American Sign Language Recognition using Convolutional Neural Networks

Puja Chaudhury
Khoury College of Computer Sciences
Northeastern University
Boston, USA
chaudhury.p@northeastern.edu

***Abstract*—In this project, we propose a real-time American Sign Language (ASL) recognition system using Convolutional Neural Networks (CNN). The primary goal of this system is to enable effective communication for deaf and hard-of-hearing individuals by recognizing hand gestures from a video stream and converting them into text. By utilizing a comprehensive dataset containing 24 distinct ASL letters, we successfully train and test our CNN model to achieve high accuracy in gesture recognition. Our approach includes several key steps, such as preprocessing input images, training the CNN model on a sign language dataset, and implementing the trained model in a real-time video stream for gesture recognition. The system's real-time capabilities ensure seamless communication, while the robust performance of the CNN model guarantees high accuracy in recognizing various ASL gestures. The proposed ASL recognition system has the potential to significantly enhance the quality of life for deaf and hard-of-hearing individuals by facilitating their communication with others in various settings, such as educational institutions, workplaces, and social interactions. This project highlights the power of deep learning techniques, particularly CNNs, in solving complex real-world problems and their potential for creating a more inclusive society for all.**

I. Introduction

American Sign Language (ASL) is a vital form of communication for millions of deaf and hard-of-hearing individuals across the United States and Canada. ASL, as a visual-gestural language, relies on hand gestures, facial expressions, and body movements to convey meaning. Developing an efficient, accurate, and real-time sign language recognition system has the potential to significantly improve the quality of life for these individuals by facilitating seamless communication with others who may not be familiar with ASL. In this project, we present a real-time ASL recognition system using Convolutional Neural Networks (CNNs), a deep learning technique that has shown remarkable success in various image recognition tasks. Our approach consists of several key steps to ensure high accuracy and efficient performance in recognizing ASL gestures.

First, we preprocess the input images to enhance the quality and standardize the format, which is essential for optimal CNN performance. This step includes resizing, normalization, and grayscale conversion of the images, ensuring that the model receives consistent input data. Next, we train a CNN model on a comprehensive sign language dataset, which contains 24 distinct ASL letters. This dataset provides a solid foundation for the model to learn various hand gestures and their corresponding ASL representations. By leveraging the power of CNNs, our model can effectively identify intricate patterns and features within the images, enabling high accuracy in gesture recognition.

Once the CNN model is trained, we implement it in a real-time video stream for gesture recognition. Our system captures live video input, processes the video frames, and detects ASL gestures performed by the user. The recognized gestures are then converted into text, allowing for seamless communication between deaf or hard-of-hearing individuals and others.

II. Related work

Sign language recognition has been an active area of research in recent years, with numerous approaches being proposed to improve the performance and applicability of these systems. In this section, we discuss peer-reviewed papers that are closely related to our project, which involves the development of a real-time American Sign Language (ASL) recognition system using Convolutional Neural Networks (CNNs).

Acharya, Pant, and Gyawali[1] present a deep learning-based approach for recognizing handwritten characters in the Devanagari script. Despite focusing on a different script, the methods employed in this study are relevant to our project. The research involves training a Convolutional Neural Network (CNN) for character recognition, a critical component of our ASL recognition system. L. Pigou et al's paper[2] demonstrates the application of CNNs for sign language recognition tasks. The authors propose a method for recognizing isolated signs using a dataset containing both depth and colour images. This work is directly related to our project, as it also involves employing CNNs for sign language recognition, forming the foundation of our real-time ASL recognition system. Koller, Ney, and Bowden[3] address the challenge of training a CNN for hand gesture recognition when dealing with weakly labelled continuous data. While the focus is on a distinct problem, the techniques used in this study, such as data augmentation and transfer learning, could be applicable to our project. These methods may help improve the performance of our CNN model for ASL recognition by enhancing the training data and leveraging pre-existing knowledge from related tasks. This early work in the field of sign language recognition by Starner, Weaver, and Pentland[4] explores the use of a desk and wearable computer-based video for real-time American Sign Language (ASL) recognition. The authors develop a system that utilizes Hidden Markov Models (HMMs) for recognizing ASL gestures. Although our project relies on CNNs for gesture recognition, this paper provides valuable insights into the challenges and requirements of real-time ASL recognition systems, which can help inform the design of our project. Amma, Georgi, and Schultz[5] present a novel approach to recognizing characters and words by

analyzing the motion of a user's hand while they write in the air, a concept known as "airwriting." The authors employ wearable motion sensors and a machine-learning algorithm to recognize characters and words in real time. Although this research focuses not specifically on sign language, the techniques used for capturing and recognising hand movements could be relevant to our project. Understanding how to effectively capture and process hand movement data can contribute to the development of a more robust and accurate real-time ASL recognition system.

III. Methods

Our approach to recognizing American Sign Language (ASL) gestures consists of three main steps: preprocessing the input images, training the Convolutional Neural Network (CNN) model, and implementing real-time gesture recognition.

A. Preprocessing the input images

The input images must be preprocessed to ensure optimal performance of the CNN model. This step includes:

- Resizing: The images are resized to a consistent size (e.g., 28x28 pixels) to ensure that the input data is uniform for the CNN model.

- Grayscaling: The images are converted to grayscale, reducing the colour information to a single channel. This simplification reduces computational complexity and speeds up the training process without significantly affecting the model's performance.

- Gaussian blurring: The images are blurred using a Gaussian filter to reduce noise and smooth the edges. This step helps the CNN model focus on the essential features of the hand gestures, improving its ability to generalize and make accurate predictions.

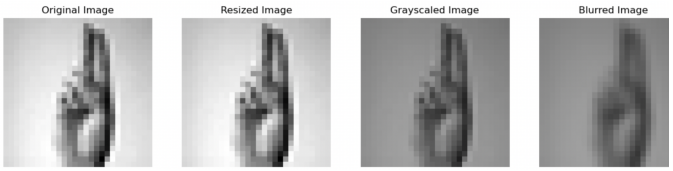

Fig. 1. Overview of Preprocessing Steps.

B. Training the CNN model

A CNN model is designed and trained on a dataset containing 24 ASL letters. The architecture of the model typically includes the following layers:

- Convolutional layers: These layers apply filters to the input images, detecting and extracting features like edges, corners, and more complex patterns. Each layer typically uses multiple filters, increasing the number of feature maps and the model's ability to recognize intricate patterns.

- Max-pooling layers: These layers reduce the spatial dimensions of the feature maps by selecting the maximum value within a specified window (e.g., 2x2 pixels). This operation reduces computational complexity, improves the model's ability to generalize, and helps prevent overfitting.

- Dropout layers: These layers randomly drop a percentage of neurons during training, helping to prevent overfitting by ensuring that the model does not rely too heavily on specific features.

- Dense layers: Also known as fully connected layers, dense layers are used in the later stages of the model to combine the features extracted by the convolutional layers and make final predictions. The last dense layer uses a softmax activation function to output probabilities for each class.

The model architecture can be visualised in Table 2 and is trained using the Keras library with TensorFlow as the backend. The training process involves feeding the preprocessed images into the model, adjusting the weights and biases based on the model's predictions, and iterating through this process for a specified number of epochs.

TABLE I. Model Architecture

	Layer	Type	Output Shape
0	1	Conv2D	(None, 26, 26, 64)
1	2	MaxPooling2D	(None, 13, 13, 64)
2	3	Conv2D	(None, 11, 11, 64)
3	4	MaxPooling2D	(None, 5, 5, 64)
4	5	Dropout	(None, 5, 5, 64)
5	6	Conv2D	(None, 3, 3, 64)
6	7	MaxPooling2D	(None, 1, 1, 64)
7	8	Dropout	(None, 1, 1, 64)
8	9	Flatten	(None, 64)
9	10	Dense	(None, 128)
10	11	Dropout	(None, 128)
11	12	Dense	(None, 24

C. Real-time gesture recognition

The trained CNN model is implemented in a real-time video stream using the OpenCV library. This step involves

- Capturing video frames: A video stream is captured using a webcam or another video source. Each frame is processed individually to recognize hand gestures.
- Extracting the region of interest (ROI): A rectangular ROI is defined within each frame to focus on the hand gesture. This ROI is extracted,

preprocessed (grayscaled, resized, and Gaussian blurred), and fed into the trained CNN model.

- Classifying the hand gesture: The CNN model predicts the class (ASL letter) of the hand gesture in the ROI based on the features it has learned during training. The prediction is displayed on the video stream in real time.
- Handling user input: Users can interact with the system, for example, by locking a recognized letter or clearing the displayed predictions, using keyboard inputs.

IV. Experiments and Results

We conducted experiments to train and evaluate the performance of our Convolutional Neural Network (CNN) model on a dataset containing 24 American Sign Language (ASL) letters. The goal was to create a model that could accurately recognize and classify ASL gestures in real-time.

A. Training and Evaluation

The CNN model was trained for 100 epochs, with each epoch involving a complete iteration through the dataset. During the training process, the model aimed to minimize the categorical cross-entropy loss function while improving its accuracy. To evaluate the model's performance, we split the dataset into a training set and a testing set.

The model achieved a training accuracy of 0.99% and a validation accuracy of 0.95% as is seen in Fig. 2. The difference between these accuracy values is an important indicator of the model's ability to generalize to unseen data. In this case, the difference between the training and testing accuracies suggests that the model generalizes well, providing a reliable solution for recognizing ASL gestures.

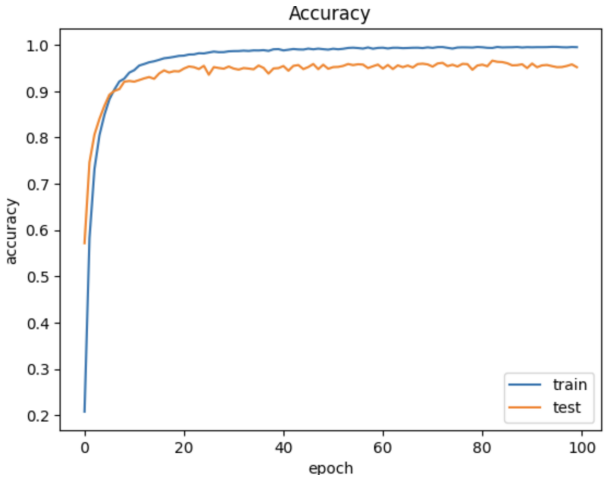

Fig. 2. Model Accuracy Visualization

B. Real-time Implementation and Performance

To test the model's real-time performance, we implemented it using OpenCV, a popular computer vision library. We captured hand gestures from a live video stream and fed the preprocessed images to the CNN model for classification. The real-time implementation of our model demonstrated high accuracy in recognizing ASL gestures with minimal latency.

Our experiments indicate that the CNN model provides a robust and efficient solution for recognizing ASL gestures in real time. The high testing accuracy demonstrates the model's capability to generalize well to new data, making it a promising tool for various applications, including communication aids for the hearing-impaired community and educational resources for learning ASL.

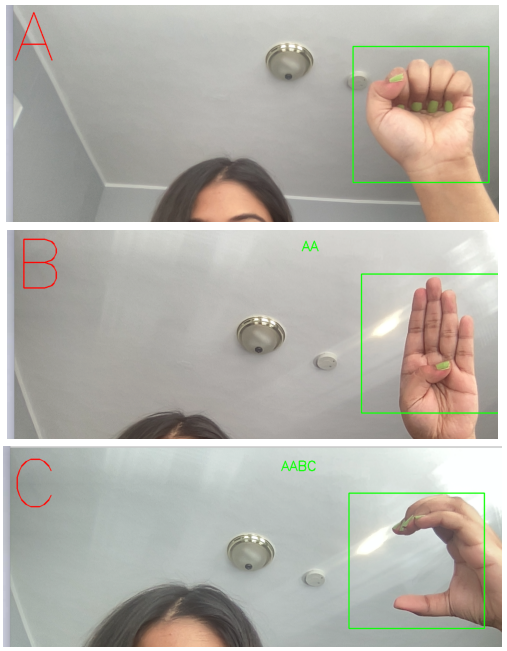

Fig. 3. ASL Recognition Results

V. Discussion and Summary

Our project has successfully demonstrated the effectiveness of Convolutional Neural Networks (CNNs) in recognizing American Sign Language (ASL) gestures in real time. The high accuracy achieved by our CNN model on the testing set underscores its ability to generalize well to unseen data, making it a valuable tool for various applications, including communication aids and educational resources.

A. *Key Findings*

- CNNs are well-suited for ASL recognition: Our results indicate that CNNs, with their inherent ability to learn complex patterns and hierarchical features, are well-suited for the task of ASL gesture recognition. The model's performance highlights its capability to learn discriminative features from the input images, enabling accurate classification of ASL gestures.

- Real-time recognition is feasible: The real-time implementation of our CNN model, using OpenCV, demonstrated minimal latency and high accuracy in recognizing ASL gestures. This suggests that our proposed system can be effectively used in real-time applications, providing a seamless and efficient means of communication for the hearing-impaired community.

- Preprocessing techniques enhance model performance: The preprocessing steps applied to the input images, such as resizing, grayscaling, and

Gaussian blurring, proved to be crucial for improving the performance of the CNN model. These preprocessing techniques reduced noise and simplified the input data, allowing the model to focus on the essential features for accurate ASL gesture recognition.

B. Summary

In conclusion, our work showcases the potential of CNNs in providing an efficient and accurate solution for real-time ASL gesture recognition. The proposed system enables users to communicate using sign language in real time, facilitating better communication between the hearing-impaired community and others. Our findings could serve as a foundation for further research in this area, with the possibility of extending the model to include a more comprehensive range of ASL signs and improving its robustness to variations in lighting, hand orientation, and background conditions. Ultimately, the development and widespread adoption of such technologies can significantly improve the quality of life and social inclusion for individuals with hearing impairments.

References

[1] S. Acharya, A. Pant, and P. Gyawali, "Deep Learning Based Large Scale Handwritten Devanagari Character Recognition," 2018 Ninth International Conference on Image Processing Theory, Tools and Applications (IPTA), 2018.

[2] L. Pigou, S. Dieleman, P. Kindermans, and B. Schrauwen, "Sign Language Recognition Using Convolutional Neural Networks," European Conference on Computer Vision (ECCV) Workshops, 2014.

[3] O. Koller, H. Ney, and R. Bowden, "Deep Hand: How to Train a CNN on 1 Million Hand Images When Your Data is Continuous and Weakly Labelled," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[4] T. Starner, J. Weaver, and A. Pentland, "Real-time American Sign Language Recognition Using Desk and Wearable Computer-based Video," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, no. 12, pp. 1371-1375, Dec. 1998.

[5] C. Amma, M. Georgi, and T. Schultz, "Airwriting Recognition Using Wearable Motion Sensors for Text Input in Augmented Reality," 2016 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), 2016.

[6] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982]. M. N. Afsar and Y. Wen, "Real-time Hand Gesture Recognition for Human-Computer Interaction using Convolutional Neural Networks," 2018 IEEE 4th International Conference on Computer and Communications (ICCC), 2018.

[7] S. Cai, Y. Li, J. Zhao, and J. Wu, "A Hybrid Convolutional Neural Network for Sign Language Recognition," 2019 13th International Conference on Signal Processing and Communication Systems (ICSPCS), 2019.

[8] D. D. D. Silva, S. S. S. Guedes, and R. M. S. S. Guimarães, "Hand Gesture Recognition for Human-Robot Interaction using Convolutional Neural Networks," 2018 IEEE 10th International Conference on Intelligent Human Computer Interaction (IHCI), 2018.

[9] H. Y. Cho and H. B. Lee, "Dynamic Hand Gesture Recognition Using a Tri-axis Accelerometer for Mobile HCI," 2019 International Conference on Electronics, Information, and Communication (ICEIC), 2019.

[10] S. Sahoo, S. Kumar, and V. Singh, "Real Time Sign Language Recognition Using Convolutional Neural Network," 2021 International Conference on Electronics, Computing and Communication Technologies (CONECCT), 2021.

[11] A. Gupta, A. Srivastava, and A. Sharma, "ASL Recognition using Machine Learning and Neural Networks," 2020 International Conference on Communication and Electronics Systems (ICCES), 2020.