AMERICAN SIGN LANGUAGE TO TEXT USING CONVOLUTIONAL NEURAL NETWORK

Yash Yadhav

School of Computer Science and Engineering (SCOPE)  
Vellore Institute of Technology  
Chennai, India

Govind G Krishnan

School of Computer Science and Engineering (SCOPE)  
Vellore Institute of Technology  
Chennai, India  
K B Poovanna  
School of Computer Science and Engineering (SCOPE)  
Vellore Institute of Technology  
Chennai, India

Allen Shijo K.  
School of Computer Science and Engineering (SCOPE)  
Vellore Institute of Technology  
Chennai, India

**Abstract**— In an effort to bridge communication gaps between individuals with hearing/speech disabilities and those without, we propose an innovative **Sign Language to Text Conversion Model**. This model enables real-time translation of American Sign Language (ASL) gestures into textual representations, facilitating seamless interaction. [Data Collection: We acquire a comprehensive dataset of ASL hand gestures from Kaggle](https://www.kaggle.com/code/drvaibhavkumar/sign-language-classification-using-cnn-acc-99). These images serve as ther training data.Pre-processing: Each image is transformed into pixel color indices, ensuring uniform representation across diverse lighting conditions and backgrounds.Convolutional Neural Network (CNN): We train a CNN model using the processed images. The model learns to recognize intricate hand gestures and their corresponding meanings.Noise Resilience: To address real-world challenges, we introduce noise during testing. Our model robustly converts noisy ASL images into accurate text, even in adverse conditions.Textual Output: The trained CNN model generates textual descriptions for each input gesture, allowing seamless communication.

[**Keywords**: Sign Language, CNN, Noise Resilience](https://www.irjmets.com/uploadedfiles/paper/volume3/issue_5_may_2021/10949/1628083441.pdf)

# Introduction

Speech impairment is a disability which affects an individual’s ability to communicate using speech and hearing. People who are affected by this use other media of communication such as sign language. Although sign language is ubiquitous in recent times, there remains a challenge for non-sign language speakers to communicate with sign language speakers or signers. With recent advances in deep learning and computer vision there has been promising progress in the fields of motion and gesture recognition using deep learning and computer vision-based techniques. The focus of this work is to create a vision-based application which offers sign language translation to text thus aiding communication between signers and non-signers. The proposed model takes video sequences and extracts temporal and spatial features from them. We then use Inception, a CNN (Convolutional Neural Network) for recognizing spatial features. The dataset used is the American Sign Language Dataset.

##### Methodology

2.1 Conv2D : Small kernels will go over the input image and detect features that are important for classification. Earlier convolutions in the model will detect simple features such as lines. Later convolutions will detect more complex features.

* 1. Batch Normalization: Like normalizing our inputs, batch normalization scales the values in the hidden layers to improve training.
  2. MaxPool3D: Max pooling takes an image and essentially shrinks it to a lower resolution. It does this to help the model be robust to translation (objects moving side to side), and also makes the model faster.
  3. Dropout: Dropout is a technique for preventing overfitting. Dropout randomly selects a subset of neurons and turns them off, so that they do not participate in forward or backward propagation in that particular pass. This helps to make sure that the network is robust and redundant, and does not rely on any one area to come up with answers.
  4. Flatten: Flatten takes the output of one layer which is multidimensional, and flattens it into a one-dimensional array. The output is called a feature vector and will be connected to the final classification layer.
  5. Dense: The first dense layer (512 units) takes the feature vector as input and learns which features will contribute to a particular classification. The second dense layer (24 units) is the final classification layer that outputs the prediction.

##### Literature review

Nowadays, people are not interested to speak in ASL when having a deaf relative or friend, or even classmate/acquaintance. Hence, deaf people are often trapped and isolated. ASL requires the use of a person’s hands so if something happens where a wrist was sprained and it disables that person from talking. For example, there was a mother who strained her wrist from signing all of her life for her deaf daughter. The doctor also made her stop signing. This caused the communication with her deaf daughter to decrease, since she had to read lips from then on.

ASL vocabulary dictionary contains thousands of sign just like words. It is very easy to get two completely different signs mixed up which leads to bad miscommunication. For example, the sign for “chocolate” and “cleve land” are similar, and they definitely don’t mean the same thing, or even close. It is very hard to follow when a conversation has and something gets mixed up.

Therefore to prevent this from happening this model is being developed.

##### Results

The present study demonstrates the efficacy of a machine learning model in converting American Sign Language (ASL) gestures into a coherent paragraph of text. This model exhibits robustness, capable of processing not only pristine visual inputs but also those afflicted with noise or extraneous data. This successful conversion from ASL to text represents a noteworthy contribution to the field of human-computer interaction, particularly with regards to bridging the communication gap between individuals who rely on ASL and those who do not.

##### Future scope

The existing **Sign Language to Text Conversion Model** faces critical limitations that necessitate further investigation. First, semantic coherence remains a challenge; occasionally, the model generates nonsensical phrases when combining words. Improving semantic consistency during translation is imperative. Second, under noisy conditions, the model’s accuracy requires enhancement. Robustness to environmental disturbances is crucial for real-world deployment. Lastly, while Convolutional Neural Networks (CNNs) serve well, exploring alternative algorithms like Artificial Neural Networks (ANNs) could yield superior results. Comparative studies are essential to identify the most effective approach. Addressing these gaps will advance sign language communication, fostering inclusivity and accessibility for all.

##### References

1. International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 9 (2018) pp. 7154-7161 © Research India Publications. Conversion of Sign Language into Text
2. Shubham Thakar, Samveg Shah, Bhavya Shah and Anant V. Nimkar. Sign Language to Text Conversion in Real Time using Transfer Learning