International Journal of Academic and Applied Research (IJAAR)
ISSN: 2643-9603
Vol. 9 Issue 8 August - 2025, Pages: 18-21

Development of Deep Learning Model For Detection of Sign Language Alphabets in an Image

Oladimeji Olayanju, Adeniji oluwashola david, Adeniyi Michael Odejayi

, Department of Computer science, University of ibadan, Nigeria

E-mail address: oladimejiolayanju@gmail.com , od.adeniji@ui.edu.ng, odejoass_ny@yahoo.com

Abstract: Sign language serves as a crucial means of communication for the deaf and hard-of-hearing community. However, the lack of widespread understanding among non-signers presents a communication barrier. This research focuses on the development of a deep learning-based model for the automatic recognition of sign language alphabets in images. The approach in this research leverages Convolutional Neural Networks (CNNs) integrated with Attention Mechanisms, specifically the Convolutional Block Attention Module (CBAM), to enhance feature extraction and improve recognition accuracy. EfficientNet is utilized as the backbone network due to its high performance and parameter efficiency. A dataset of Sign Language gestures is collected from a repository and preprocessed to train and evaluate the model. Experimental results demonstrate the effectiveness of the developed approach in accurately classifying sign language alphabets in an image, contributing to advancements in assistive technology and bridging the communication gap between signers and non-signers. The model was deployed on a web-based interface for real-time usability, making sign language recognition more accessible and practical for everyday interactions. In the experiment, the multi- model was tested using appropriate metrics in which accuracy of 99.90%, AUC of 99.99%, loss of 0.03%, recall of 99.90%, precision of 99.90% were obtained and independent test set that was used on the model showed impressive results. It was observed from this study that the developed approach performed excellently well in predicting sign language alphabets in an image. We recommend the applicability of this work in different sector for effective communication.

Keywords—Detection, sign language, deep learning, Convolutional Neural Networks

1. Introduction

Innovation is the motivating force that changes the way we live. Since its inception, Computer Science has been a developing research field. This development has been driven by improvements in technology, with increase in hardware capabilities and growth in software efficiency. Due to this, many lives have been influenced greatly and improved by computers that have become literally universal in present societies. Currently, we are in the era of Big Data (Duan et al., 2019). The advancement of technology and accessibility of such massive quantity of data have given us chance to make use of it in ways that can transform our lives greatly. In this regard, Machine Learning (ML) is gradually turning into the major framework for mining information from data (Rajan et 2020). It has been in use in a variety of fields, from the industrial sector to fraudulent activity detection. In this era, the applications of these technologies seem to be almost limitless. Conventionally, machine learning was restricted to work with small datasets. By implication, the ideas from Image Processing field to tackle real world challenges were not realistic. This is so because there is lack of sufficient data to train with. Similarly, we lack the sufficient computational power for the algorithms to learn. Nevertheless, currently, the computing power has developed greatly (Hennessy & Patterson, 2019); huge amount of data is now at our disposal, innovative algorithms for data processing have been developed, and the network infrastructure and the datacenters have been smoothly integrated with data generators and processors. In particular, the field of computer vision is progressing due to the improvement in machine learning and deep learning. Several research areas have advanced to the point that algorithms can be used to automate some real-life tasks that were formerly performed manually by humans. In these automation difficulties, computer vision is becoming an essential instrument. Consequently, there is currently a high need for well-functioning computer vision systems. Computer vision is an area of Artificial Intelligence (AI) with the objective of getting information from images (Thompson et al., 2020). The image data may be in any form and from various sources: video frames, cameras snaps, or high-resolution data from medical imaging equipment, and so on. Computer vision makes use of the ideals and models derived from numerous fields including machine learning, cognitive sciences, psychology, pattern recognition, and so on, for creating applications to handle very specific issues (sabeenian et al 2020). Because a single image can contain several items, one of the most important tasks of a computer vision system is to trace and identify the objects inside it. (Babatunde et al., 2015) However, according to recent statistics by Forbes, over 70 million people worldwide are deaf, with more than 80% of them residing in underdeveloped and developing countries. Hence, sign language is often used for communication by people with impaired hearing and speech to express their thoughts and emotions. However, non-signers find it extremely difficult to understand, therefore trained sign language interpreters are needed during medical, legal appointments, educational and training sessions. To address this, there is a need to develop models to recognize alphabets gestures in sign language. Sign languages are natural languages that have been developed through the evolution of contact between the hearing impaired but have not been invented by any system . With the implementation of sign language in schools, hearing Teachers and students can communicate through both linguistic and non-linguistic ways that can aid in creating an interactive environment for hearing-impaired and hard-of-hearing students and thus enhance the effectiveness of academic learning (Wadhawan, et al, 2020). The communication between a person from an impaired community with a person who does not understand sign language could be a tedious task. owever, with rapid breakthroughs that image processing and recognition have achieved in variety of applications including medical imaging, surveillance systems, remote sensing, among others recently The internet service driven network is a new approach to the provision of network computing that concentrates on the services you want to provide as adopted in (Adeniji et al 2008), However there are no adequate provision for quality of service (QoS) in OpenFlow using Flow Label to reduce bits required as a field to match packets in internet protocol six (IPv6)(olabisi et al 2019) coupled with the reborn of Convolutional Neural Networks and improvement in machine learning and deep learning in the field of computer vision, various works have revealed that machine learning and data analytics can be used to recognize and classify sign languages (Rakshit,, et al 2024).

2. METHODOLOGY

2.1 Data Description

The aim of this work is to use the cutting-edge CNN framework for sign language recognition, there is a need for dataset that will be proper to use. Hence, sign language images with the label(s). The dataset used in this study is derived from the Sign Language Alphabet Dataset created by Akash Nagaraj (2018) and published on Kaggle. The original dataset consists of 87,000 images of Sign Language alphabets, organized into 29 classes. These classes include 26 letters (A Z) and three additional categories: SPACE, DELETE, and NOTHING, which are useful for real time applications in sign language recognition. Each image in the dataset is 200x200 pixels in size. For this research, the dataset has been modified by reducing the number of images per class to 700, resulting in 18,200 images. This preprocessing step ensures a more balanced and computationally manageable dataset while maintaining the diversity necessary for effective training of the deep learning model. Furthermore, this work aims to work exclusively on Fingerspelling Sign Language. These requirements are majorly for the training dataset, which constrained the choices and limited the sources from which the data is obtained. For this research, Sign Language with background images were considered.

Figure 2.1 Sample of Sign Language with background image.

2.2 Model development with CBAM (Convolutional Block Attention Module)

Various researchers have contributed and are contributing towards improving the methods that will be effective in recognizing sign language in an image, However, drawbacks still exists inform of computational complexity and background challenges in the images. With respect to this, we used Efficient Net to extract features from the sign language images and CBAM(Convolutional Block Attention Mechanism) to select the relevant features. CBAM(Convolutional Block Attention Mechanism, is a lightweight and effective attention mechanism used in deep learning particularly for convolutional Neural Networks. CBAM entails two attention modules: • Channel Attention Module: This focuses on “What” information is important by applying across the channel dimension of the feature maps • Spatial Attention Module: This focuses on “Where” the important information is by applying attention across spatial dimension (height and width) of the feature maps.

By incorporating CBAM, the model effectively learns to highlight the most discriminative features in sign language images, reducing the impact of background noise and improving classification accuracy. Once the refined features are extracted using EfficientNet and enhanced through CBAM, they are passed into a fully connected network for classification. At this stage, the extracted CNN features will be trained by adding custom dense layers and an output layer suitable for a multi-class classification problem. Since this research addresses a multi-class classification problem, the model is designed to classify each input into a single category among multiple possible classes.

International Journal of Academic and Applied Research (IJAAR)
ISSN: 2643-9603
Vol. 9 Issue 8 August - 2025, Pages: 18-21

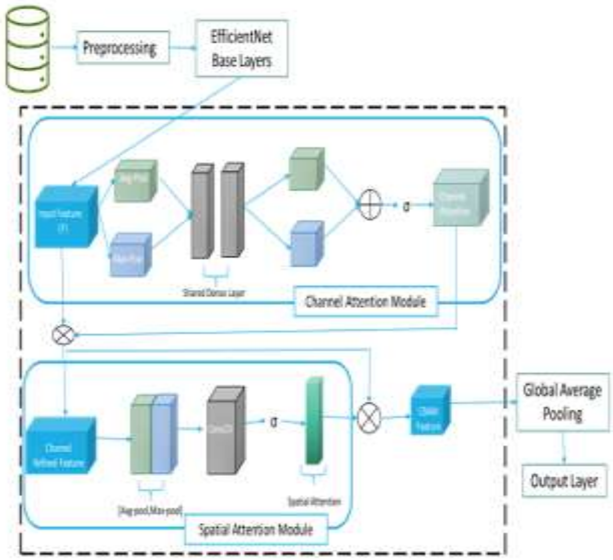

Figure 2.2 CBAM Architecture

This module can be inserted into CNN architecture and help improve performance in task like object detection, classification and segmentation by making the model focus on the most relevant features and regions. The development of a robust multi-class classification framework ensures that the system can accurately distinguish between different sign language gestures, improving recognition performance and practical applicability.

3 Results and Discussions

The results of the sign language recognition model are hereby presented in this section. The results were validated with test dataset and independent test set.

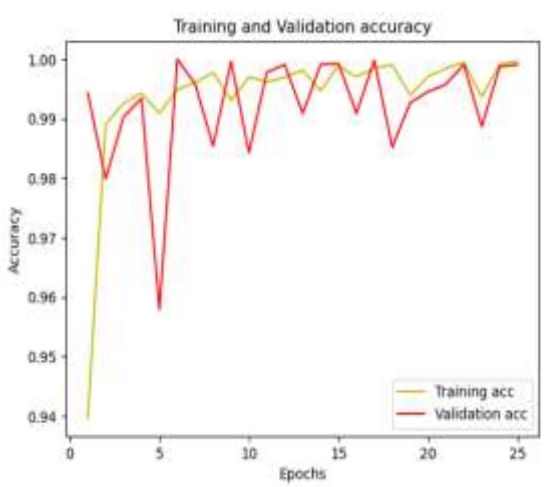

Figure 3.1 Training and Validation Accuracy Result Visualization

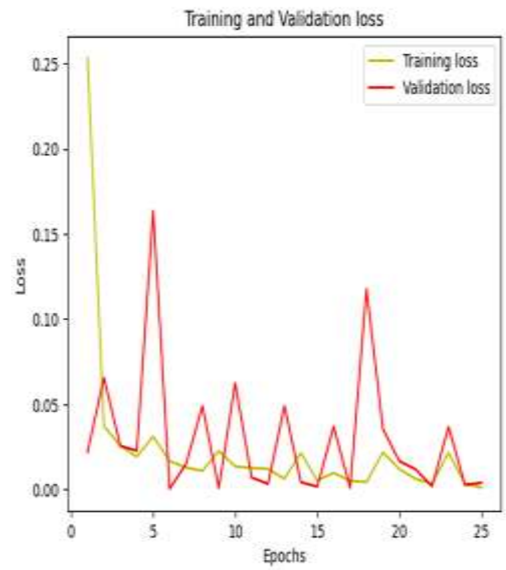

Figure 3.2 Training and Validation Loss Result Visualization

Table 3.1 gives the summary of the results evaluation using the standard metrics. Table 3.1: Summary of Model Performance.

Metric	Score
Recall	0.9990842490842491
Precision	0.999095160110236
AUC	0.9999996184024263
F1 Score	0.9990841302325169
Training Accuracy	99.93%
Validation Accuracy	99.90%
Training Loss	0.0028
Validation Loss	0.0303

The table 3.2 shows the result of the model showing accuracy, Recall, precision and F1score.

Table 3.2 below presents the performance metrics of various convolutional neural network (CNN) models on sign language classification task. Each model, including Customized CNN, VGG16, InceptionV3, ResNet50V2, and DenseNet121, is evaluated based on accuracy, recall, precision, and F1-score.

Table 3.2: Model performance comparison on test Set

Model	Accuracy	Recall	Precision	F1-score
Our Model (EfficientNet + CBAM	99.90%	99.90%	99.90%	99.90%
VGG16	99.34%	0.99	0.99	0.99
Inception V3	81.62%	0.87	0.82	0.81
ResNet50	99.70%	1.0	1.0	1.0
DenseNet121	99.80%	1.0	1.0	1.0
CNN and Tensorflow	99.77%	-	-	-
DeafTech Vision(DTV-CNN)	99.87%	-	-	-

Table 3.2 above presents the performance metrics of various convolutional neural network (CNN) models on sign language classification task. Each model, including Customized CNN, VGG16, InceptionV3, ResNet50V2, and DenseNet121, is evaluated based on accuracy, recall, precision, and F1-score. The developed model achieves outstanding results across all metrics, with nearly perfect scores close to 1.00, indicating exceptional performance in correctly classifying instances. VGG16 also demonstrates high accuracy and robust performance across all metrics, while ResNet50V2 and DenseNet121 exhibit similar results. However, InceptionV3 falls short in comparison, with lower accuracy and performance metrics overall.

4. Conclusion

In the experiment, the multi-class model was tested using appropriate metrics in which accuracy of 99.90%, AUC of 99.99%, loss of 0.26%, recall of 99.90%, precision of 99.90% were obtained and independent test set that was used on the model showed impressive results.

5.ACKNOWLEDGMENT (HEADING 5)

The authors wish to thank the Department of Computer Science, University of Ibadan for the support in this research work.

6.REFERENCES

1. R. G. Rajan and M. Judith Leo, "American Sign Language Alphabets Recognition using Hand Crafted and Deep Learning Features," 2020 International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India, 2020, pp. 430-434, doi: 10.1109/ICICT48043.2020.9112481.
2. Duan, Y., Edwards, J. S., & Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda. International Journal of Information Management, https://doi.org/10.1016/j.ijinfomgt.2019.01.021
3. Hennessy, J. L., & Patterson, D. A. (2019). Computer Architecture: A Quantitative Approach. In Morgan Kaufmann, San Francisco, CA, sixth edition. https://doi.org/10.1088/1751 8113/44/8/085201
4. Thompson, N. C., Greenewald, K., Lee, K., & Manso, G. F. (2020). The Computational Limits of Deep Learning. 4, 2–5. http://arxiv.org/abs/2007.05558.
5. Babatunde, O. H., Armstrong, L., Leng, J., & Diepeveen, D. (2015). A survey of computer-based vision systems for automatic identification of plant species. Journal of Agricultural Informatics, 6(1), 61–71. https://doi.org/10.17700/jai.2015.6.1.152.
6. Sabeenian, R. S., Bharathwaj, S. S., & Aadhil, M. M. (2020). Sign language recognition using deep learning and computer vision. J Adv Res Dyn Control Syst, 12(5), 964-968..
7. Wadhawan, A., & Kumar, P. (2020). Deep learning-based sign language recognition system for static signs. Neural computing and applications, 32(12), 7957-7968.
8. Adeniji, S.D, Khatun, S, Borhan, MA,Raja, RSA , (2008).A design proposer on policyframework in IPV6 network‖. 2008 IEEE International Symposium on Information Technology.Vol 4,pp 1-6.
9. Olabisi,A. A., Adeniji,, O. D. Abeng Enangha (2019) A Comparative Analysis of Latency, Jitter and Bandwidth of IPv6 Packets Using Flow Labels in Open Flow Switch in Software Defined Network‖ Afr. J. MIS, Vol.1, Issue 3, pp. 30-36.
10. Rakshit, P., Paul, S., & Dey, S. (2024). Sign language detection using convolutional neural network. Journal of Ambient Intelligence and Humanized Computing, 15(4), 2399-2424.