CHAPTER I
INTRODUCTION

1.1 Background

Language is one of the fundamental aspects of human life. Language is an arbitrary system of sound symbols used by communities to cooperate, interact, and express themselves [1]. Through language, humans can convey messages, thoughts, and feelings to others, thereby fostering communication and cooperation among individuals [2]. Thus, language serves as the primary means of communication and plays a vital social role in society [3].

However, not everyone can communicate verbally. One group facing such barriers is people with hearing impairments (the deaf) [4]. Deafness is a condition of partial or total hearing loss that results in limitations in language use and speech [5]. In daily life, this group uses sign language as their primary means of communication, relying on hand movements, facial expressions, body language, and lip movements to convey meaning [6]. Sign language serves not only as a communication tool but also as the identity of the deaf community.

In Indonesia, there are two recognized sign language systems: the Indonesian Sign Language System (SIBI) and Indonesian Sign Language (BISINDO). SIBI is the official, government-standardized sign language used in special education schools, while BISINDO is more widely used in the daily lives of the deaf community [7]. Linguistically, sign language has its own distinct phonological, morphological, syntactic, and semantic systems that differ from spoken languages because it employs a visual-gestural modality [8]. This language is regarded as an effective non-verbal communication system as it relies on kinetic symbols in the form of hand movements and facial expressions [9].

Understanding of sign language remains limited among the general public [7]. This has driven the development of technologies capable of recognizing sign language through computer vision [10]. Computer vision is inspired by how the human eye recognizes objects [11]. One modern method for object detection is the Detection Transformer (DETR), which offers a transformer-based architecture and does not require anchor boxes or non-maximum suppression (NMS) because it can predict objects and bounding boxes directly [12].

Various studies have demonstrated the effectiveness of the Detection Transformer (DETR) in diverse scenarios. Study [13] used DETR to detect vehicle logos, achieving an AP50 of 0.952. Study [14] integrated DETR with EasyOCR to read numbers on analog kWh meters and achieved an mAP of 0.968. Study [15] applied DETR to detect explicit content in anime characters and achieved an AP50 of 0.875.

Although DETR is favored for its simple architecture and the elimination of manual components such as NMS, like most deep learning models, its performance remains highly dependent on the availability and diversity of training data [12]. To address the challenge of data scarcity, data augmentation approaches become crucial. Study [16] provides an interesting comparison in a limited-data scenario using YOLOv8 by applying on-the-fly augmentation methods, yielding an mAP of 0.674, a highly competitive result with only a slight difference compared to conventional methods achieving an mAP of 0.699. These findings indicate that on-the-fly augmentation can maintain model performance without the burden of storing physical data.

Based on this effectiveness, this study employs the Detection Transformer (DETR) by adopting on-the-fly augmentation techniques, a dynamic augmentation process performed during training without saving new image results. The variety of augmentations generated depends on the number of epochs, as each image may undergo different augmentations in every training iteration [17].

In addition to data augmentation, evaluating model performance is also crucial to determine how well the system can be implemented in real-world conditions. Although DETR has high computational complexity [12], this study incorporates frame-by-frame input from a camera to evaluate detection stability under real-world usage conditions, by reducing the number of transformer layers and object queries in the DETR architecture. Thus, this study tests the model not only on static images but also in real-world usage scenarios.

Based on this, the research hypothesis is that the application of DETR with on-the-fly augmentation is capable of producing accurate and stable BISINDO alphabet detection, both on static images and in real-world use. The web-based implementation is expected to provide an easily accessible practical solution, while

demonstrating the potential of this technology in supporting communication for the deaf community.

1.2 Research Question

This study's research questions provide guidance for implementing the DETR model to detect the Indonesian Sign Language (BISINDO) alphabet and words, as follows:

1. How can the DETR method combined with on-the-fly data augmentation be implemented to detect the Indonesian Sign Language (BISINDO) alphabet and words?
2. How can DETR's performance be evaluated using PyCOCOTools, considering metrics such as accuracy, mean Average Precision (mAP), and detection stability?
3. How can the DETR model be optimized or adapted to achieve near real-time detection?
4. How robust is the model to variations in input, such as lighting conditions, hand positions, and different users?
5. How can the DETR model be integrated into a web-based system that is interactive, accessible, and user-friendly for non-technical users?

1.3 Research Objectives

The objectives of this study are outlined as follows, providing clear guidance and direction for the research process:

1. To design and implement the DETR method with on-the-fly data augmentation for detecting the BISINDO alphabet and words.
2. To evaluate the performance of the DETR model using PyCOCOTools, including accuracy, mean Average Precision (mAP), and detection stability.
3. To explore ways to optimize the DETR model to achieve near real-time detection.
4. To assess the model's robustness to variations in input, including differences in lighting, hand positions, and users.
5. To develop a web-based system utilizing the DETR model that is interactive, accessible, and user-friendly for non-technical users.

1.4 Research Benefits

Based on the background, research questions, and objectives outlined above, the benefits of this study are summarized as follows:

1. Providing an accurate method for detecting the BISINDO alphabet and words using DETR with on-the-fly data augmentation techniques.
2. Evaluating the DETR model's performance using PyCOCOTools, including accuracy, mean Average Precision (mAP), and detection stability.
3. Providing insights and strategies for optimizing the DETR model to achieve near real-time detection.
4. Producing an interactive and user-friendly web-based system prototype as a reference for BISINDO recognition system development.
5. Contributing to deep learning research in sign language recognition, supporting inclusive communication for the deaf community, and serving as a reference for future studies.

1.5 Scope of the Study

To ensure the research remains focused and well-defined, the scope of this study is as follows:

1. The research data consists of images of the BISINDO alphabet and words, collected using a laptop camera in a controlled environment with a plain background (white wall), with subjects wearing black clothing to maintain consistent hand-object contrast.
2. The letters J, R, and 17 dynamic (movement-based) words are represented as static poses for image processing purposes.
3. The method used in this study is the Detection Transformer (DETR), and no comparisons with other detection methods are conducted.
4. Model testing was conducted on the test set and in a real-world usage scenario using a laptop camera, in an environment similar to the training data. Testing included variations in lighting and subjects not present in the training data but did not cover environmental conditions significantly different from the training data.
5. The scope of detection in this study is limited to 43 BISINDO classes, consisting of letters A through Z and static BISINDO words, namely Apa, Ayo, Baik Hati,

Bodoh, Cerewet, I Love You, Kamu, Keras Kepala, Makan, Menangis, Nama, Pintar, Saya, Setia, Siapa, Sombong, and Tidur. This study does not include detection of numbers or BISINDO sentences.