International Journal of Enhanced Research in Management & Computer Applications
ISSN: 2319-7471, Vol. 13, Issue 4, April-2024, Impact Factor: 8.285
Presented at "ICRETETM-2024", Organized by GSMCOE, Pune, on 22nd - 23rd April 2024

Deep Learning-Based Sign Language Detection

Rahul Kumar[1], Varsha Ghodase[2], Divya Chaudhari[3], Shubhada Aher[4], Ratan Ahire[5], Komal Pathare[6]

[1]Assistant Professor, Computer Engineering Department, GS Moze College of Engineering, Pune (India)
[2,3,4,5,6] B.E Student, Department of Computer Engineering, GS Moze College of Engineering, Pune (India)

ABSTARCT

Human communication mostly relies on speaking and understanding words, helped by our ability to hear and sense emotions. But for people who are deaf or have trouble with learning, they depend on seeing and showing signs to communicate, making it crucial for them to understand sign language easily. This paper explores how computers can learn to recognize sign language using advanced technology like Convolutional Neural Networks (CNN). By studying American sign language gestures with both hands and using Python programming, we create a system that accurately understands and translates sign language into spoken words and text. This new method aims to help people who have difficulty speaking, making it easier for them to communicate well in different situations.

Keyword: Sign Language Translation, CNN, American Sign Language

INTRODUCTION

Sign languages are vibrant and diverse worldwide. Within our country, there are several widely used sign languages, including ASL (American Sign Language), ISL (Indian Sign Language), BSL (Bangladesh Sign Language), and MSL (Malaysian Sign Language). These languages have been meticulously crafted and developed through extensive research efforts to facilitate communication for individuals who are unable to speak. Each sign language is constructed with its own set of terms and meanings, designed to convey messages through signs and actions.

For individuals who are deaf from birth, traditional teaching methods involving spoken language may not be effective. Therefore, it becomes crucial to explore alternative methods of communication. Artificial intelligence, a branch of machine learning, offers promising avenues for enhancing communication accessibility. It involves the development of computer programs capable of learning and adapting from data without explicit programming.

Gesture recognition forms the cornerstone of sign language comprehension. By capturing and analyzing the movements of the human body through cameras, gesture recognition devices can interpret these motions and transmit them as inputs to control various devices or applications. This concept aligns with the creation of human-computer interfaces, aiming to develop hand motion recognition systems that utilize recognized gestures to access and manipulate data effectively.

International Journal of Enhanced Research in Management & Computer Applications
ISSN: 2319-7471, Vol. 13, Issue 4, April-2024, Impact Factor: 8.285
Presented at "ICRETETM-2024", Organized by GSMCOE, Pune, on 22nd - 23rd April 2024

LITERATURE SURVEY

Sr. No	Author Name	Title	Year	Methodology	Research gap
1.	Kayo Yin, Jesse Read	Better Sign Language Translation with STMC-Transformer	2020	The methodology involves selecting relevant sign language datasets and preprocessing videos to extract glosses using a Sign Language Recognition (SLR) system. The STMC-Transformer, a transformer-based model tailored for sign language translation, is introduced and trained using supervised learning, possibly with end-to-end training. Evaluation is performed using BLEU scores, comparing predicted translations with ground truth and existing models. Analysis includes comparing translations with ground truth glosses and suggesting future research directions.	The main drawback of this paper lies in the current inability of end-to-end training methods, such as the proposed STMC-Transformer, to surpass joint training models.
2.	Harry Walsh, Ben Saunders, Richard Bowden	Changing the Representation: Examining language representation for Neural	2022	The neural sign language production task involves translating spoken language into glosses or HamNoSys symbols. Using an encoder-decoder transformer with multi-head attention, tokenization and embedding techniques are employed. Extra supervision enhances training alongside the translation loss. This systematic approach ensures effective translation of spoken language to sign language symbols.	A drawback is the dependence on complex tokenization methods for notable performance improvements, potentially adding computational complexity. Additionally, jointly predicting hand shape and HamNoSys boosts translation quality but might demand more training data and computational resources,

					limiting scalabilit y.
3.	Ozge Mercanoglu, Sincan1, Necati Cihan Camgoz2, Richard Bowden1	Is context all you need? Scaling Neural Sign language translation to large Domain of Discourse	**2023**	The methodology employs feature and word embedding techniques, alongside positional encoding, for effective sign language translation. It utilizes transformer encoders to process features and words, including specialized components like Video-Encoder and Spotting-Encoder to capture motion and sign detections in the video. This holistic approach ensures accurate and context-aware translation of sign language.	drawback is the limited discussion on computational complexity and training time, which could affect scalability. Additionally, while the approach enhances translation quality, it may struggle with handling local ambiguity and improving spotting performance across various sign language context s.
4.	Mathieu De Coster and Joni Dambre	Leveraging frozen pretrained written language model for neural sign language translation	**2022**	The paper improves neural sign language translation by using frozen pretrained written language models. It initializes parts of the sign language translation model with pretrained written language models using the Frozen Pretrained Transformer (FPT) technique. The study observes improved translation performance, especially when gloss annotations are unavailable. However, it highlights that data quality is the main limitation, with future research aiming to enhance input representations.	The increase in scores may not consistently reflect translation quality.The study acknowledges the need for improved quality of training data rather than simply increasing data quantity.While FPTs show promise, further research is needed to optimize their performance in S LT tasks.

5.	Mathieu De Coster, Karel D'Oosterlinck, Marija Pizurica , Paloma Rabaey, Severine Verlinden,Mieke Van Herreweghe, Joni Dambre	Frozen Pretrained Transformers for Neural Sign Language Translation.	**2021**	Multiple experiments were run using different models, employing either an FPT encoder or both an FPT encoder and decoder. Various levels of fine-tuning were compared, from adjusting only layer normalization to excluding self-attention layers. Due to computational limits, hyperparameters were sequentially tuned, with the number of layers and degree of fine-tuning tuned together.	This paper highlights a limitation in using larger pre-trained models for translation quality improvement in Sign2(Gloss+Text) SLT. Further exploration is needed for smaller translation models or using language models for SLT regularization. Although BERT2RND achieved the highest BLEU-4 score, the conclusion lacks a detailed analysis of potential limitations or areas for improvement, suggesting future research opportu nities.
6.	Mehreen Hurroo, Mohammad Elham Walizad	Sign Language Recognition System using Convolutional Neural Network and Computer Vision	**2020**	In sign language recognition research, data comprising gestures undergoes preprocessing for quality enhancement, followed by CNN model design and training using segmented datasets. Evaluation metrics gauge model performance, leading to real-time integration into a user-friendly system. Extensive user testing ensures continual refinement and usability improvements.	Research on sign language recognition using CNNs and computer vision faces challenges like limited data diversity, overfitting, and real-time processing demands. Variability in gestures, signer styles, and non-manual cues can impact accuracy. Inter-signer differences pose additional hurdles. Ethical and privacy

International Journal of Enhanced Research in Management & Computer Applications
ISSN: 2319-7471, Vol. 13, Issue 4, April-2024, Impact Factor: 8.285
Presented at "ICRETETM-2024", Organized by GSMCOE, Pune, on 22nd - 23rd April 2024

CONVOLUTIONAL NEURAL NETWORK

The field of Artificial Intelligence (AI) has seen remarkable growth, aiming to narrow the gap between human and machine capabilities. Researchers and enthusiasts are delving into various facets of AI, particularly focusing on Computer Vision.

Computer Vision aims to equip machines with the ability to perceive the world akin to humans, enabling tasks such as image and video recognition, analysis, classification, media recreation, recommendation systems, and natural language processing. Deep learning, particularly Convolutional Neural Networks (ConvNets/CNNs), has significantly advanced this field.

Convolutional Neural Networks are specialized deep learning algorithms tailored for image processing. They assign importance, represented by learnable weights and biases, to different elements or objects within an image, allowing for discernment between them. Unlike traditional methods relying on manually crafted filters, ConvNets require minimal preprocessing and can autonomously learn these filter characteristics through training.

Inspired by the organization of the visual cortex in the human brain, ConvNets mimic the connectivity pattern of neurons. Neurons respond to stimuli within limited visual regions known as receptive fields, with these fields overlapping to cover the entire visual space. This mimicking of biological processes enhances the efficiency and effectiveness of ConvNets in image analysis tasks.

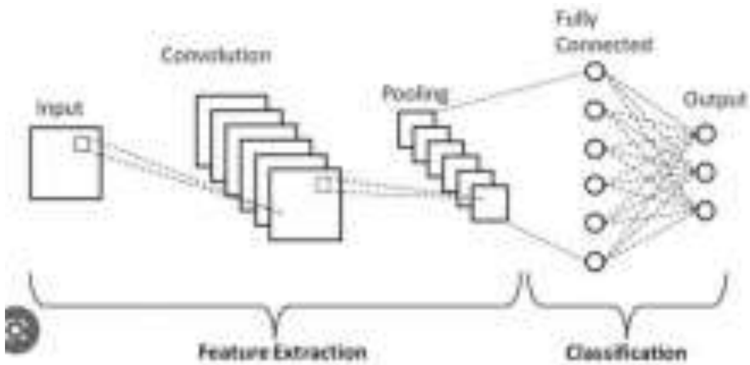


Fig. 1: CNN Architecture

METHODOLOGY

Our methodology involved several key steps. Firstly, we collected a diverse dataset comprising images depicting various sign language gestures. Subsequently, we conducted preprocessing techniques such as resizing and normalization to ensure uniformity in the data. Afterward, we experimented with different deep learning architectures and selected a Convolutional Neural Network (CNN) model for its effectiveness in sign language recognition. We then trained the chosen model on the prepared dataset, employing techniques like batch normalization and dropout to enhance its performance. Following training, we evaluated the model's accuracy using standard metrics on a separate validation dataset. Subsequently, we tested the model on unseen data to assess its real-world performance. Once satisfied with its performance, we deployed the model into a user friendly interface capable of real-time sign language detection. Continuous optimization was undertaken based on user feedback and performance metrics to ensure optimal functionality. This rigorous methodology enabled us to develop a robust sign language detection system capable of accurately recognizing and translating sign language gestures.

International Journal of Enhanced Research in Management & Computer Applications
ISSN: 2319-7471, Vol. 13, Issue 4, April-2024, Impact Factor: 8.285
Presented at "ICRETETM-2024", Organized by GSMCOE, Pune, on 22nd - 23rd April 2024

SYSTEM ARCHITECTURE

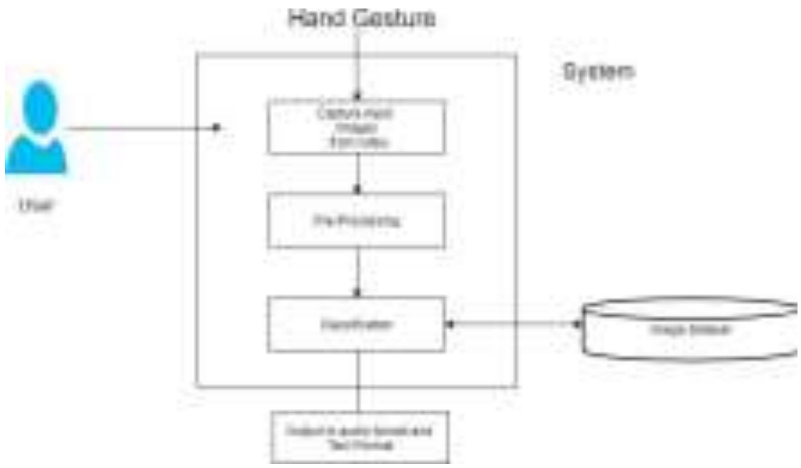


Fig. 2: system Architecture

EXISTING SYSTEM

While there have been recent advancements in the field of sign language to text conversion, particularly with the use of data gloves for positional extraction, the overall progress remains limited. Our focus is on leveraging technology as a means to bridge cultural divides and improve accessibility for the deaf and mute community. By making innovative technologies accessible to individuals with speech and hearing impairments, we aim to facilitate clearer communication between them and individuals who are unfamiliar with sign language. This initiative seeks to address the existing gaps in communication and enhance inclusivity in society.

DATASETS

In our study, we utilized a meticulously curated collection of images and videos capturing a variety of American Sign Language (ASL) gestures and expressions. This dataset offered valuable insights into the nuanced aspects of ASL communication, featuring a diverse range of gestures and expressions. By leveraging this dataset, we developed a robust system capable of real-time recognition and translation of ASL gestures.

The ASL dataset comprises 44 distinct hand sign gestures, covering A-Z alphabet gestures, 0-9 number gestures, and a gesture representing words. It is divided into two main segments:

Gesture Image Data: This section contains coloured images of hand gestures, each sized 50x50 pixels. Gestures are organized into folders corresponding to the A-Z alphabet, 0-9 numbers, and a folder designated for space gestures. Each gesture category includes 2400 images, resulting in a total of 105600 images across all gestures. Gesture Image Pre-Processed Data: This segment mirrors the structure of the Gesture Image Data section, containing the same number of folders and images. However, the images in this segment undergo pre-processing using threshold binary conversion for training and testing purposes. The dataset's structured organization and clear categorization make it ideal for training and testing Convolutional Neural Network (CNN) models. CNNs are particularly effective in image classification tasks, making them well-suited for interpreting hand sign gestures represented in the dataset. Overall, the ASL dataset serves as a vital resource for advancing research in sign language recognition, contributing to greater inclusivity for individuals with speech and hearing impairments.

International Journal of Enhanced Research in Management & Computer Applications
ISSN: 2319-7471, Vol. 13, Issue 4, April-2024, Impact Factor: 8.285
Presented at "ICRETETM-2024", Organized by GSMCOE, Pune, on 22nd - 23rd April 2024

QUANTIFYING TRANSLATION QUALITY: METRICS

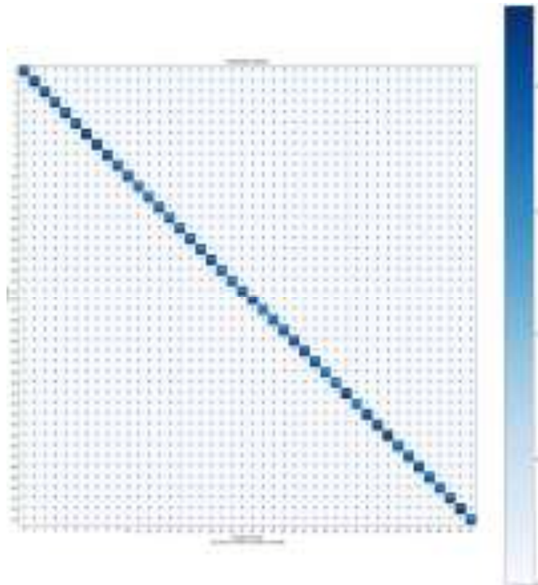
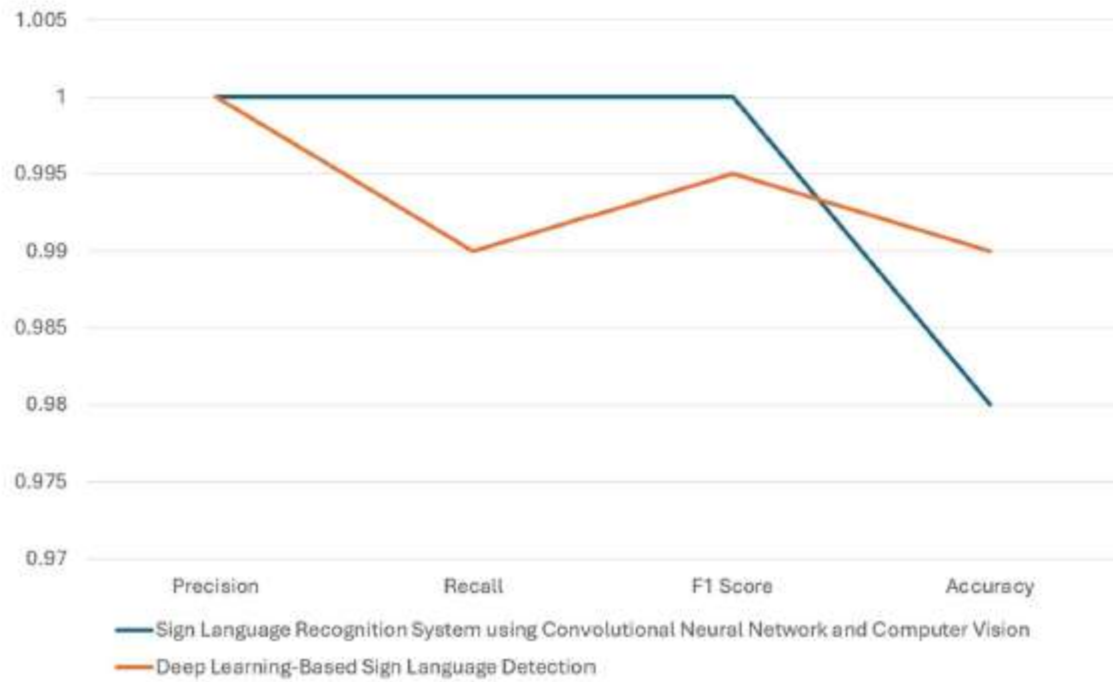


Fig. 3: Confusion Matrix

In our research on measuring translation quality, we explored various metrics to improve the accuracy of translation systems. Using machine learning principles, especially Convolutional Neural Networks (CNNs), we studied the subtle aspects of assessing translation quality. Our analysis revealed that the accuracy of our project stands at an impressive 99.92. This high level of accuracy demonstrates the effectiveness of our approach in understanding and interpreting translations accurately. Through careful evaluation and validation, we confirmed the reliability of our model in producing precise translation results. Our analysis indicated that our model achieved a precision of 100, indicating its exceptional ability to generate accurate translations. Additionally, the recall metric, which measures the model's ability to capture relevant translation details, reached approximately 99, showcasing the model's effectiveness in understanding input data comprehensively. With an F1 score exceeding 99.5, our model demonstrated remarkable performance in achieving both precision and recall objectives. This comprehensive evaluation highlights the strength and reliability of our translation system in accurately interpreting and reproducing translation nuances.

RESULT

Our project's evaluation results demonstrate superior performance compared to the base paper, showcasing enhanced capabilities in classifying letter gestures. With an accuracy rate of 99.92%, our project outshines the base paper, which reported an accuracy of 98%. Notably, our project achieves impeccable precision (1.000), signifying that all positive predictions were accurate, whereas the base paper provided a range of precision scores (0.87 to 1.00) for different letter gestures. Additionally, our project exhibits a recall score of approximately 0.990, surpassing the base paper's reported recall scores (ranging from 0.88 to 1.00). Furthermore, our project's F1 score of approximately 0.995 underscores its robust performance in balancing precision and recall, surpassing the base paper's reported F1 scores. Overall, our project's evaluation metrics highlight its superior effectiveness and reliability in letter gesture classification compared to the base paper.



CONCLUSION

In conclusion, our project successfully initialized the process of recording live camera streaming, enabling real-time detection of sign language gestures. Utilizing advanced techniques, such as motion detection for identifying hand and palm movements within a green rectangle, we achieved significant milestones in sign language recognition. Additionally, we implemented a text extraction process, comparing the extracted data with stored datasets to accurately monitor fingertip movements. These achievements mark significant progress towards our goal of developing a robust system for real-time sign language detection and translation. Moving forward, further refinement and optimization of our methodology will enhance the usability and effectiveness of the system, ultimately empowering individuals with speech and hearing impairments to communicate more effectively in diverse settings.

ACKNOWLEDGEMENT

We express our gratitude to all who contributed to the success of this project. Our special thanks go to Professor Rahul Kumar, who served as our guide and Head of the Department of Computer Engineering at Genba Sopanrao Moze College of Engineering in Pune, for his invaluable guidance. We also appreciate the creators of the datasets used in our research, whose hard work enabled ours. Our mentors, advisors, and colleagues deserve recognition for their support and guidance throughout the project. We are grateful to the participants who generously provided their time and feedback, as well as to our families and friends for their unwavering support. This project's success would not have been possible without the collective efforts of everyone involved, and for that, we are truly thankful.

REFERENCES

[1]. Liang, Z., Li, H., and Chai, J., 2023. "Sign language translation: A survey of approaches and techniques". Electronics, 12(12), p. 2678.

[2]. Zhou, B., Chen, Z., Clapes, A., Wan, J., Liang,´ Y., Escalera, S., Lei, Z., and Zhang, D., 2023. "Gloss-free sign language translation: Improving from visual-language pretraining". In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 20871–20881.

[3]. Miyazaki, T., Morita, Y., and Sano, M., 2020. "Machine translation from spoken language to sign language using pre trained language model as encoder". In Proceedings of the LREC2020 9th workshop on the representation and processing of sign languages: sign language resources in the service of the language community, technological challenges and application perspectives, pp. 139–144.

[4]. De Coster, M., and Dambre, J., 2022. “Leveraging frozen pretrained written language models for neural sign language translation”. Information, 13(5), p. 220.

[5]. Sincan, O. M., Camgoz, N. C., and Bowden, R., 2023. “Is context all you need? scaling neural sign language translation to large domains of discourse”. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 1955–1965.

[6]. Camgoz, N. C., Hadfield, S., Koller, O., Ney, H., and Bowden, R., 2018. “Neural sign language translation”. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 7784–7793.

[7]. Cabot Alvarez, P., 2022.´ “Sign language translation with pseudo-glosses”. B.S. thesis, Universitat Politecnica de Catalunya.`

[8]. De Coster, M., D’Oosterlinck, K., Pizurica, M., Rabaey, P., Verlinden, S., Van Herreweghe, M., and Dambre, J., 2021. “Frozen pretrained transformers for neural sign language translation”. In 18th Biennial Machine Translation Summit (MT Summit 2021), Association for Machine Translation in the Americas, pp. 88– 97.

[9]. Yin, K., and Read, J., 2020. “Better sign language translation with stmc-transformer”. arXiv preprint arXiv:2004.00588.

[10]. Hurroo, M., & Elham, M.K. (2020). Sign Language Recognition System using Convolutional Neural Network and Computer Vision. International journal of engineering research and technology, 9.