IT FOR SOCIETY, Vol. 09, No. 01
ISSN 2503-2224

BISINDO (*Bahasa Isyarat Indonesia*) Sign Language Recognition Using Deep Learning

Ricky Setiawan[1], Yustina Yunita[2], Fajri Fathur Rahman[3], Hasanul Fahmi[1]
Faculty of Computing
President University
Cikarang, Indonesia
hasanul.fahmi@president.ac.id

***Abstract*—Sign language recognition plays a crucial role in facilitating communication for individuals with hearing impairments. This paper presents a deep learning-based approach for recognizing Bahasa Isyarat Indonesia (BISINDO), the sign language used in Indonesia. The proposed system employs convolutional neural networks (CNNs) and recurrent neural networks (RNNs) to automatically extract features from sign language gestures and classify them into corresponding linguistic units. The dataset used for training and evaluation consists of annotated BISINDO sign language videos. Preprocessing techniques such as normalization and augmentation are applied to enhance the robustness of the model. Experimental results demonstrate the effectiveness of the proposed approach in accurately recognizing BISINDO sign language gestures, achieving state-of-the-art performance compared to existing methods. The developed system shows promising potential for real-world applications in enhancing communication accessibility for the hearing-impaired community in Indonesia.**

Keywords: ***BISINDO, CNN, Deep Learning Model, MobileNetV2, Sign Language, Real-time detection.***

I..INTRODUCTION

Bahasa Isyarat Indonesia (BISINDO), often known as Indonesian Sign Language, is a significant method of communication for Indonesia's deaf community that focuses on hand gestures, facial emotions, and body movements. This project presents a deep learning strategy that uses Convolutional Neural Networks (CNNs) and MobileNetV2 to recognize BISINDO signs in real time. To assist in accurate recognition, a carefully selected dataset of 26 BISINDO alphabet images received severe preprocessing. Using CNNs and MobileNetV2, the model performs well at reading BISINDO sign signals. This project represents a significant advancement in the use of advanced technology, especially deep learning approaches, to support lesser-known sign languages such as BISINDO. The effective deployment of this technology has affected research on the creation of assistive tools that allow deaf people to communicate in a natural way. The results of this study show that new technology, particularly deep learning models, can be used to bridge communication gaps and combine sign languages into modern technological solutions.

Sign language is a comprehensive communication system that uses gestures, facial expressions, and body postures to enable hearing-impaired people to communicate effectively. Many countries, with over 70 million deaf individuals worldwide, have established their own sign languages. BISINDO (Bahasa Isyarat Indonesia or Indonesian Sign Language) is widely utilized by the deaf community in Indonesia. The advent of digital communication tools and platforms has increased the demand for automatic sign language recognition systems.

While advances in sign language identification have been made for widely recognized languages such as ASL (American Sign Language), BISINDO remains neglected in technology. To overcome this, there is the need to create a strong deep learning model capable of identifying and interpreting BISINDO in real-time from video input.

II. LITERATURE REVIEW

2.1 BISINDO

Bahasa Isyarat Indonesia, commonly referred to as BISINDO, represents a collection of distinct deaf sign languages prevalent in Indonesia, especially on the island of Java. Its foundation is rooted in American Sign Language, but

unified language, especially when lobbying for its official recognition and implementation in educational settings, the reality is that its variants across cities might not always be mutually comprehensible.

In the realm of sign language within Indonesia, a

notable duality exists. The country recognizes two primary sign languages: Sistem Isyarat Bahasa Indonesia (SIBI) and Bahasa Isyarat Indonesia (BISINDO). While educational institutions officially adopt SIBI, its practical application among the deaf community is limited. The reason being, SIBI is a direct translation of the spoken Indonesian language, incorporating its grammatical intricacies, including prefixes and suffixes, making it less intuitive for the deaf. Conversely, BISINDO offers a more natural approach, translating spoken Indonesian words and pairing them with contextual expressions. The preference for BISINDO over SIBI is not just a matter of ease but also stems from the deaf community's strong recommendation. They advocate for BISINDO's recognition as the official Indonesian sign language, replacing SIBI.

A pivotal study conducted by Isma in 2012 [1] shed light on this linguistic diversity. The research highlighted that the sign languages employed in Jakarta and Yogyakarta, though related, are fundamentally different. They share approximately 65% of their lexicon, but their grammatical structures differ, suggesting an ongoing divergence. The differences are so pronounced that during a meeting in Hong Kong, Isma's participants had to switch to Hong Kong Sign Language for effective communication. One notable distinction is the word order: Yogyakarta predominantly uses a verb-final (SOV) structure, while Jakarta leans towards a verb-medial (SVO) arrangement, especially when the subject or object roles of the noun phrases are interchangeable. The study did not delve into the sign language variations of other Indonesian cities.

2.2 Preprocessing

Preprocessing serves as a pivotal phase in refining data for optimal training, employing techniques to enhance image quality by eliminating obstacles and irregularities [2]. This crucial step ensures that the data is appropriately normalized, contributing to effective training and achieving superior accuracy. Noteworthy in preprocessing is background subtraction, a method based on the RGB color standard (Red, Green, Blue), where adjustments to these values result in the desired image outcomes [3] [4] .

In the realm of literature, the study by Avola et al. [5] on 'Exploiting Recurrent Neural Networks and Leap Motion Controller for the Recognition of Sign Language and Semaphoric Hand Gestures' underscores the importance of preprocessing in refining data quality, particularly for applications involving sign language recognition. The study demonstrates how preprocessing techniques contribute to optimal training, paving the way for effective gesture recognition using advanced neural network architectures.

Another notable contribution comes from "Continuous Chinese Sign Language Recognition with CNN-LSTM" [6], where preprocessing plays a crucial role in preparing data for the recognition of continuous sign language gestures. The study highlights the significance of normalization through preprocessing in achieving accurate recognition results using Convolutional Neural Networks (CNN) and Long Short-Term Memory networks (LSTM). In the field of medical imaging, "Image preprocessing of abdominal CT scan to improve visibility of any lesions in kidneys" [7], explores preprocessing techniques to enhance the visibility of lesions in abdominal CT scans. The study emphasizes the role of preprocessing in improving image quality for medical diagnostics.

These studies collectively reinforce the importance of preprocessing in diverse domains, demonstrating its role in optimizing data quality for effective training and accurate model outcomes.

2.4 CNN (Convolutional Neural Network)

The Convolutional Neural Network (CNN) has emerged as a significant model for object recognition [8]. Since 2012, CNN has been instrumental in supporting object recognition tasks . CNN's strength lies in its ability to distribute data across various layer frames [9]. Over time, CNN has undergone several enhancements, leading to the development of variants like VGG16, ResNet, and the contemporary 3D-CNN, which is a trending area of study. As depicted in Figure 1, CNN processes data by converting it into layers comprising max pooling and fully connected layers [10].

Depending on the dataset, CNN utilizes both high-resolution and low-resolution layers.

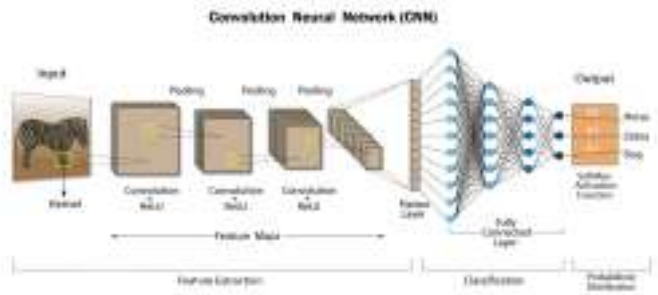

Figure 1. Convolutional Neural Network (CNN) Architecture

The convolutional layer, a fundamental component of this, processes the input image with a smaller-sized filter or kernel to extract salient features such as edges, textures, and patterns. This convolution operation involves the filter moving over the input image, computing the dot product at each step, resulting in a feature map that signifies the presence of specific features in the original image. Two crucial parameters in this process are the stride, which determines the filter's movement across the image, and padding, which adjusts the image's spatial dimensions by adding zeros around its border. This ensures that spatial resolution is maintained post-convolution. Another vital layer within the feature extraction process is the pooling layer. Its primary role is to reduce the spatial dimensions of the feature map, ensuring the network remains less sensitive to minor translations and variations in the image. Common pooling operations include max pooling, which takes the maximum value from a set of values in the feature map, min pooling, which takes the minimum, and average pooling, which computes the average of the values.

Following feature extraction, the multidimensional feature maps are transformed into a one-dimensional vector through a flattening process. This vector is then introduced to the fully connected layer, which functions similarly to a traditional Multi-Layer Perceptron (MLP). Here, every neuron from the preceding layer connects to every neuron in the subsequent layer, responsible for the

final classification or regression task. Through mechanisms like forward propagation, the output is computed, and backpropagation adjusts the network's weights based on the error. In summary, CNNs, with their combination of convolutional, pooling, and fully connected layers, offer an efficient and adaptive approach to image recognition and related tasks.

III. RESEARCH METHOD

3.1 Dataset

In this project, we used CNN and MobileNetV2 to develop a Bisindo (Bahasa Isyarat Indonesia) language recognition system. Figure 2 illustrates the BISINDO (Bahasa Isyarat Indonesia) for each alphabet that will be recognized. We require a set of datasets that accurately represent sign language motion in Bisindo (Bahasa Isyarat Indonesia) to complete this project. As a result, we decided to use a Kaggle dataset specifically for this purpose. The dataset was captured with various lighting and backgrounds, and it had been pre-processed with a rotation + 30 degree or even a horizontal flip, as demonstrated in Figure 3.

We use image data as the primary data type in this project, which consists of images that represent 26 of the alphabet in Bisindo sign language movement. The dataset contains 350 images in each class as well as 90 testing images in each class.

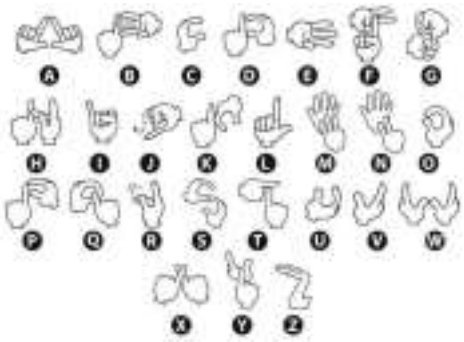

Figure 2. BISINDO Sign Language for all alphabets

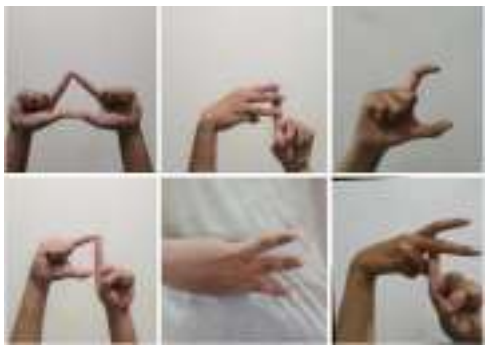

Figure 3. Example of the dataset. From the top left to right (A, B, C) and bottom (D, E, F)

3.2 Convolutional Neural Network (CNN)

In this study, we present an advanced deep learning model for BISINDO (Indonesian Sign Language) recognition, integrating the principles of Convolutional Neural Networks (CNN) with advanced techniques in machine learning. Our model is anchored by the MobileNetV2 architecture, a CNN renowned for its efficiency, especially in mobile applications, and pre-trained on the extensive ImageNet dataset. This choice ensures a robust feature extraction capability crucial for interpreting sign language images.

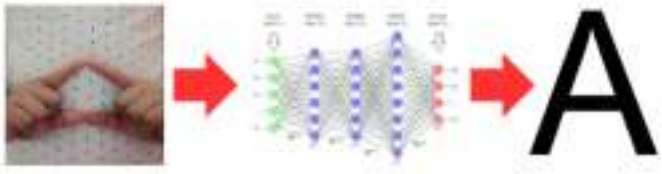

Figure 4. Illustration of CNN on BISINDO Sign Language Translation

3.3 Model Architecture

We adapt MobileNetV2 for our specific application by adjusting the input shape to process 150x150 pixel images and freezing its layers to leverage the pre-trained weights effectively, a strategy known as transfer learning.

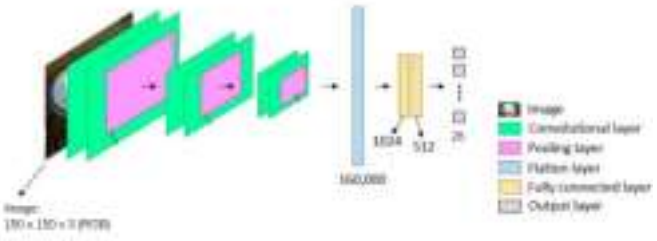

Figure 5. Illustration of architecture and training models of Convolutional Neural Networks

Beyond the base model, our architecture is enhanced with additional layers tailored for sign language recognition. A GlobalAveragePooling2D layer follows to minimize overfitting by simplifying the output from the preceding convolutional layers. This is complemented by a fully connected dense layer with 1024 neurons (using ReLU activation), which serves to interpret the complex patterns in the data. A dropout layer with a rate of 0.5 further aids in reducing overfitting. The architecture culminates in a softmax-activated dense layer with 26 neurons, each representing a class in the BISINDO alphabet, thus facilitating multi-class classification.

This framework we use reflects an integration of CNN principles with transfer learning, data augmentation, and strategic fine-tuning. Particularly pertinent for BISINDO sign language recognition [11].

IV. RESULT

System can detect input images with accuracy 93% for an average 26 classes. The classification example can be seen on figure 6-9.

Figure 6. Real time image of Class V detected as Class V

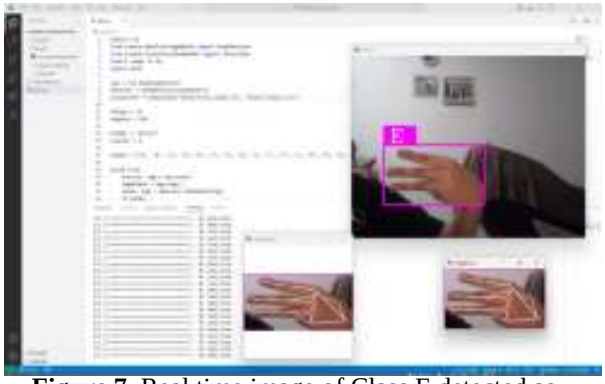

Figure 7. Real time image of Class E detected as Class E

Figure 8. Real time image of Class L detected as Class L

Figure 9. Real time image of Class O detected as Class O

V. CONCLUSION

This project enhances the field of sign language recognition by focusing on the Indonesian deaf community's use of Bahasa Isyarat Indonesia (BISINDO), also known as Indonesian Sign Language. This project successfully constructed a deep learning model capable of real-time recognition of BISINDO indications by leveraging Convolutional Neural Networks (CNNs) and MobileNetV2 architecture.

The combination of CNN concepts and the MobileNetV2 architecture produced impressive results in reading BISINDO sign language motions. The model showed significant capabilities, setting the framework for accurate and effective recognition [12].

This research not only solves a technological gap in detecting lesser-known sign languages, but it also demonstrates the promise of leveraging sophisticated technology, especially deep learning models, to assist the deaf community. The use of this technology will have an impact on the development of assistive devices that allow hearing-impaired people to communicate naturally.

While the project reached significant milestones, further adjustments and developments are required to improve the model's accuracy across all BISINDO sign language levels. Future research will try to improve real-time detection and interpretation of BISINDO by refining the model, exploring larger datasets, and delving into alternative approaches.

This project highlights the significance of technical breakthroughs in supporting successful communication for the deaf community, as well as the possibilities for implementing sign languages into modern technological solutions.

ACKNOWLEDGEMENT

We extend our sincere appreciation to President University, our academic institution, and the Faculty of Computer Science for providing an environment conducive to research and learning. Special gratitude is directed towards Asst. Prof. Dr. Hasanul Fahmi, our esteemed lecturer in Deep Learning, for his invaluable guidance and unwavering support throughout the research process and writing.

We would also like to express our gratitude to Ms. Cutifa Safitri, the Head of the Informatics Study Program at the Faculty of Computer Science, for her support and contributions to our academic journey.

Finally, heartfelt thanks go to our families and friends for their continuous encouragement and understanding. Their love and support have played a pivotal role in our ability to successfully complete this research.

REFERENCES

[1] Isma, S. T. P. (2012). Signing varieties in Jakarta and Yogyakarta: Dialects or separate languages. Master of Art Thesis, The Chinese University of Hong Kong.

[2] Perumal, S., & Velmurugan, T. (2018). Preprocessing by contrast enhancement techniques for medical images. International Journal of Pure and Applied Mathematics, 118(18), 3681–3688.

[3] Solichin, A., & Harjoko, A. (2013). Metode Background Subtraction untuk Deteksi Obyek Pejalan Kaki pada Lingkungan Statis. Seminar Nasional Teknologi Informasi 2013, 1–6.

[4] Alginahi, Y. (2010). Preprocessing Techniques in Character Recognition. Intech, 10, 1–21. https://doi.org/10.5772/9776.

[5] Avola, D., Bernardi, M., Cinque, L., Foresti, G.L., & Massaroni, C. (2019). Exploiting Recurrent Neural Networks and Leap Motion Controller for the Recognition of Sign Language and Semaphoric Hand Gestures. IEEE Transactions on Multimedia, 21(1), 234–245. https://doi.org/10.1109/TMM.2018.2856094.

[6] Yang, S., & Zhu, Q. (2017). Continuous Chinese sign language recognition with CNN-LSTM. Ninth International Conference on Digital Image Processing (ICDIP 2017), 10420(100), 104200F. https://doi.org/10.1117/12.2281671.

[7] Bindu, G.H., Reddy, P.V.G.D. Prasad, & Murty, M. Ramakrishna. (2018). Image preprocessing of abdominal CT scan to improve visibility of any lesions in kidneys. Journal of Theoretical and Applied Information Technology, 96(8), 2298–2306.

[8] Karambakhsh, A., Kamel, A., Sheng, B., Li, P., Yang, P., & Feng, D.D. (2019). Deep gesture interaction for augmented anatomy learning. International Journal of Information Management, 45(October 2017), 328–336. https://doi.org/10.1016/j.ijinfomgt.2018.03.004.

[9] Ariesta, M.C., Wiryana, F., Suharjito, & Zahra, A. (2018). Sentence Level Indonesian Sign Language Recognition Using 3D Convolutional Neural Network and Bidirectional Recurrent Neural Network. 2018 Indonesian Association for Pattern Recognition International Conference (INAPR), 16–22. https://doi.org/10.1109/INAPR.2018.8627016.

[10]Molchanov, P., Gupta, S., Kim, K., & Kautz, J. (2015). Hand gesture recognition with 3D convolutional neural networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 1–7. https://doi.org/10.1109/CVPRW.2015.7301342.

[11]Syulistyo, A. R., Hormansyah, D. S., & Saputra, P. Y. (2020). SIBI (Sistem Isyarat Bahasa Indonesia) translation using Convolutional Neural Network (CNN). IOP Conference Series: Materials Science and Engineering, 732(1), 012082. https://doi.org/10.1088/1757-899X/732/1/012082.

[12]Köpüklü, O., Gunduz, A., Kose, N., & Rigoll, G. (2019).