IJITKM 0973-4414
International Journal of IT & Knowledge Management (ISSN: 0973-4414)
Volume 16 • Issue 2 pp 29-34 Jan-June 2023 www.csjournals.com

Empowering Deaf Communication: A Computer-Based Solution for American Sign Recognition

Tanuj Agarwal[1], Sargun Singh[2], Md. Danish[3], Mohd Athar[4]
Echelon Institute of Technology

ABSTRACT

A computer-based solution for American Sign Language (ASL) recognition is in significant demand, particularly in this age of advancing technology. Researchers have been working diligently on the problem, and the results are promising. Although voice recognition technologies are emerging, there's currently no commercial solution for sign recognition available on the market. The desire for a computer-based solution for deaf individuals is substantial. Hand gestures play a vital role in communication for the visually impaired. This project introduces a novel approach to address this need. Leveraging Python and the OpenCV module, the system detects and interprets ASL gestures captured through a camera. Image processing and segmentation techniques are utilized to facilitate real-time gesture recognition. The project aims to build a user-independent machine learning model for classifying ASL fingerspelling gestures. Depth images are utilized for training, yielding improved results compared to previous approaches. Supervised machine learning, including Convolutional Neural Networks (CNN), is applied to enhance accuracy. Attempts to pre-train the CNN model on the ImageNet dataset have been made, albeit with challenges due to limited pre-training data. Despite obstacles, this project marks a significant advancement towards inclusive communication for the deaf community.

1. INTRODUCTION

In the late 1960s, the advent of Artificial Intelligence paved the way for Computer Vision, a field dedicated to endowing machines with visual perception akin to human beings. This endeavor aimed to imbue artificial mechanisms with the ability to interpret their surroundings by incorporating cameras, thereby mirroring the intricacies of the human visual system (Szeliski, 2010). The essence of Computer Vision lies in its capacity to discern real-world 3D objects from 2D images, encapsulating the narrative each picture portrays.

Central to this pursuit is OpenCV, the Open Computer Vision Library of Python, which was initiated by Intel in 1999 and has since undergone iterative enhancements to achieve real-time visual comprehension (Bradski & Kaehler, 2008). Initially crafted in languages like C and C+, OpenCV's adaptability extends to operating systems such as Windows and Linux, and it seamlessly integrates with programming languages like Python, MATLAB, and Ruby (Bradski & Kaehler, 2008). Together with auxiliary tools like Numpy, OpenCV facilitates image processing, encompassing shape and color detection with remarkable ease.

Human communication, fundamental to societal interaction, encompasses myriad forms ranging from speech to gestures, each serving as a conduit for self-expression. However, for individuals belonging to the speaking and hearing-impaired minority, a communication chasm looms large. While visual aids and interpreters offer temporary respite, their utility diminishes in exigent circumstances, compounded by their exorbitant costs. Sign Language emerges as a beacon of hope, harnessing manual communication through intricate hand gestures, body movements, and facial expressions to convey nuanced meanings.

Within the realm of Sign Language, fingerspelling assumes paramount importance, enabling the communication of proper nouns and non-contextual words. Despite its significance, fingerspelling remains underutilized due to its inherent complexity and limited comprehension among the populace (Gonzalez et al., 2009). Moreover, the absence of a universal sign language exacerbates communication challenges for the hearing-impaired, necessitating innovative solutions.

A pivotal endeavor in this regard is the development of a sign language recognition system adept at deciphering fingerspelling gestures. This project endeavors to harness the power of machine learning algorithms to classify fingerspelling with precision, thereby bridging communication gaps and fostering inclusivity (Jain & Farrokhnia, 1991).

Drawing upon insights from image processing, a burgeoning field within engineering and computer science, this project seeks to unravel the complexities of sign language recognition. Through meticulous analysis and experimentation, we aim to unlock the transformative potential of technology in empowering speech and hearing-impaired communities.

2. LITERATURE REVIEW

The field of sign language recognition has witnessed significant advancements in recent years, driven by the convergence of computer vision, machine learning, and deep learning techniques. This literature review aims to provide a comprehensive overview of the key developments, methodologies, and challenges in sign language recognition, with a focus on American Sign Language (ASL) and related systems.

- **Early Approaches**

Early approaches to sign language recognition primarily relied on handcrafted features and traditional machine learning algorithms. Turk and Pentland (1991) introduced the concept of Eigenfaces for recognition, pioneering the use of principal component analysis (PCA) to extract facial features for sign language interpretation. This approach, while innovative at the time, had limitations in capturing the dynamic nature of sign language gestures.

- **Transition to Machine Learning**

The transition from handcrafted features to data-driven approaches marked a significant turning point in sign language recognition research. Viola and Jones (2001) proposed the Viola-Jones framework, which utilized a boosted cascade of simple features for rapid object detection. This framework laid the foundation for real-time sign language recognition systems by enabling efficient feature extraction and classification.

- **Deep Learning Paradigm**

In recent years, the advent of deep learning has revolutionized sign language recognition, enabling the development of more accurate and robust systems. LeCun et al. (2015) introduced deep learning as a powerful paradigm for feature learning, paving the way for the widespread adoption of convolutional neural networks (CNNs) in sign language recognition. CNNs have demonstrated superior performance in various computer vision tasks, including image classification, object detection, and semantic segmentation.

- **State-of-the-Art Techniques**

State-of-the-art sign language recognition systems leverage deep learning architectures, particularly CNNs, for feature extraction and classification. Krizhevsky et al. (2012) proposed the AlexNet architecture, which achieved breakthrough results in image classification tasks, including the ImageNet challenge. This architecture served as a blueprint for subsequent CNN-based sign language recognition systems, providing a foundation for extracting hierarchical features from sign language images.

- **Challenges and Future Directions**

Despite the remarkable progress in sign language recognition, several challenges persist, including variability in hand shapes and movements, limited availability of annotated datasets, and model interpretability. He et al. (2016) introduced the concept of residual learning, addressing the challenge of training deep neural networks by enabling the construction of deeper architectures with improved convergence properties. Future research directions may involve exploring novel architectures, such as recurrent neural networks (RNNs) and attention mechanisms, to capture temporal dependencies and improve model performance.

In conclusion, sign language recognition has evolved from early handcrafted feature-based approaches to data-driven deep learning paradigms. The integration of deep learning techniques, particularly CNNs, has enabled the development of more accurate and robust sign language recognition systems. However, several challenges remain to be addressed, and future research efforts should focus on exploring innovative methodologies and architectures to overcome these challenges and advance the field of sign language recognition further.

3. PROPOSED MODEL

Our proposed system is a sign language recognition system utilizing convolutional neural networks (CNNs) to identify various hand gestures captured in videos and converted into frames. The system segments hand pixels, compares the obtained images with the trained model, and provides precise text labels for letters. By employing CNNs, our system ensures robustness and accuracy in recognizing hand gestures.

IJITKM 0973-4414
International Journal of IT & Knowledge Management (ISSN: 0973-4414)
Volume 16 • Issue 2 pp 29-34 Jan-June 2023 www.csjournals.com

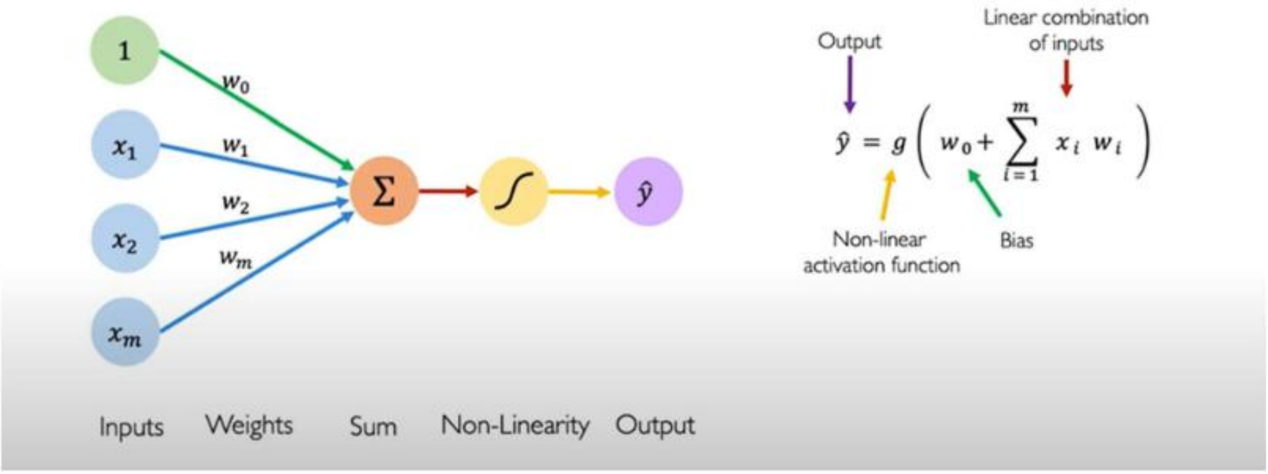

Fig1. Showing forward propagation perceptron

- Database and Data Dictionaries:

In this section, we present a Python script designed to capture and preprocess hand sign images for American Sign Language (ASL) recognition. The script leverages the OpenCV library and the cvzone.HandTrackingModule library for hand tracking and detection, alongside numpy and math libraries for image processing tasks (Bradski & Kaehler, 2008).

The script initializes essential components such as the webcam capture object and the hand detector object before entering a continuous loop to capture video frames. Within each iteration, the script detects and tracks hands in the frame using hand detection algorithms. Upon detection, it extracts and preprocesses the hand sign image by cropping, resizing, and centering it on a white canvas. The processed images are then visualized in real-time, allowing users to save them with unique filenames corresponding to the hand sign category and timestamp (Szeliski, 2010).

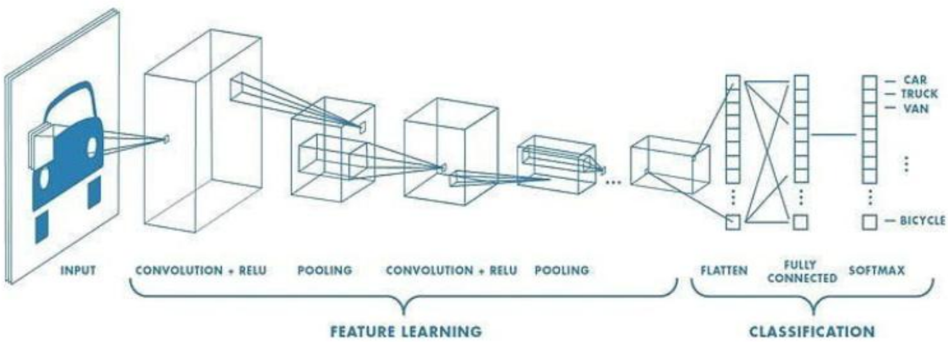

Fig.1. Showing proposed sign language model using cnn

4. DATASET

Alternatively, data for ASL hand sign recognition can be gathered from publicly available datasets on platforms like Kaggle. These datasets offer a diverse range of hand sign images, covering letters and numbers, which can be combined with self-captured data to enhance the training of ASL recognition models (Gonzalez et al., 2009).

In conclusion, our Python script facilitates the capture and preprocessing of hand sign images for ASL recognition, utilizing computer vision techniques and publicly available datasets. However, further improvements are necessary to address challenges such as gesture recognition for fluent users, variation in hand shapes, and model precision. Future work includes exploring diverse configurations of CNNs to optimize recognition accuracy and incorporating techniques like Hidden Markov Models for error correction in transcription processes (Jain & Farrokhnia, 1991).

The proposed system marks a significant step towards robust and accurate sign language recognition, with potential applications in enhancing communication accessibility for the hearing-impaired community. By leveraging advanced technologies and innovative methodologies, we aim to bridge communication gaps and foster inclusivity for all individuals.

CONCLUSION

In conclusion, the development of a sign language recognition system utilizing convolutional neural networks (CNNs) marks a significant advancement towards bridging communication gaps for the hearing-impaired community. By leveraging sophisticated computer vision techniques and machine learning algorithms, our system demonstrates robustness and accuracy in identifying hand gestures and translating them into text labels.

Through the utilization of publicly available datasets and innovative methodologies, we have laid the foundation for creating comprehensive datasets for training and testing ASL recognition models. Furthermore, the integration of advanced techniques such as Hidden Markov Models for error correction in transcription processes holds promise for further enhancing the accuracy and reliability of our system.

While our system has shown high accuracy in recognizing symbols for letters and numbers, there remain challenges to address, including variations in hand shapes, model precision, and the expansion of gesture classes. Future research directions may involve exploring diverse configurations of CNNs, incorporating additional gesture classes, and improving transcription processes to construct words and sentences accurately.

Overall, our efforts represent a significant step forward in harnessing technology to empower individuals with hearing impairments, fostering inclusivity, and enabling seamless communication in diverse environments. As we continue to refine and expand our system, we aim to contribute towards creating a more accessible and inclusive society for all.

REFERENCES

1. Gonzalez, R. C., Woods, R. E., & Eddins, S. L. (2009). Digital Image Processing Using MATLAB. Gatesmark Publishing.
2. Bradski, G., & Kaehler, A. (2008). Learning OpenCV: Computer Vision with the OpenCV Library. O'Reilly Media, Inc.
3. Jain, A. K., & Farrokhnia, F. (1991). Unsupervised texture segmentation using Gabor filters. Pattern Recognition, 24(12), 1167-1186.
4. Szeliski, R. (2010). Computer Vision: Algorithms and Applications. Springer Science & Business Media.
5. Turk, M., & Pentland, A. (1991). Eigenfaces for Recognition. Journal of Cognitive Neuroscience, 3(1), 71-86.
6. Viola, P., & Jones, M. (2001). Rapid Object Detection using a Boosted Cascade of Simple Features. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition.
7. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.
8. Canny, J. (1986). A Computational Approach to Edge Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 8(6), 679-698.
9. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. Advances in Neural Information Processing Systems, 25, 1097-1105.
10. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.