Indonesian Journal of Data and Science

Volume 6 Issue 2
ISSN 2715-9936
https://doi.org/10.56705/ijodas.v6i2.247

Research Article

Integration of Yolov8 And Instance Segmentation in The Chinese Sign Language (CSL) Recognition System

Mikel Ega Wijaya [1], Anik Nur Handayani [2,*]

[1] Universitas Negeri Malang, Malang, Indonesia, egawijaya701@gmail.com
[2] Universitas Negeri Malang, Malang, Indonesia, aniknur.ft@um.ac.id

Correspondence should be addressed to Anik Nur Handayani; aniknur.ft@um.ac.id

Received 12 January 2025; Accepted 23 May 2025; Published 31 July 2025

Abstract:

This research aims to develop an advanced recognition system for Chinese Sign Language (CSL) by integrating YOLOv8 and instance segmentation techniques. Communication through sign language is essential for the deaf community, and although CSL has been standardized in China, recognizing complex hand movements remains a significant challenge. YOLOv8 is employed for real-time object detection, while instance segmentation is used to provide more detailed analysis of hand gestures. This integration seeks to improve hand gesture recognition under varying lighting and background conditions, which is crucial for more effective communication between the deaf community and the wider society. The study evaluates the system's performance using common metrics such as Mean Average Precision (mAP), precision, recall, and F1-score. The findings indicate that the non-segmentation model performs better than the segmentation model in terms of precision, recall, and mAP, especially when trained with a larger dataset ratio. The non-segmentation model provides faster and more accurate detection, while the segmentation model, despite using the same amount of data, shows potential for more detailed recognition of gestures. Although the segmentation model shows improvements in the F1-score with more detailed accuracy, the non-segmentation model remains superior in overall detection speed and accuracy. This research highlights the importance of integrating YOLOv8 and instance segmentation for improving CSL recognition, with better results on the non-segmentation model for more effective communication for the deaf.

Keywords: Chinese Sign Language, YOLOv8, Instance Segmentation, Hand Gesture Recognition, Deep Learning.

1. Introduction

In everyday life, communication plays an important role in interactions between individuals [1]. For most people, verbal communication through spoken language is the primary way to exchange information [2]. However, individuals with hearing impairments, especially those who are deaf, face barriers to effective participation in verbal communication in social settings. Sign language is a solution for the deaf community. Sign language allows them to communicate using hand movements, facial expressions, and body movements [3]. As an independent language, sign language has a structure equivalent to spoken language and is an important part of the deaf culture [4]. In China, the deaf community uses Chinese Sign Language (CSL), which has undergone various standardization efforts to ensure uniformity in communication. In addition, a finger spelling system with 30 symbols is also used to convey words that do not have specific sign movements [5] In 2019, the number of people with hearing loss in China reached 426.5 million, an increase of 90.1% since 1990 [6], highlighting the importance of developing an effective communication system for the deaf

Technological developments have driven innovation in the introduction of sign language, which is now being applied to smart devices to help deaf people communicate. This technology facilitates communication in various aspects of life such as education and work [7]. With increasingly accurate systems, it is hoped that communication barriers can be minimized, making interaction between the deaf community and the wider community smoother. Globalization has also driven advances in communication, including accessibility for the deaf community, facilitating communication between regions and cultures even if they are separated by long distances [8]. However, the main challenge in the introduction of sign language still lies in the identification of complex hand movements.

The hand is the part of the body most often used to convey signals in communication [9]. Hand movements in sign language can be categorized into two types: static and dynamic. Static movements refer to a fixed hand position, while dynamic movements involve a series of sequential changes in hand position [10]. Identifying these two types of movements is a major challenge in sign language recognition systems, especially when the hand objects overlap or have similar shapes but different meanings. To overcome this challenge, YOLOv8 (You Only Look Once) and Instance Segmentation-based methods are potential solutions for improving the accuracy of sign language recognition.

You Only Look Once (YOLO) is a deep learning model developed to detect objects directly and in real-time [11]. YOLOv8, which uses a Convolutional Neural Network (CNN), is capable of detecting and recognizing sign language in real-time with high precision, including in terms of mean Average Precision (mAP) [12]. These results are in line with research [13] that implements YOLO in detecting Indonesian sign language, showing that YOLO has superior performance in this task. On the other hand, [14] emphasizes that instance segmentation plays an important role in improving computing efficiency in deep learning-based object detection. Based on these findings, this study integrates YOLOv8 and instance segmentation to recognize the letters of the Chinese Sign Language (CSL) alphabet from static hand images. Edge detection is also applied to strengthen the visual characteristics of the hand shape, enabling the system to work more accurately in classifying each displayed sign.

This research aims to develop a Chinese Sign Language (CSL) recognition system based on YOLOv8 and Instance Segmentation to clarify hand movements and improve the accuracy of gesture recognition in various lighting and background conditions, thus supporting more effective communication for the deaf.

2. Method:

This research develops a Chinese sign language (CSL) recognition system by integrating YOLOv8 and instance segmentation through a structured approach. The method includes data collection, labelling, augmentation, dataset division, training, validation, and model testing and evaluation. Each stage is designed to have good accuracy in hand gesture recognition, with performance evaluation using relevant metrics to assess the effectiveness of the system in accurately detecting and interpreting gestures.

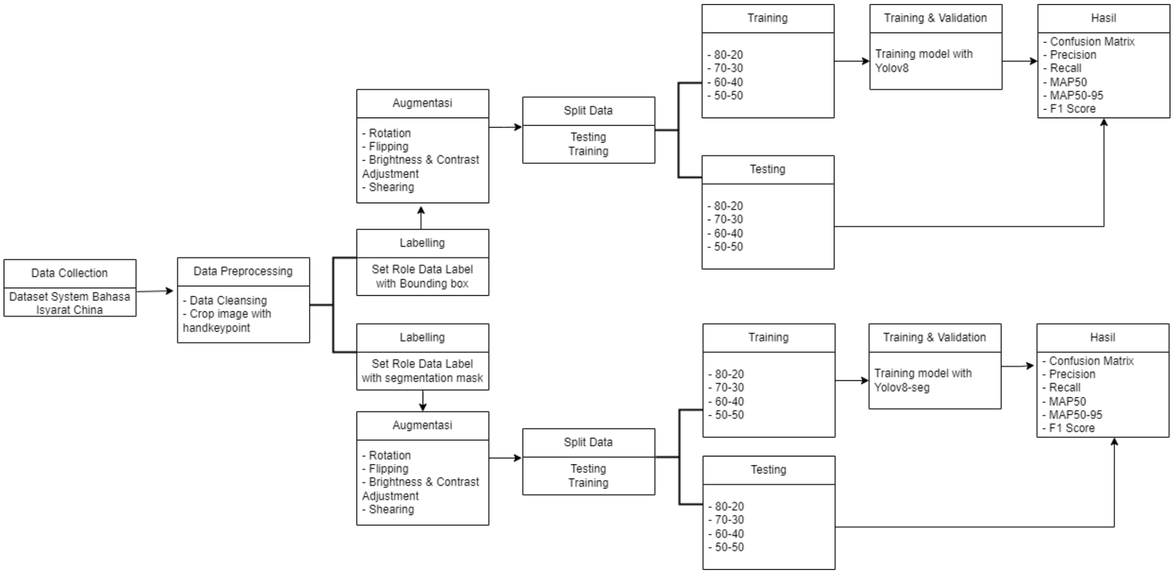

Figure 1. Research Flow Chart

Data Collection

The first stage in this research is data collection, which includes the acquisition process to obtain the required information [15]. This data collection process aims to ensure that the information used in the research is of good quality and relevant. In this research, the Chinese sign language (CSL) hand image dataset was created independently by referring to the official standard of hand gestures in Chinese sign language, and supported by references from previous research. The creation of this dataset is very important to produce data that is in accordance with the rules and regulations that apply in Chinese sign language. This dataset consists of a series of hand gestures that represent the Chinese Sign Language alphabet, so it can be used to train and test sign language recognition models.

In addition, the data collected includes a variety of hand positions, viewing angles, as well as different lighting conditions to improve the generalization ability of the model in detecting and recognizing signs better. The collection of data covering several conditions aims to ensure that the developed model has good robustness in dealing with different situations. Thus, the sign language recognition model trained using this dataset is expected to function optimally in various real-world conditions. This data is used to test and train the Chinese Sign Language (CSL) model [5].

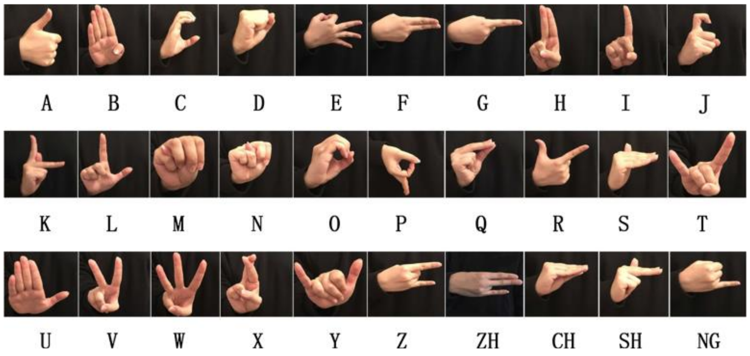

Figure 2. Chinese Sign Language (Source: [5])

Data Pre-processing

The data preprocessing stage is carried out to improve the quality of the dataset before it is used in model training [16]. This process begins with data cleansing, Data cleansing is a crucial process to improve data quality by transforming unstructured data into reliable data [17]. Researchers sorted and deleted images that did not meet the quality criteria, such as images that were too blurry, had a lot of noise, or insufficient lighting. This step aims to ensure that only clear and representative images are used in the model training and testing process.

After the data cleaning process, all images in the dataset will be further processed with a cropping technique based on hand keypoint detection. The hand keypoint technique is proven to be effective in detecting objects with high accuracy and precision, especially in identifying key features of the hand [18], [19]. Thus, the main features in sign language can be more easily recognized by the model. In addition, all processed images are adjusted to a uniform size to ensure consistency in the training process and improve the model's performance in accurately detecting hand gestures.Yolo v8 architecture

Labelling

The labeling process is a crucial step in data preparation that involves assigning a specific role to each attribute in the dataset [20]. The labeling process is done by annotating the image dataset [21]. This process is essential for training detection and segmentation models. These annotations identify and tag objects, so that the model can learn the cue patterns accurately. In this research, the labeling process is done using Roboflow, a platform that simplifies annotation as well as dataset management. Roboflow allows users to label images in a more efficient and structured manner, which is very helpful in processing large and complex datasets.

Each image in the dataset was labeled with two different types of annotations according to the needs of the model. For training using YOLOv8, each hand gesture is assigned a bounding box, which serves to determine the location of objects in the image. Meanwhile, for training with YOLOv8-Seg, the image is given a segmentation mask, which marks specific areas of the fingers as well as the hand shape in more detail. YOLOv8-seg is used in this study because it is a variant of the YOLO (You Only Look Once) model specifically designed for instance segmentation tasks. With this approach, the model can not only recognize the presence of hands in the image, but is also able to understand the shape and structure of gestures more precisely, thus improving the performance in detecting and interpreting Chinese sign language.

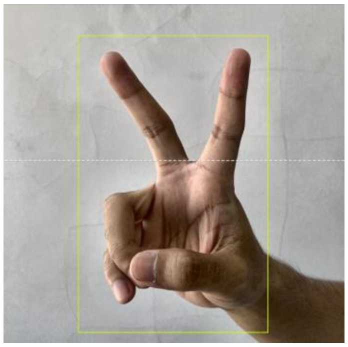

Figure 3. Bounding Box Labelling Example

Figure 4. Labelling Segmentation Example

Augmentation

The data augmentation process was carried out using Roboflow. Data augmentation is a technique commonly used to increase the diversity of training samples by synthesizing new data from an existing dataset [22], [23]. This process is crucial in machine learning as it helps make models more robust by introducing variations of the original data. In this study, the augmentation process involved tripling the number of images from the original dataset, which helped increase the size and diversity of the dataset. Consequently, the model can be trained with a broader range of inputs, improving its ability to generalize and perform better on unseen data.

Augmentation has been proven effective in various studies for improving model generalization, especially in situations where datasets are limited [22]. This technique helps reduce overfitting, a common issue when training models on small datasets with limited variation. By expanding the dataset with augmented images, the model is exposed to more diverse examples, enabling it to learn more generalizable features. This is particularly important for enhancing the model's ability to generalize and interpret data effectively across different scenarios.

Figure 5. Example of Image Augmentation

Split Data

The separation of the dataset into training and testing subsets is a crucial step in data analysis and machine learning to ensure that the model learns effectively and is tested appropriately. This division is generally done by dividing the dataset into two subsets: training (training set) to train the model, and testing (testing set) to evaluate its performance [24]. Variations of division ratios such as 50:50, 60:40, 70:30, and 80:20 are used to analyze the impact of the amount of training data on model performance. Larger ratios, such as 80:20, allow the model to learn more, while balanced ratios, such as 50:50, provide a larger proportion of test data for more accurate evaluation.

Proper dataset division prevents overfitting, where the model adapts too much to the training data [25], and ensures that the model learns effectively [26]. A proper split between training and testing data also helps the model generalize better to unseen data. This ratio variation helps to analyze the balance between the amount of training data used and the accuracy achieved by the model. By adjusting the ratio, researchers can find an optimal balance between learning from data and ensuring the model's performance on new, unseen data.

Evaluation Matrix

Evaluate the performance of the model in recognizing Chinese Sign Language (CSL) using the main evaluation metric, Mean Average Precision (mAP), with a focus on mAP@50. mAP is a metric used to measure the overall performance of object detection [27], including object detection and segmentation tasks. mAP measures the accuracy of the model in detecting objects at various Intersection over Union (IoU) levels. Meanwhile, mAP50 (Mask) specifically measures how well the model detects and depicts the shape of objects in detail with a mask [28]. This metric is calculated by calculating the Average Precision (AP) for each class, which measures the area under the precision-recall curve.

a) IoU (Intersection over Union)
IoU measures the similarity between two sets, namely between model predictions and ground truth. The IoU value is calculated using the formula:

$$IoU(A,B) = \frac{|A \cap B|}{|A \cap B|}$$

b) Precision dan Recall
Precision and recall are used to measure the effectiveness of the model in detecting objects, calculated using the following formula:

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positive}$$

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negative}$$

c) Average Precision (AP)
Average Precision is calculated by ranking predictions based on the model's confidence score and calculating the area under the precision-recall curve. The formula for calculating AP for one class is:

$$AP = \sum_{n=1}^{N} (Recall_n - Recall_{n-1}) \times Precision_n$$

d) Mean Average Precision (mAP)

mAP is calculated by averaging the AP value for all classes. In this study, mAP@50 is calculated with an IoU threshold of 0.5, which means that the prediction will be considered correct if the IoU between the prediction and the ground truth is greater than 0.5.

$$mAP = \frac{AP_{IoU=0.5} + AP_{IoU=0.55} + \cdots + AP_{IoU=0.95}}{k}$$

e) F1-Score

The F1-score is a metric that combines precision and recall into a single number to provide a more balanced picture of model performance. The F1-score is calculated using the formula:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

The model evaluation was carried out using metrics to analyze the accuracy of hand movement detection in Chinese Sign Language and the model's ability to distinguish gestures. mAP@50 and mAP50-95 provide an overview of the model's performance in various IoU thresholds, which is important for assessing object detection and segmentation. The F1-score is used to assess the balance of precision and recall in gesture recognition, measuring the extent to which the model accurately and completely detects hand movements.

3. Results and Discussion

Environment and Parameters

Model training and inference were conducted using Google Colab with a Tesla T4 GPU (16GB VRAM), which provided the necessary computational power for training the YOLOv8 and YOLOv8-SEG models. The Chinese Sign Language (CSL) image dataset was trained with a resolution of 640 x 640 pixels. The training process was set to 100 epochs, with early stopping applied at 50 epochs to prevent overfitting and improve the model’s ability to generalize. The training process utilized the default YOLOv8 settings for all parameters, ensuring consistency with standard practices in object detection and segmentation tasks. The use of a Tesla T4 GPU significantly accelerated the training and inference times, which was essential given the size of the dataset and the complexity of the models.

The chosen resolution of 640 x 640 pixels for the images was based on the balance between computational efficiency and maintaining sufficient detail for the model to learn meaningful features from the images. This resolution is commonly used in many image recognition tasks, allowing for faster processing while still providing enough information for accurate object detection. By using early stopping at 50 epochs, the training process was optimized to ensure the model did not overfit to the training data, thus improving its potential for real-world applications. The entire setup on Google Colab, along with the Tesla T4 GPU, ensured that the models could be trained efficiently, making it possible to conduct multiple experiments within a reasonable timeframe.

Training Method

The YOLOv8 and YOLOv8-seg training methods are focused on the detection and segmentation of hand movements in Chinese sign language. Training is carried out using a processed and labeled dataset, with YOLOv8 for object detection and YOLOv8-seg for hand area segmentation. The process involves parameter selection, such as the number of epochs, and optimization techniques for learning efficiency. The model is evaluated periodically to prevent overfitting, using metrics such as mAP, precision, recall, and F1-score. Testing is carried out to measure the model's ability to recognize and classify hand movements in varying conditions, including data not seen during training.

Results

Model	Skenario	Precision	Recall	mAP50	mAP50-95	F1-score
Non-Segmentation	50:50	0.98	0.973	0.994	0.854	0.97
Non-Segmentation	60:40	0.982	0.993	**0.995**	0.868	**0.99**
Non-Segmentation	70:30	0.98	**0.991**	**0.995**	**0.88**	0.98
Non-Segmentation	80:20	**0.987**	0.972	0.991	**0.88**	0.98
Segmentation	50:50	0.838	0.699	0.817	0.806	0.71

Model	Skenario	Precision	Recall	mAP50	mAP50-95	F1-score
Segmentation	60:40	0.756	0.779	0.874	0.863	0.74
Segmentation	70:30	0.772	0.782	0.855	0.842	0.75
Segmentation	80:20	**0.859**	**0.834**	**0.901**	**0.896**	**0.82**

Table 3.4 shows the results of the model evaluation based on two scenarios: Non-Segmentation and Segmentation with various dataset division ratios. The Non-Segmentation model shows excellent performance in all evaluation metrics, with Precision and Recall reaching 0.98 at a 50:50 ratio and continuing to increase to 0.987 at an 80:20 ratio. The mAP50 and mAP50-95 metrics for non-segmentation also show high results, indicating excellent accuracy in detecting objects at stricter IoU thresholds. The F1-score of the non-segmentation model is also very stable, reaching 0.97 at a 50:50 ratio, which indicates an excellent balance between precision and recall.

Meanwhile, the segmentation model shows lower performance than non-segmentation, especially at a smaller data division ratio (50:50). Precision and Recall in segmentation are lower, with a Precision value at a 50:50 ratio of only 0.838 and Recall of 0.699. However, as the training data ratio increases, the results improve, with the F1-score reaching 0.82 at a ratio of 80:20. Although the mAP50 and mAP50-95 values are lower than for non-segmentation, the segmentation model shows potential in detecting objects in more detail, although it requires more training data to achieve optimal results.

In general, segmentation models tend to have lower accuracy compared to non-segmentation models. Segmentation has a more complex task [29] compared to non-segmentation, where the model must not only detect objects but also separate and mark object areas with high precision. This task increases the difficulty in handling noise, label inaccuracy, and object variation in images, which can reduce model accuracy. In addition, segmentation models are more prone to overfitting because they focus on small features such as object contours and require more intensive data augmentation. Even if the number of data, epochs, and other parameters are the same, segmentation models require more training data and computing time to achieve optimal results, which explains why segmentation models have lower accuracy on metrics such as Precision, Recall, and mAP compared to non-segmentation.

Testing

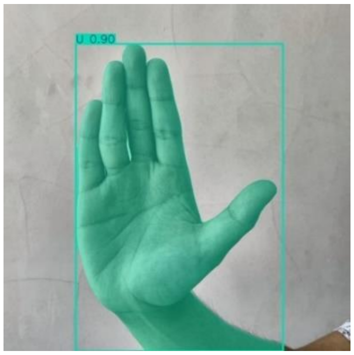

Figure 6. Example of Segmentation Test Results

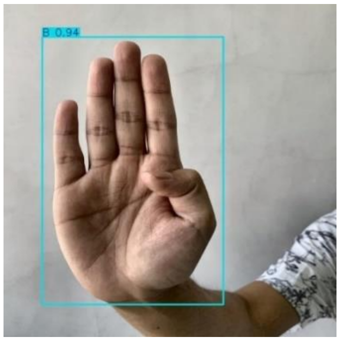

Figure 7. Example of Non-Segmentation Test Results

Testing was conducted to evaluate the performance of the Segmentation and Non-Segmentation models in detecting hands in images representing Chinese Sign Language gestures. The first image, tested with the Segmentation model, resulted in a confidence score of 0.90. Although lower than the Non-Segmentation model, the segmentation model offers more in-depth detection by separating objects from the background and identifying object areas in more detail. Meanwhile, the second image, tested with the Non-Segmentation model, resulted in a higher confidence score of 0.94. The Non-Segmentation model is more efficient at detecting the object as a whole using a bounding box, but does not separate the object in detail. Although the Non-Segmentation model has a higher score, the Segmentation model provides an advantage in more in-depth detection accuracy, although it requires more computing time and resources.

Although this study shows excellent accuracy in the Non-Segmentation and Segmentation models, there are several limitations to note. This study is limited to the recognition of static hand movements in Chinese Sign Language (CSL), where the model is only trained using images of hands with a white background. This causes limitations in detecting dynamic hand movements that involve continuous changes in position. In addition, recognition of images with varying backgrounds or uncontrolled lighting conditions has not been tested, which can affect the accuracy of the model in real-world situations. The results of this test show that the choice of model depends on the needs of the application, whether it prioritizes speed of detection or accuracy in detecting object details.

4. Conclusion

This study compares the performance of Non-Segmentation and Segmentation models in the recognition of Chinese Sign Language (CSL) using YOLOv8 and instance segmentation. The Non-Segmentation model shows excellent performance in all evaluation metrics, with high Precision, Recall, mAP50, and F1-score, especially at larger data division ratios, such as 80:20. These results show that the Non-Segmentation model is more efficient in detecting objects as a whole, with high accuracy in various IoU threshold conditions. On the other hand, the Segmentation model shows lower performance, especially at a 50:50 ratio, but its performance improves as the training data increases. Although the Precision and Recall values are lower, the Segmentation model shows an improvement in the F1-score which reaches 0.82 at a ratio of 80:20, indicating its ability to detect objects in more detail.

Overall, although Non-Segmentation has an advantage in terms of detection speed and accuracy, the Segmentation model offers more potential in detecting objects in greater depth and with high precision, although it requires more training data and computing time. These results show that the choice of model depends on the needs of the application, whether it prioritizes detection speed or object detail accuracy.

Acknowledgments

The completion of this research and the compilation of the related journal would not have been possible without the guidance, support, and encouragement from various parties. Therefore, the author would like to express his deepest gratitude to the Department of Electrical Engineering & Informatics and the Supervisor who provided valuable direction and support during the research process

References:

[1] K. Emmorey, “Ten Things You Should Know About Sign Languages,” *Curr Dir Psychol Sci*, vol. 32, no. 5, pp. 387–394, Oct. 2023, doi: 10.1177/09637214231173071.

[2] A. N. Handayani, M. I. Akbar, H. Ar-Rosyid, M. Ilham, R. A. Asmara, and O. Fukuda, “Design of SIBI Sign Language Recognition Using Artificial Neural Network Backpropagation,” in *2022 2nd International Conference on Intelligent Cybernetics Technology & Applications (ICICyTA)*, IEEE, Dec. 2022, pp. 192–197. doi: 10.1109/ICICyTA57421.2022.10038205.

[3] T. N. Fitria, “The Use Of Sign Language As A Media For Delivering Information On National Television News Broadcasts,” *ELP (Journal of English Language Pedagogy)*, vol. 9, no. 1, pp. 118–131, Jan. 2024, doi: 10.36665/elp.v9i1.764.

[4] C. Jo, “Overview of Chinese Sign Language,” *Chinese Language and Literature*, Dec. 2023, doi: 10.57237/j.cll.2023.05.002.

[5] X. Jiang, S. C. Satapathy, L. Yang, S.-H. Wang, and Y.-D. Zhang, "A Survey on Artificial Intelligence in Chinese Sign Language Recognition," *Arab J Sci Eng*, vol. 45, no. 12, pp. 9859–9894, Dec. 2020, doi: 10.1007/s13369-020-04758-2.

[6] H. Wang, Y. Chen, Z. Yang, L. Zhu, Y. Zhao, and T. Tian, "Estimation and projection of the burden of hearing loss in China: findings from the Global Burden of Disease Study 2019," *Public Health*, vol. 228, pp. 119–127, Mar. 2024, doi: 10.1016/j.puhe.2024.01.004.

[7] X. Chen, L. Su, J. Zhao, K. Qiu, N. Jiang, and G. Zhai, "Sign Language Gesture Recognition and Classification Based on Event Camera with Spiking Neural Networks," *Electronics (Basel)*, vol. 12, no. 4, p. 786, Feb. 2023, doi: 10.3390/electronics12040786.

[8] D. Wiryany, S. Natasha, R. Kurniawan, J. I. Komunikasi, and M. Bandung, "Perkembangan Teknologi Informasi Dan Komunikasi Terhadap Perubahan Sistem Komunikasi Indonesia," 2022.

[9] - Ridwang, A. A. Ilham, I. Nurtanio, and - Syafaruddin, "Dynamic Sign Language Recognition Using Mediapipe Library and Modified LSTM Method," *Int J Adv Sci Eng Inf Technol*, vol. 13, no. 6, pp. 2171–2180, Dec. 2023, doi: 10.18517/ijaseit.13.6.19401.

[10] 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, 2023, doi: 10.14569/IJACSA.2023.0140435.

[11] Mr. J. Rajasekhar, "Understanding Yolo: Real-Time Object Detection Explained," *Interantional Journal Of Scientific Research In Engineering And Management*, vol. 08, no. 07, pp. 1–9, Jul. 2024, doi: 10.55041/IJSREM36359.

[12] S. Tyagi, P. Upadhyay, H. Fatima, S. Jain, I. Avinash, and K. Sharma, "American Sign Language Detection using YOLOv5 and YOLOv8," 2023, doi: 10.21203/rs.3.rs-3126918/v1.

[13] S. Daniels, N. Suciati, and C. Fathichah, "Indonesian Sign Language Recognition using YOLO Method," *IOP Conf Ser Mater Sci Eng*, vol. 1077, no. 1, p. 012029, Feb. 2021, doi: 10.1088/1757-899X/1077/1/012029.

[14] Z. Q. Zhao, P. Zheng, S. T. Xu, and X. Wu, "Object Detection with Deep Learning: A Review," Nov. 01, 2019, *Institute of Electrical and Electronics Engineers Inc.* doi: 10.1109/TNNLS.2018.2876865.

[15] J. C. Caicedo *et al.*, "Data-analysis strategies for image-based cell profiling," *Nat Methods*, vol. 14, no. 9, pp. 849–863, Sep. 2017, doi: 10.1038/nmeth.4397.

[16] K. Maharana, S. Mondal, and B. Nemade, "A review: Data pre-processing and data augmentation techniques," *Global Transitions Proceedings*, vol. 3, no. 1, pp. 91–99, Jun. 2022, doi: 10.1016/j.gltp.2022.04.020.

[17] M. Guo *et al.*, "Normal Workflow and Key Strategies for Data Cleaning Toward Real-World Data: Viewpoint," *Interact J Med Res*, vol. 12, p. e44310, Sep. 2023, doi: 10.2196/44310.

[18] A. N. Handayani, S. Amaliya, M. I. Akbar, M. Z. Wiryawan, Y. W. Liang, and W. C. Kurniawan, "Hand Keypoint-Based CNN for SIBI Sign Language Recognition," *International Journal of Robotics and Control Systems*, vol. 5, no. 2, pp. 813–829, 2025, doi: 10.31763/ijrcs.v5i2.1745.

[19] A. N. Handayani, M. I. Akbar, H. A. Rosyid, M. Arifianto, O. Fukuda, and R. A. Asmara, "K-NN for Hand Keypoints Detection Based on SIBI Standard," in *2024 International Conference on Electrical and Information Technology (IEIT)*, IEEE, Sep. 2024, pp. 72–77. doi: 10.1109/IEIT64341.2024.10763365.

[20] A. Onan, "SRL-ACO: A text augmentation framework based on semantic role labeling and ant colony optimization," *Journal of King Saud University - Computer and Information Sciences*, vol. 35, no. 7, p. 101611, Jul. 2023, doi: 10.1016/j.jksuci.2023.101611.

[21] R. Wulanningrum, A. N. Handayani, and A. P. Wibawa, "Perbandingan Instance Segmentation Image Pada Yolo8," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 11, no. 4, pp. 753–760, Aug. 2024, doi: 10.25126/jtiik.1148288.

[22] G. Li, L. Zhou, Q. Tong, Y. Ding, X. Qi, and H. Liu, "A Data Augmentation Approach to Sentiment Analysis of MOOC Reviews," *International Journal of Advanced Computer Science and Applications*, vol. 15, no. 8, 2024, doi: 10.14569/IJACSA.2024.01508122.

[23] A. N. Handayani, T. Andriyanto, D. F. Azizah, M. Z. Wiryawan, and H. A. Rosyid, “Comparison of ResNet-50 and EfficientNet-B0 Method for Classification of Indonesian Sign Language System (SIBI) Using Multi Background Dataset,” in *2024 International Conference on Computer Engineering, Network, and Intelligent Multimedia (CENIM)*, IEEE, Nov. 2024, pp. 1–6. doi: 10.1109/CENIM64038.2024.10882836.

[24] V. R. Joseph, “Optimal ratio for data splitting,” *Statistical Analysis and Data Mining: The ASA Data Science Journal*, vol. 15, no. 4, pp. 531–538, Aug. 2022, doi: 10.1002/sam.11583.

[25] Q. H. Nguyen *et al.*, “Influence of Data Splitting on Performance of Machine Learning Models in Prediction of Shear Strength of Soil,” *Math Probl Eng*, vol. 2021, pp. 1–15, Feb. 2021, doi: 10.1155/2021/4832864.

[26] A. Nurhopipah and U. Hasanah, “Dataset Splitting Techniques Comparison For Face Classification on CCTV Images,” *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 14, no. 4, p. 341, Oct. 2020, doi: 10.22146/ijccs.58092.

[27] C.-J. Zhang *et al.*, “Evaluation of the YOLO models for discrimination of the alfalfa pollinating bee species,” *J Asia Pac Entomol*, vol. 27, no. 1, p. 102195, Mar. 2024, doi: 10.1016/j.aspen.2023.102195.

[28] J. Chen, C. Ji, J. Zhang, Q. Feng, Y. Li, and B. Ma, “A method for multi-target segmentation of bud-stage apple trees based on improved YOLOv8,” *Comput Electron Agric*, vol. 220, p. 108876, May 2024, doi: 10.1016/j.compag.2024.108876.

[29] E. Lee, B. Park, M.-H. Jeon, H. Jang, A. Kim, and S. Lee, “Data augmentation using image translation for underwater sonar image segmentation,” *PLoS One*, vol. 17, no. 8, p. e0272602, Aug. 2022, doi: 10.1371/journal.pone.0272602.