ISSN: 2963-2536 (media online)

Proceedings of Malikussaleh International Conference on Multidisciplinary Studies (MICoMS)

Organized by center of Excellence of Technology Natural Polymer and Recycle Plastics

International Conference Proceedings 00001 (2024) DOI: https://doi.org/10.29103/micoms.v4.2024

Indonesian Sign Language (BISINDO) Alphabet Detection System Using YOLO (You Only Look Once) Algorithm

Andra Munandar [1], Zara Yunizar [2] and Sujacka Retno [3,]

[1] Department of Informatic, Universitas Malikussaleh, Bukit Indah, Lhokseumawe, 24353, Indonesia, andra.200170236@mhs.unimal.ac.id

[2] Department of Informatic, Universitas Malikussaleh, Bukit Indah, Lhokseumawe, 24353, Indonesia, zarayunizar@unimal.ac.id

[3] Department of Informatic, Universitas Malikussaleh, Bukit Indah, Lhokseumawe, 24353, Indonesia, sujacka@unimal.ac.id

Correspondence: andra.200170236@mhs.unimal.ac.id

Abstract: This research aims to develop an Indonesian Sign Language (BISINDO) alphabet detection system using the YOLOv5 algorithm, an efficient and fast deep learning-based object detection model. The dataset used consists of BISINDO alphabet images enriched through data augmentation techniques such as rotation, flipping, and brightness adjustment. The evaluation results show that the YOLOv5s model achieved very good performance, with an average precision of 85.2%, recall of 89.3%, F1-score of 87.2%, and mean average precision (mAP) of 87.1%. The confusion matrix also indicates the model's ability to differentiate each BISINDO alphabet with high accuracy. The training data testing showed the model successfully achieved consistent decreases in all loss components, such as a decrease in train box loss from 0.06 to 0.015, and validation loss converging towards 0.002 for object loss and class loss. The real time testing also shows that the YOLOv5-based BISINDO alphabet detection system can perform well and consistently, indicating the practical application potential of this system to facilitate communication between people with hearing/speech disabilities and the general public. Overall, this research has resulted in an accurate and real-time implementable BISINDO alphabet recognition system.

Keywords: Sign Language, YOLO, Object Detection , BISINDO

1. Introduction

reaching 212,240 individuals as of March 2022 [1]. One group of individuals with disabilities comprises those who are deaf and speech-impaired. These individuals often face communication barriers due to their limited ability to hear and speak [2]. Therefore, they rely on Sign Language as their primary means of communication. Sign language itself utilizes hand gestures, facial expressions, and body language as substitutes for verbal communication [3].

One of the most commonly used sign languages in Indonesia is BISINDO (Indonesian Sign Language), which was developed within the deaf community to facilitate everyday communication [4]. However, despite its practicality, sign language remains underutilized in social life due to various constraints, particularly in terms of understanding among the general public and the lack of knowledge and awareness about sign language [5]. This often results in difficulties for deaf and speech-impaired individuals when interacting with the general public, hindering their participation in various aspects of social, economic, and educational life [6]. An example in the educational aspect can be seen in a survey conducted by the University of Indonesia's Disability Service Center (2022), which revealed that 78% of deaf students experience difficulties interacting in the learning process due to minimal understanding of sign language among lecturers and students.

On the other hand, the advancing technology in pattern recognition and object detection has opened new opportunities to address communication barriers for people with disabilities, including the deaf and speech-impaired [7]. One technology currently widely used is the You Only Look Once (YOLO) algorithm, known for its ability to detect objects quickly and accurately in real-time. This algorithm is designed to recognize various objects in images or videos with high efficiency, making it highly potential for application in sign language recognition [8]. Based on these considerations, this research proposes the development of a BISINDO alphabet recognition system using YOLOv5, which is expected to serve as an innovative solution in facilitating communication between people with disabilities and the general public. The main focus of this research is the recognition of the BISINDO alphabet, which forms the foundation for word and sentence formation in sign language. This system aims to provide a strong foundation for further development of vocabulary and sentence detection, as well as assist in the process of understanding BISINDO for both people with disabilities and those who wish to learn sign language.

2. Materials and Methods

2.1. Indonesian Sign Language

Indonesian Sign Language (BISINDO) is a sign language developed by the Indonesian deaf community through GERKATIN [9]. Unlike SIBI, which adapts ASL, BISINDO has evolved naturally in accordance with Indonesian culture, utilizing simpler hand gestures, facial expressions, and body language, making it more effective for daily communication [10].

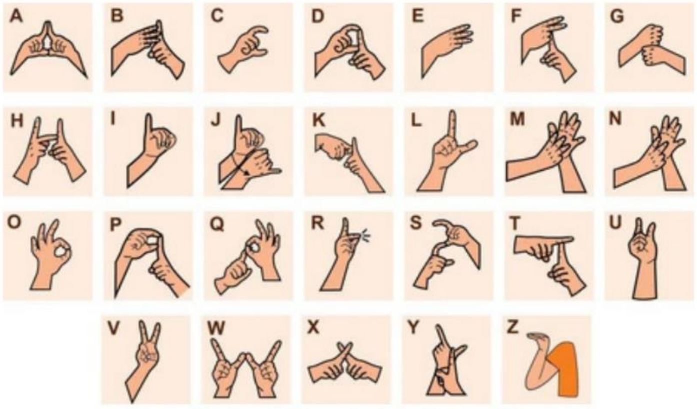


Figure 1. Alphabets in BISINDO

2.2. YOLO

You Only Look Once (YOLO) is an object detection algorithm that employs a single-stage detection approach to process images in a single pass [11]. YOLOv5, as the latest version, introduces improvements through Cross Stage Partial Network (CSP) and Path Aggregation Network (PAN), which enhance detection accuracy [12].

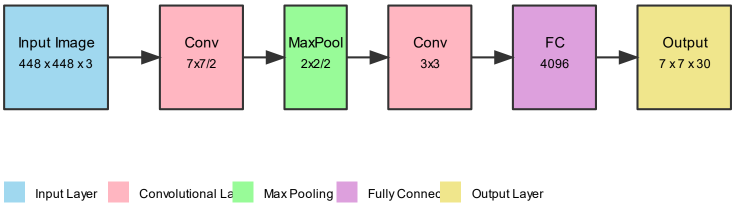


Figure 2. YOLO Architecture

2.3. *Previous Research*

Previous research has explored sign language recognition using computer vision and deep learning, such as a SIBI recognition system with YOLOv5 achieving 77% accuracy, and a hand detection system for Indonesian sign language using CNN and YOLO with 89% accuracy [13].

This current research aims to build upon these prior efforts by developing a real-time recognition system focused on BISINDO (Indonesian Sign Language) using the efficient and accurate YOLO algorithm. The goal is to address limitations of previous work, including the lack of high accuracy real-time systems and the need for expanded BISINDO datasets.

2.3. *Data Collection*

The dataset used in this research consists of a collection of BISINDO alphabet images (A-Z) obtained from the Kaggle platform. The dataset comprises 11,500 image data divided into 26 alphabet classes. These image data display various hand gesture variations representing the alphabets in BISINDO.

The dataset was enriched through data augmentation processes using several techniques such as image rotation, brightness adjustment, horizontal flipping, noise addition, and scaling. Subsequently, the dataset was labeled using the Roboflow application to create bounding boxes and annotations for each image. The labeling process resulted in files in YOLO (.txt) format, which contain information about the object coordinates and classes for each image to be used in the training process. An example of the dataset can be seen in **Figure 3** below :

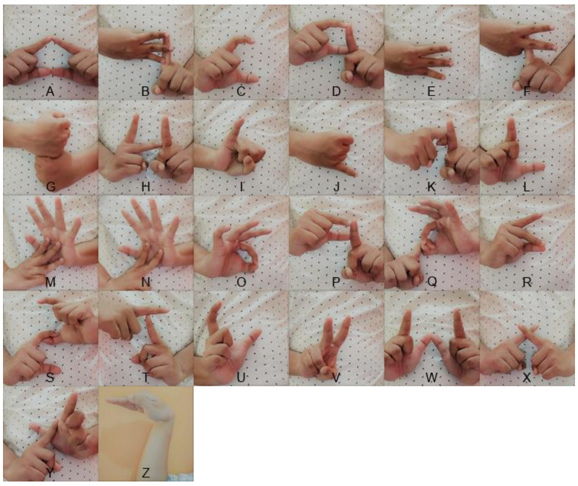


Figure 3. Dataset Image

2.3. *Data Preprocessing*

In this stage, several data augmentation techniques were employed to enrich the dataset's variation. The data augmentation phase is an essential step to modify and expand the dataset so that the model can recognize a wider range of BISINDO alphabet gesture variations more effectively.

2.3.1. Data Resize

All images in the dataset are resized to 640 x 640 pixels in the resize stage. This process is important to uniform the input size that will be processed by the YOLOv5 model. With a consistent size, the model training process becomes more efficient and the recognition results are optimised, ensuring compatibility across various devices, enhancing data preprocessing consistency, and supporting robust model performance.

2.3.2. Data Rotation

The rotation stage is performed by rotating the dataset image in the range of -18 to 18 degrees. This process generates a variety of new viewpoints that help the model to be more adaptive in recognising gestures. This angle variation is important considering that in real use, the angle of image capture may vary.

2.3.3. Data Flip

In the data flip stage, the images in the dataset are mirrored horizontally to add variety to the data. This technique helps the model recognise BISINDO alphabetic gestures from various viewpoints, considering that in practice users may show gestures from different directions.

2.3.4. Brightness Adjustment

Brightness Adjustment At this stage, the image brightness level is modified with a range of -20% to +20%. This adjustment aims to improve the model's ability to recognise gestures in different lighting conditions. This brightness variation helps the model to be more robust to changes in lighting conditions during implementation in the real environment.

2.3.5. Data Annotation

Each image in the dataset was then annotated using the roboflow tool. This process involves placing a bounding box around the hand area and labelling it according to the alphabet shown. The annotation results are saved in a .txt format file that contains the bounding box coordinate information and the class label. This annotation file will be used as a guide for the model in the training process to learn the characteristics of each BISINDO alphabet.

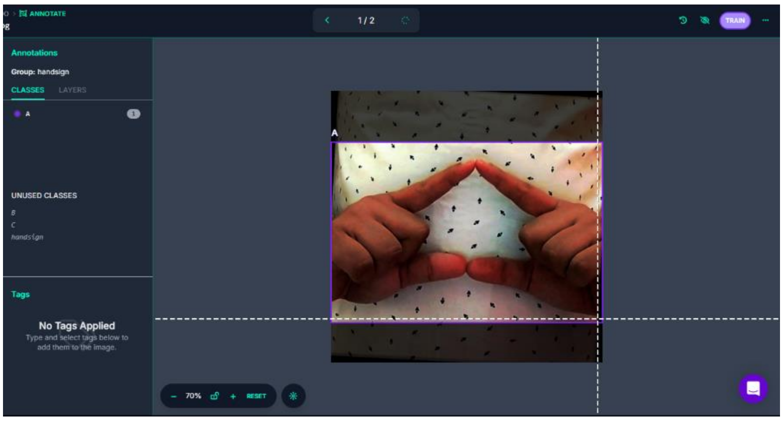


Figure 4. Data Annotation

2.4. *Model Implementation*

In this research, YOLOv5 was used to detect BISINDO alphabets. YOLOv5 is a deep learning model specifically designed for object detection using a single-stage detector approach, where the detection and classification processes are performed in a single processing step. The model has a structure consisting of a CSPDarknet backbone for feature extraction, a PANet neck for feature aggregation, and a head for object detection.

The YOLOv5s model architecture was chosen as the base model because it is a lightweight variant of the YOLOv5 architecture. This model employs a Cross Stage Partial Network (CSP), which allows for more efficient feature extraction by reducing the number of parameters and computations required. The structure also enables the model to detect objects at various scales through the Feature Pyramid Network (FPN).

The YOLOv5 model's detection mechanism divides the input image into grid cells, where each cell is responsible for detecting objects whose centers fall within that cell. Each cell predicts the bounding box, confidence score, and class probability. In the context of this research, the predicted classes are the 26 BISINDO alphabets (A-Z).

2.5. *Evaluation*

The evaluation stage was conducted to assess the performance of the BISINDO alphabet recognition system developed using the YOLOv5 algorithm. The evaluation process involved testing the model on the test dataset and calculating relevant evaluation metrics. These metrics included the accuracy of bounding box detection, prediction of object presence, prediction of object class, as well as visualization of model performance through the confusion matrix. The use of these evaluation metrics allowed for a comprehensive assessment of the system's capabilities in accurately detecting and classifying the BISINDO alphabets.

2.5.1. Box Loss

Box Loss measures how well the model can predict the bounding box coordinates of the detected object. These coordinates include the centre position (x, y) as well as the width (w) and height (h) of the bounding box.

$$Box\ Loss = \sum(x_{pred} - x_{true})^2 + \sum(y_{pred} - y_{true})^2 + \sum(w_{pred} - w_{true})^2 + \sum(h_{pred} - h_{true})^2 \quad (1)$$

Where $xpred, ypred, wpred, hpred$ are the predicted values of the model, and x_true, y_true, w_true, h_true are the true values of the bounding box coordinates. A lower Box Loss value indicates that the model has a better ability to accurately predict the bounding box.

2.5.2. Object Loss

Obj Loss measures the ability of the model to predict the presence of objects in each grid cell. The model must be able to distinguish between grid cells that contain objects and those that do not.

$$Obj\ Loss = \sum(obj_{pred} - obj_{true})^2 \quad (2)$$

Where $xpred, ypred, wpred, hpred$ are the predicted values of the model, and x_true, y_true, w_true, h_true are the true values of the bounding box coordinates. A lower Box Loss value indicates that the model has a better ability to accurately predict the bounding box.

2.5.3. Class Loss

Class Loss measures the accuracy of the model in predicting the class of the detected object. For BISINDO alphabet recognition, the predicted class includes 26 letters from A to Z.

$$Class\ Loss = \sum(class_{pred} - class_{true})^2 \quad (3)$$

Where $class_{pred}$ is the prediction probability distribution for each class, and $class_{true}$ a is the true label of the object class.

2.5.3. Confussion Matrix

The next stage of system testing uses a Confusion Matrix to evaluate classification results. The Confusion Matrix records the number of test data that are correctly and incorrectly classified. In this test, a Multi-Class Confusion Matrix is used to compare prediction results with actual data. From these results, the values of Accuracy, Precision, and Recall are calculated. Accuracy measures the model's accuracy level in correctly classifying data based on the correspondence between predicted and actual values. Below are the equations for calculating accuracy, precision, and recall.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \qquad Precision = \frac{TP}{TP+FP} \qquad Recall = \frac{TP}{TP+FN} \tag{4}$$

3. Results and Discussion

3.1. Model Implementation

The model used in this research is YOLOv5s, a deep learning-based object detection model designed for efficiency and speed. The model training is done using Google Colab platform with NVIDIA T4 GPU runtime to speed up the computational process. The dataset used consists of images of the Indonesian Sign Language (BISINDO) alphabet from A to Z, with a total of 150 images for each alphabet, making a total of 3,900 images. The model training process was conducted for 50 epochs, with the aim of optimising detection accuracy without causing overfitting. The implementation involved data augmentation, division of the dataset into training and validation data, and hyperparameter settings to ensure optimal model performance.

3.2. Analysis of Model Performance

Performance analysis of the YOLOv5s model was conducted to evaluate its ability to detect Indonesian Sign Language (BISINDO) alphabets based on several standard evaluation metrics including Precision, Recall, Mean Average Precision (mAP), Intersection over Union (IoU), and Loss. Additionally, a Confusion Matrix was used to identify prediction distributions for each category. This performance analysis includes data from both model training and testing processes.

The performance evaluation results in Figure 6 show satisfactory performance with average precision of 0.852 (85.2%), recall of 0.893 (89.3%), and F1-score of 0.872 (87.2%). The model demonstrates good stability with a precision standard deviation of 0.02 across classes, where the majority of classes achieved precision and recall above 0.85. Overall, the model achieved a Mean Average Precision (mAP) of 0.871 and balanced accuracy of 0.882, indicating that the YOLOv5 model has successfully learned the characteristics of each BISINDO alphabet.

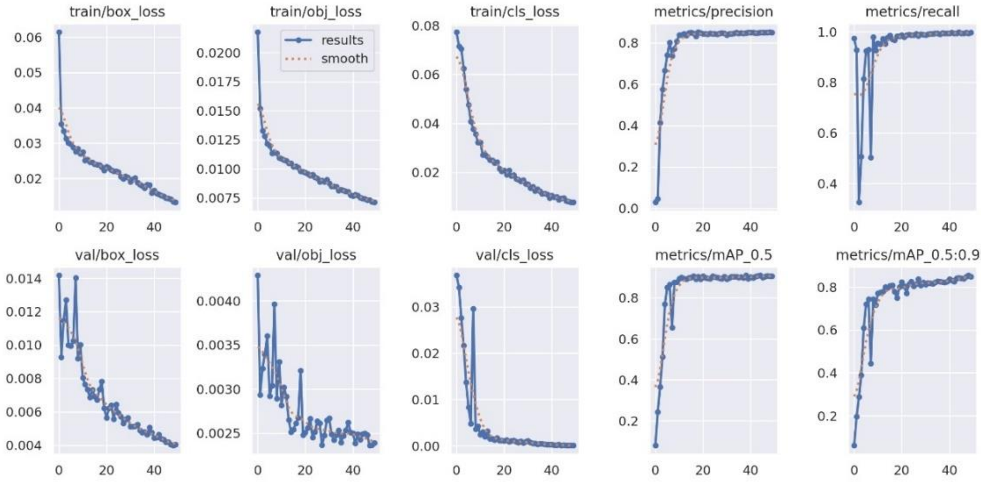


Figure 5. Evaluations Model

Figure 7 shows the model's performance across several testing metrics for easier analysis. The graphs indicate consistent decreases in all loss components. Train box loss decreased from 0.06 to 0.015, while validation box loss stabilized at 0.004. Object loss and class loss showed similar downward trends, with validation loss converging near 0.002. Performance metrics showed significant improvements, with precision and recall achieving stable values above 0.8 after epoch 20. Mean Average Precision (mAP@0.5) reached 0.85 and mAP@0.5:0.95 stabilized at 0.75, indicating the model successfully achieved optimal performance.

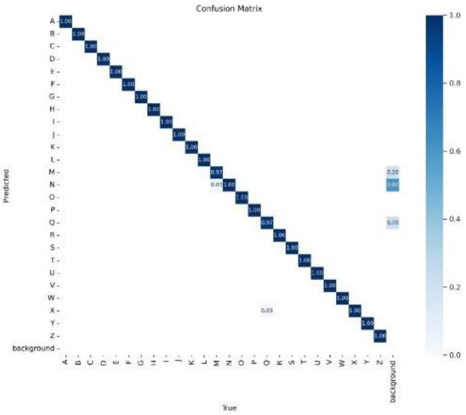


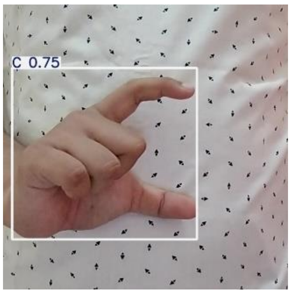
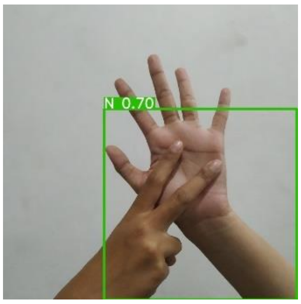
Figure 6. Confussion Matrix

Additionally, the Confusion Matrix shows excellent classification performance with most classes achieving perfect accuracy (1.00). Minor notes include letters M and Q having 0.97 accuracy, and slight classification errors between letters N and X at 0.03. Overall, the matrix demonstrates the model's excellent ability to distinguish each BISINDO alphabet, with minimal classification errors.

3.2. Model Analysis Result

The trained YOLOv5s model can effectively detect Indonesian Sign Language (BISINDO) alphabets from A to Z with high performance. This detection process includes identifying letter positions (bounding box) in images and their corresponding letter classification.

Table 1. Test Data Result

No	Test Image	Actual Label	Predicted Label
1		C	C
2		N	N

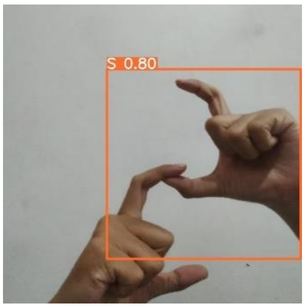
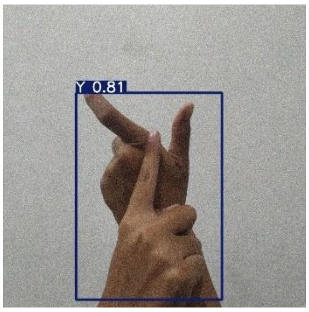
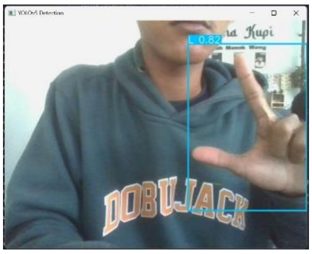
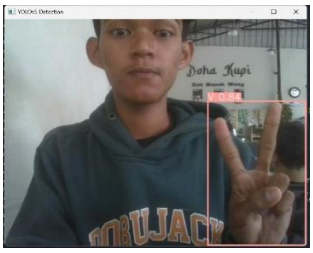
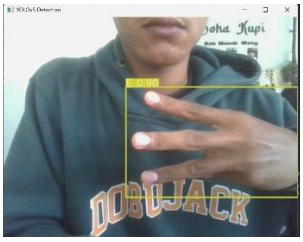
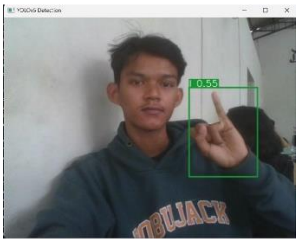
3		S	S
4		Y	Y

Table 2. Real-Time Prediction Result

No	Test Image	Actual Label	Predicted Label
1		L	L
2		V	V
3		E	E

4		I	I

Based on the modeling results in tables 1 and 2, testing using training data successfully classified hand gestures with consistent accuracy. In real-time testing, the system also demonstrated good performance in detecting hand gestures directly. This success in real-time detection indicates that the model has good generalization capability and can adapt to varying lighting conditions, angles, and hand positions in real-world usage.

The consistency between actual and predicted labels in both types of testing shows that the system has good reliability in detection. This serves as a positive indicator that the implemented YOLO model can be used as a practical solution for Indonesian sign language alphabet recognition.

4. Conclusions

a) This research developed a BISINDO (Indonesian Sign Language) alphabet detection system using the YOLOv5 algorithm, an efficient and fast deep learning-based object detection model.

b) The research used a BISINDO alphabet image dataset enriched through data augmentation techniques including rotation, flipping, and brightness adjustments. This dataset was used to train and evaluate the YOLOv5s model.

c) Evaluation results show the YOLOv5s model achieved excellent performance, with average precision of 85.2%, recall of 89.3%, F1-score of 87.2%, and mean average precision (mAP) of 87.1%. The confusion matrix also indicated the model's ability to distinguish each BISINDO alphabet with high accuracy.

d) Testing on training data showed the model achieved consistent decreases in all loss components, with train box loss decreasing from 0.06 to 0.015, and validation loss converging near 0.002 for object loss and class loss.

e) Real-time testing also demonstrated that the YOLOv5-based BISINDO alphabet detection system can work effectively and consistently, indicating the potential practical application of this system to facilitate communication between deaf/mute individuals and the general public.

5. Referrence

[1] F. A. Setiawan, Z. A. Rohmah, and G. F. Laxmi, “Indonesian Sign Language (BISINDO) Alphabet Detection Using the You Only Look Once (YOLO) Algorithm Version 8,” in 2024 International Conference on Computer, Control, Informatics and its Applications (IC3INA), 2024, pp. 388–393.

[2] J. Emerson and P. Enderby, “Concerns of speech-impaired people and those communicating with them,” Health Soc. Care Community, vol. 8, no. 3, pp. 172–179, 2000.

[3] M. Zakariah, Y. A. Alotaibi, D. Koundal, Y. Guo, and M. Mamun Elahi, “Sign language recognition for Arabic alphabets using transfer learning technique,” Comput. Intell. Neurosci., vol. 2022, no. 1, p. 4567989, 2022.

[4] S. Isnaniah, T. Agustina, and F. Annisa, “The Use of Sign Language in Deaf Indonesian Classrooms in Surakarta,” KEMBARA J. Keilmuan Bahasa, Sastra, dan Pengajarannya, vol. 9, no. 2, pp. 468–481, 2023.

[5] M. De Meulder and H. Haualand, “Sign language interpreting services: A quick fix for inclusion?,” Transl. Interpret. Stud., vol. 16, no. 1, pp. 19–40, 2021.

[6] R. S. Rathod, Silent Voices, Stronger Bonds: Strategies for Communication with Hearing and Speech Impaired. Laxmi Book Publication, 2024

[7] A. Hussain, N. Saikia, and C. Dev, “Advancements in Indian Sign Language Recognition Systems: Enhancing Communication and Accessibility for the Deaf and Hearing Impaired,” Asian J. Electr. Sci., vol. 12, no. 2, pp. 37–49, 2023.

[8] M. Alaftekin, I. Pacal, and K. Cicek, “Real-time sign language recognition based on YOLO algorithm,” Neural Comput. Appl., vol. 36, no. 14, pp. 7609–7624, 2024.

[9] T. N. Fitria, “THE USE OF SIGN LANGUAGE AS A MEDIA FOR DELIVERING INFORMATION ON NATIONAL TELEVISION NEWS BROADCASTS,” ELP (Journal English Lang. Pedagog., vol. 9, no. 1, pp. 118–131, 2024.

[10] I. G. A. O. Aryananda and F. Samopa, “Comparison of the Accuracy of The Bahasa Isyarat Indonesia (BISINDO) Detection System Using CNN and RNN Algorithm for Implementation on Android,” MALCOM Indones. J. Mach. Learn. Comput. Sci., vol. 4, no. 3, pp. 1111–1119, 2024.

[11] T. Diwan, G. Anirudh, and J. V Tembhurne, “Object detection using YOLO: Challenges, architectural successors, datasets and applications,” Multimed. Tools Appl., vol. 82, no. 6, pp. 9243–9275, 2023.

[12] J. E. Gallagher and E. J. Oughton, “Surveying You Only Look Once (YOLO) Multispectral Object Detection Advancements, Applications And Challenges,” arXiv Prepr. arXiv2409.12977, 2024.

[13] S. Daniels, N. Suciati, and C. Fathichah, “Indonesian sign language recognition using yolo method,” in IOP Conference Series: Materials Science and Engineering, 2021, vol. 1077, no. 1, p. 12029.