand motion, and multi-hand interference. This positions our approach as a distinct solution within the landscape of hand recognition systems, especially in scenarios requiring selective attention and control.

Based on the research and testing, it is known that the development of MPH-YOLO showed good and significant success.

6. Conclusion

Based on the test results involving 800 video samples containing visual input of single and multiple hand gestures, the modified MPH-YOLO model achieved an accuracy of 97.5%. These findings demonstrate the model's effectiveness in isolating the intended hand object using visual markers. However, this study also highlights a potential area for future research: enabling palm recognition without relying on markers. Future development will focus on identifying the primary hand object in multi-hand scenarios using markerless detection strategies.

Transparency:

The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.

Copyright:

© 2025 by the authors. This open-access article is distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

References

[1] A. P. Dhote and V. R. Parihar, "Gesture recognition," in *Proceedings of the Icici (pp. 512–516)*, 2017.

[2] A. P. Dhote, V. R. Parihar, R. P. Sonar, and R. S. Nage, "A novel system of real time hand tracking and gesture recognition," presented at the International Conference on Inventive Computing and Informatics (ICICI), 512–516, 2017.

[3] N. Patel and S. J. He, "A survey on hand gesture recognition techniques, methods and tools," *International Journal of Research in Advent Technology*, vol. 6, no. 6, 2018.

[4] M.-C. Roh, D. Kang, S. Huh, and S.-W. Lee, "A virtual mouse interface with a two-layered Bayesian network," *Multimedia Tools and Applications*, vol. 76, pp. 1615-1638, 2017.

[5] P. Xu, "A real-time hand gesture recognition and human-computer interaction system," *arXiv preprint arXiv:1704.07296*, 2017.

[6] U. V. Solanki and N. H. Desai, "Hand gesture based remote control for home appliances: Handmote," presented at the World Congress on Information and Communication Technologies (pp. 419-423). IEEE, 2011.

[7] A. Chaudhary and J. Raheja, "Light invariant real-time robust hand gesture recognition," *Optik*, vol. 159, pp. 283-294, 2018.

[8] A. Chaudhary, *Robust hand gesture recognition for robotic hand control*. Singapore: Springer, 2018.

[9] W.-P. Fang, "An intelligent hand gesture extraction and recognition system for home care application," presented at the Sixth International Conference on Genetic and Evolutionary Computing, 457–459, 2012.

[10] C. N. Nyaga and R. D. Wario, "Sign language gesture recognition through computer vision," presented at the IST-Africa Week Conference (IST-Africa) (pp. Page-1). IEEE, 2018.

[11] J. Xu, X. Zhang, and M. Zhou, "A high-security and smart interaction system based on hand gesture recognition for internet of things," *Security and Communication Networks*, vol. 2018, no. 1, p. 4879496, 2018.

[12] J. Wang and S. Payandeh, "Hand motion and posture recognition in a network of calibrated cameras," *Advances in Multimedia*, vol. 2017, no. 1, p. 2162078, 2017.

[13] F. Zhang *et al.*, "Mediapipe hands: On-device real-time hand tracking," *arXiv preprint arXiv:2006.10214*, 2020.

[14] J. Redmon, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016.

[15] M. Soeleman, Arief Soeleman and C. Supriyanto, "Deep learning model for unmanned aerial vehicle-based object detection on thermal images," *Revue d'Intelligence Artificielle*, vol. 37, no. 6, pp. 1441–1447, 2023. https://doi.org/10.18280/ria.370608

[16] T. Grzejszczak, M. Kawulok, and A. Galuszka, "Hand landmarks detection and localization in color images," *Multimedia Tools and Applications*, vol. 75, pp. 16363-16387, 2016.

[17] J. Lin, X. Ruan, N. Yu, and J. Cai, "Multi-cue based moving hand segmentation for gesture recognition," *Automatic Control and Computer Sciences*, vol. 51, pp. 193-203, 2017.

[18] A. Tagliasacchi, M. Schröder, A. Tkach, S. Bouaziz, M. Botsch, and M. Pauly, "Robustarticulated-icp for real-time hand tracking," *Computer Graphics Forum*, vol. 34, no. 5, pp. 101–114, 2015.

[19] A. I. Maqueda, C. R. del-Blanco, F. Jaureguizar, and N. García, "Human–computer interaction based on visual hand-gesture recognition using volumetric spatiograms of local binary patterns," *Computer Vision and Image Understanding*, vol. 141, pp. 126-137, 2015.

[20] W. H. Chun and T. Höllerer, "Real-time hand interaction for augmented reality on mobile phones," in *Proceedings of the 2013 International Conference on Intelligent User Interfaces*, 307–314, 2013.

[21] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "Yolox: Exceeding yolo series in 2021," *arXiv preprint arXiv:2107.08430*, 2021.

[22] X. Long *et al.*, "PP-YOLO: An effective and efficient implementation of object detector," *arXiv preprint arXiv:2007.12099*, 2020.

[23] B. Ye, S. Jin, B. Li, S. Yan, and D. Zhang, "Dual histogram equalization algorithm based on adaptive image correction," *Applied Sciences*, vol. 13, no. 19, p. 10649, 2023.

[24] A. Mujahid *et al.*, "Real-time hand gesture recognition based on deep learning YOLOv3 model," *Applied Sciences*, vol. 11, no. 9, p. 4164, 2021. https://doi.org/10.3390/app11094164

[25] T.-P. Chang, H.-M. Chen, S.-Y. Chen, and W.-C. Lin, "Deep learning model for dynamic hand gesture recognition for natural human-machine interface on end devices," *International Journal of Information System Modeling and Design*, vol. 13, no. 10, pp. 1-23, 2022.

[26] S. Wibowo and I. Sugiarto, "Hand symbol classification for human-computer interaction using the fifth version of YOLO object detection," *CommIT (Communication and Information Technology) Journal*, vol. 17, no. 1, pp. 43-50, 2023.

[27] A. Kapitanov, K. Kvanchiani, A. Nagaev, R. Kraynov, and A. Makhliarchuk, "HaGRID--hand gesture recognition image dataset," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 4572–4581, 2024.

[28] J. Materzynska, G. Berger, I. Bax, and R. Memisevic, "The jester dataset: A large-scale video dataset of human gestures," presented at the Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, 0, 2019.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 7: 665-675, 2025
DOI: 10.55214/25768484.v9i7.8707
© 2025 by the authors; licensee Learning Gate

[29] D. Berrar, "Cross-validation. In S. Ranganathan, M. Gribskov, K. Nakai, & C. Schönbach (Eds.)," *Encyclopedia of Bioinformatics and Computational Biology*, vol. 1, pp. 542–545, 2019. https://doi.org/10.1016/B978-0-12-809633-8.20349-X

[30] C. Budzinski, H. L. Wu, E. Sarraf, S. Miller, and J. Moore, "Accuracy of mediapipe visual hand tracking for use in medical training procedures," in *Proceedings of the 2024 Design of Medical Devices Conference, DMD*, 2024, pp. 8–11, doi: https://doi.org/10.1115/DMD2024-1039.

[31] B. Alsharif, E. Alalwany, and M. Ilyas, "Transfer learning with YOLOV8 for real-time recognition system of American Sign Language Alphabet," *Franklin Open*, vol. 8, p. 100165, 2024.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 7: 665-675, 2025
DOI: 10.55214/25768484.v9i7.8707
© 2025 by the authors; licensee Learning Gate