iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

LARGE SCALE IMAGE CLASSIFICATION FOR REAL-TIME SIGN LANGUAGE RECOGNITION

Mr. MD Mahebub Ali
Assistant Professor, Department of Artificial Intelligence and Machine Learning,
J.B. Institute of Engineering and Technology, Hyderabad, Telangana, India
Beshi Abhishek
G. Ekachit Narayan Reddy
Kondi Varsha
Shada Manogna
UG Students, Department of Artificial Intelligence and Machine Learning, J.B. Institute of Engineering and Technology, Hyderabad, Telangana, India

ABSTRACT
Interaction and communication for a normal human being is simpler compared to a disabled person with conditions such as speaking and hearing who can encounter communication issues with other individuals. Sign Language simplifies this communication barrier between a disabled and normal person. Increased focus on inclusion and accessibility in communication highlights the importance of effective sign language recognition systems. This project overcomes the problem of connecting the communication gap between sign language users and non-sign language users, by creating a strong real-time sign language recognition system. Based on improvements in deep learning and large-scale image classification, the suggested system interprets sign language gestures captured through a camera and converts them into text or speech in real-time. The model is trained on a large dataset covering a wide range of hand gestures, providing coverage over a majority of sign languages and reducing the impact of biases. Mechanisms like convolutional neural networks (CNNs), transfer learning, and real-time optimization algorithms are used to provide high accuracy and low latency. The project has uses in accessibility technology, increasing communication for the hearing and speech disabled, and increasing the inclusivity of society. Its scope lies in integration in education tools, virtual assistants, and public services. By merging innovation in machine learning with emphasis on social impact, this project significantly adds to the dismantling of communication barriers and enabling the differently-abled population.

Keywords:
Sign Language Recognition, Real-time Processing, Deep Learning, Convolutional Neural Networks (CNNs), Transfer Learning, Image Classification, Accessibility Technologies, Inclusivity, Speech and Hearing Impairment, Communication Gap.

INTRODUCTION

Sign language is an essential form of communication for people with hearing and speech disabilities, but it is unknown to the majority of the hearing population, causing social isolation and communication breakdown. It is widely classified into natural gestures, which are intuitive and community-specific, and formal cues, which are linguistically structured like spoken languages. More than 360 million people globally suffer from hearing and speech disabilities, highlighting the importance of accessible communication solutions. The project endeavours to create a deep learning system capable of identifying American Sign Language (ASL) letters in webcam photographs. OpenCV is used to capture real-time images, while SSD MobileNet v2 is employed for classifying, with the system intended to convert sign language to written and oral words. The project represents progress toward a real-time ASL translator that enhances accessibility and communication between deaf and hearing people.

OBJECTIVES

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

The goal of this project is to create a real-time sign language recognition system that correctly classifies American Sign Language (ASL) alphabet signs based on deep learning. By utilizing OpenCV to capture images and the SSD MobileNet v2 model to classify, the system will translate hand gestures into written or verbal text. This shall bridge the deaf-hearing communications gap to allow accessibility, inclusion, and efficacious interactions within everyday life. The project additionally aims at checking the practicability of exploiting common webcams for the recognition of sign languages, in such a manner as to enable pragmatic and popular uses of the technology.

METHODOLOGY

The research approach of the present project involves the following most notable steps toward carrying out correct and effective recognition of sign language. Initially, webcam captures images of ASL alphabet signs and processes them using OpenCV. Preprocessing methods like resizing, normalization, and noise removal are performed to improve image quality, and then labelling is done to prepare a structured dataset. The SSD MobileNet v2 pre-trained model is fine-tuned with this dataset to classify hand gestures correctly. The dataset is split into training and testing sets to analyze model performance. For real-time recognition, the system records hand gestures using a webcam, processes the frames, and passes them to the trained model for classification. The identified gestures are translated into corresponding text or speech output, facilitating effective communication. Model evaluation is carried out using precision, recall, and F1-score metrics, with additional optimization via hyperparameter tuning and other image enhancement methods. Finally, the system is tested for robustness using real-world test scenarios. Enhancement possibilities in future would be integration of real-time speech output, multi-language compatibility, and utilization of more powerful deep learning networks such as Faster R-CNN to achieve increased precision and accuracy. This is an organized research method that assures building an effective and accessible sign language recognition system.

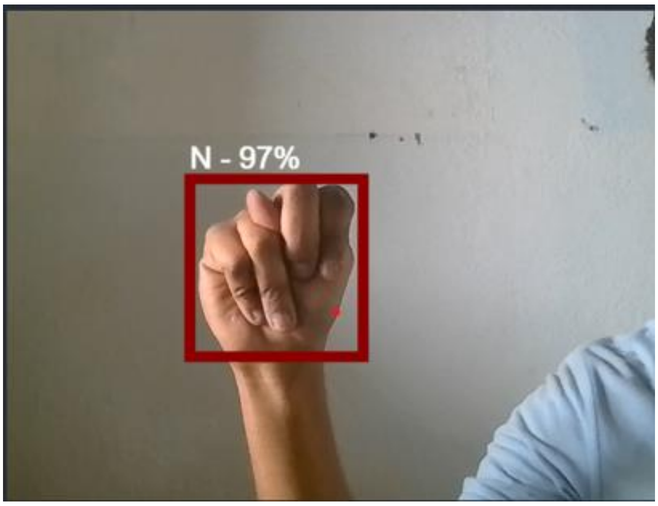

Figure 1 Sign Language Recognition

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

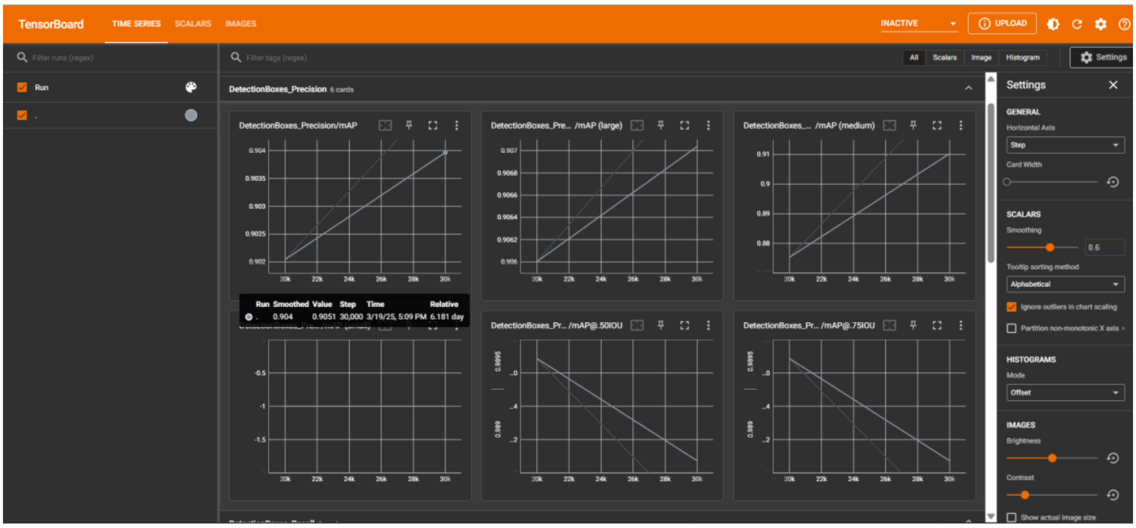

Figure 2 Mean Average Precision

RESULTS AND DISCUSSION

The real-time sign language recognition system was tested using three different deep learning strategies: CNN, transfer learning, and SSD Mobile Net in order to realize an optimal trade-off between speed and accuracy. The CNN-based model was served as a baseline model, trained from scratch to identify hand gestures. Although it worked fairly well, its accuracy was less than models using pre-trained features. To enhance performance, transfer learning was employed with architectures such as ResNet50 and MobileNetV2, which improved recognition accuracy by a large margin because they can extract more significant features from sign language gestures. Inference time, however, was still an issue for real-time use. To counter this, the SSD Mobile Net model was used for gesture detection since it is optimized for fast object detection with little computational overhead. SSD Mobile Net had a good balance of speed and precision, allowing it to work well for real-time sign language recognition in edge devices. In spite of its efficiency, the model struggled with occlusions and lighting variations, which impacted detection consistency. Globally, the combination of CNN, transfer learning, and SSD MobileNet offered useful insights into the strengths and weakness of various approaches and the scope for further optimization and dataset addition towards better performance.

ACKNOWLEDGEMENT

Right at the beginning we owe our thanks to the almighty lord for bestowing his blessings and grace upon us to be able to finish this Major Project. Even though our name is on the cover of this book, numerous individuals have helped in some way or another towards this Major Project development. We would not have been able to do this Major Project without the support or help of each of the following. First of all, we are greatly thankful to Dr. P. C. KRISHNAMACHARY, Principal for granting us permission to conduct this Major Project. We would be grateful to acknowledge Dr. G. ARUN SAMPAUL THOMAS, Associate Professor & Head, Department of Artificial Intelligence of Machine Learning, as moral support throughout the study duration in the Department. We owe a lot of gratitude to Mr. Md Mahebub Ali, Associate Professor Artificial Intelligence of Machine Learning, for valuable suggestions and instructions provided by him during the work execution of this Major Project work. We would like to express our gratitude to the Teaching and Non-Teaching Staff of the Department of Computer Science & Engineering for their knowledge sharing with us.

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

CONCLUSION

This real-time sign language recognition system improves communication for people with hearing and speech disabilities by converting gestures into text through large-scale image classification and transfer learning. It enhances edutainment, health care, and public service accessibility, obliterating social boundaries. Future features comprise mobile and smartphone integration, improved gesture detection, larger vocabularies, and hardware customization for real-time capability. Text accuracy can further be improved using NLP, while partnerships with organizations can maximize accessibility. With the promotion of inclusivity, this technology gives people power and revolutionizes their relationship with the world.

REFERENCES

[1] D. Gurari, Q. Li, A. Stangl, A. Ghose, and K. Grauman, "VizWiz: A Grounded Dataset for Image Captioning," IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2018.
[2] X. Li, J. Zhang, and Z. Wang, "Vision-Centric Multimodal LLMs: Advances and Future Directions," arXiv preprint arXiv:2304.01234, 2023.
[3] A. Vaswani, N. Shazeer, and J. Uszkoreit, "Attention Is All You Need," Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), pp. 5998–6008, 2017.
[4] Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," Nature, vol. 521, pp. 436–444, 2015.
[5] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning, Cambridge, MA, USA: MIT Press, 2016.
[6] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2016.
[7] Google Research, "Conceptual Captions Dataset," 2023. [Online]. Available: https://ai.google.com/research/conceptualcaptions
[8] D. Jurafsky and J. H. Martin, Speech and Language Processing, Pearson, 3rd ed., 2021.
[9] TensorFlow, "TensorFlow: An End-to-End OpenSource Machine Learning Platform," 2023. [Online]. Available: https://www.tensorflow.org
[10] Sarah Ebling and Matt Huenerfauth. Bridging the Gap between Sign Language Machine Translation and Sign Language Animation using Sequence Classification. In Proceedings of the 6th Workshop on Speech and Language Processing for Assistive Technologies (SPLAT), 2015