Classification of Indonesian Sign Language (BISINDO) on Video using YOLO

1st Devana Gema Falesta
School of Computing
Telkom University
Bandung, Indonesia
devana@student.telkomuniversity.ac.id

2st Tjokorda Agung Budi Wirayuda
School of Computing
Telkom University
Bandung, Indonesia
cokagung@telkomuniversity.ac.id

3st Febryanti Sthevanie
School of Computing
Telkom University
Bandung, Indonesia
sthevanie@telkomuniversity.ac.id

***Abstract*— Indonesian Sign Language (BISINDO) is one type of sign language used by people with disabilities in Indonesia to communicate. However, there is still a gap between people with disabilities and non-disabled people in terms of communication, especially in the use of technology. This research aims to develop a video-based BISINDO classification system using the You Only Look Once version 11 (YOLOv11) model, which is expected to help bridge communication between people with disabilities and non-disabled people. The BISINDO+ dataset consists of 26 classes with 6,389 images aggregated from Kaggle, and a personal dataset created by the author with diverse backgrounds was developed in this research. The augmentation and hyperparameter tuning effectively increase model performance using accuracy, precision, recall, and mAP metrics. The results of this study show YOLOv11 achieved the best performance on the validation set, with a precision of 0.998, recall of 1.000, and mAP of 0.995, showing a precision difference of 0.7 and a recall difference of 0.4 compared to YOLOv8, making it slightly superior in object detection accuracy. Evaluation results on test data show YOLOv11 achieved a precision of 0.997, recall of 1.000, and mAP of 0.995. Real-time testing yielded 100% accuracy at a distance of 30 cm and decreased to 97% at a distance of 60 cm. The accuracy reached 100% on a plain background, while it declined to 92% on a patterned background due to visual interference.**

Keywords—BISINDO, YOLOv11, object detection, YOLOv8, Real-time

I. Introduction

Sign language is vital for people with hearing and speech disabilities. According to information from the Disability Information Management System of the Ministry of Health of the Republic of Indonesia, in March 2022, the number of persons with disabilities in Indonesia reached 212,240 individuals. This number has increased over the past two years, from 197,582 individuals in March 2020 to 207,604 individuals in March 2021. Meanwhile, the number of individuals with hearing and speech disabilities in March 2022 was 19,392, or 9.14% of the total number of persons with disabilities in Indonesia [1]. In daily life, individuals with hearing and speech disabilities face challenges in communicating with non-disabled individuals. A technology-based solution is urgently needed to assist communication between individuals with hearing and speech disabilities and non-disabled individuals, fostering a comfortable and equitable social life.

Language is an essential factor in communication that allows humans to interact with each other, share information, and express themselves. According to research by Arifah et al. [2], sign language is a communication tool for people with disabilities, especially the deaf and hard of hearing. Sign language has a unique grammatical structure, vocabulary, and rules like spoken language. BISINDO is one of the distinctive sign languages created by the Indonesian Deaf Welfare Movement (GERKATIN) and developed by the deaf community. Sign language is performed by utilizing hand movements and facial and body expressions to form certain symbols or gestures as a substitute for letters or words. Therefore, technology development to recognize sign language can be advanced by leveraging sophisticated gesture recognition processes, enabling seamless interpretation of hand movements and expressions to bridge communication gaps effectively.

YOLO has been proven to perform well in sign language detection, with consistently high accuracy. Several studies have shown that recent models, such as YOLOv8, YOLOv10, and YOLOv11, can achieve mean average precision (mAP) values of over 94% [3][4][5]. These results prove the effectiveness of YOLO in real-time gesture detection. However, some studies are limited to less diverse background variations; based on the sample datasets in these studies, the data variation tends to be monotonous, and there is no background testing. This can be problematic as it needs to reflect actual field conditions, which may have broader and more complex variations.

In this research, the author focuses on developing a BISINDO dataset that incorporates diverse background variations to support the development of a classification model capable of achieving high performance in sign language recognition, even in environments with non-uniform or dynamic backgrounds. The developed dataset is a combination of publicly available datasets and self-acquired data. YOLO-based deep learning, specifically YOLOv8, is utilized as the performance baseline, while YOLOv11 is employed to enhance efficiency regarding computational resource requirements. YOLOv8 has 11.2 million parameters, while YOLOv11 has only 9.4 million parameters but still shows excellent results. A study on YOLOv11 reveals that, despite its smaller model size, its performance remains robust, with improvements in certain aspects, demonstrating its ability to maintain high accuracy and efficiency while reducing computational resource demands [4].

The structure of this paper is organized as follows: Section 2 provides a comprehensive overview of the existing research on sign language classification. Section 3 details the methodologies employed, including the datasets and performance metrics for evaluating the models. Section 4 presents an analysis of the experimental results and a thorough discussion of the implications and insights derived from the findings. The final section concludes the paper and proposes potential avenues for future research and development.