EE 541 – Computational Introduction to Deep Learning

Project Proposal

Abid Hassan

April 18, 2025

Project Title: CV for American Sign Language

Topic summary

The overarching goal of this project is to use deep neural networks to read American Sign Language. We will train the model to accurately classify hand signs that correspond to the ASL alphabet and commonly used gestures, as shown in the figure 1. ASL recognition is a critical first step in enabling real-time communication assistance for Deaf and hard-of-hearing individuals. ASL gestures involve complex hand articulations, subtle finger positions, and dynamic spatial-temporal cues that differ significantly from typical gesture recognition tasks. It is important to interpret such communication with a high degree of certainty, but humans tend to make mistakes in these scenarios. Our goal is to train a deep learning model to automate this process with high accuracy.

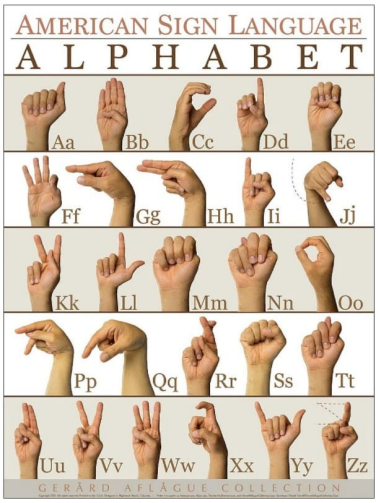

Figure 1: Alphabets of American Sign Language

Introduction

American Sign Language (ASL) is a primary mode of communication for millions of Deaf and individuals with hearing aids in USA. However, communication barriers often arise between ASL users

and non-signers, especially in real-time settings such as customer service, education, and healthcare. Automating the recognition of ASL gestures can bridge this gap, enabling more inclusive, accessible, and equitable communication.

A successful ASL recognition model would benefits those who are deaf or needs hearing aids, public servants and educators (like when president is address with person who is deaf), tv-reporters, software engineers to develop inclusive applications etc.

In academic setting this work directly makes contribution towards gesture recognition and human-computer interaction. Moreover, big tech giants like Google, Apple, and Microsoft can integrate ASL recognition into accessibility features for smartphones, smart glasses, or AR/VR platforms.

The foundation of this proposal lies in use of Convolutional Neural Networks (CNNs) for image-based hand gesture recognition. It involves training a model in supervised setting using cross-entropy loss function. The input data may need pre-processing/data augmentation, to improve the performance.

Related Work

There exist a plethora of work in this domain. Early work by Pigou et al. 2014 demonstrated the effectiveness of Convolutional Neural Networks (CNNs) for classifying isolated signs from hand gestures. Their work laid the groundwork for applying deep learning to visual-language understanding tasks. Building on this, Mollahosseini, Chan, and Mahoor 2016 used deeper CNN architectures for facial expression recognition, which, although not specific to ASL, introduced ideas relevant to interpreting fine-grained, person-specific gestures and expressions—both crucial components in ASL communication.

Koller et al. 2015 integrated CNNs with Hidden Markov Models (HMMs) to recognize continuous sign sequences, introducing a hybrid architecture that learned temporal dependencies on top of using spatial information. Later,Camgoz et al. 2018 proposed an end-to-end neural sign language translation framework using a combination of CNNs and Recurrent Neural Networks (RNNs) to interpret sign language in videos.

Dataset description

The dataset used for this project is available over kaggle and is of size 1.1 GB. The dataset contains images from 29 classes (26 alphabets, SPACE, DELETE and NOTHING). Each class contains 3000 images in the training set and each image is a 200 x 200 RGB image.

The dataset folder contains two sub-directories (one for train and one for test). DO NOT use the testing set for training. The training and testing directories contain one folder for each class. Each of the class folders has 3000 images in them. It is important to note that the test set only contains one image per each class. This is to encourage the usage of real world images.

We will measure model performance using classification accuracy.

Architecture Investigation Plan

Using more powerful and expressive model like EfficientNet, ResNext, Vision Transformers etc is beyond the score of this project. I will simply use ResNet-18 / ResNet-34 to classifying the ASL classification problem.

Estimated Compute Needs

I have personal Apple MacBook with 96 GB of RAM. I can simply use the computing capacity in my laptop to perform all the experiments.

Likely Outcome and Expected Results

In the case of a successful project, we expect our final output measure to demonstrate high classification accuracy ($\geq$95%) on held-out test sets of American Sign Language (ASL) alphabet. This accuracy would indicate that the model has learned distinguishable features from the input data.

The most likely reasons for project failure includes, lack of sufficient and diverse training data, model overfitting, poor choice of model architecture, or noisy labels.

References

Camgoz, Necati Cihan et al. (2018). "Neural sign language translation". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7784–7793.

Koller, Oscar et al. (2015). "Deep learning of mouth shapes for sign language". In: *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 85–91.

Mollahosseini, Ali, David Chan, and Mohammad H Mahoor (2016). "Going deeper in facial expression recognition using deep neural networks". In: *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, pp. 1–10.

Pigou, Lionel et al. (2014). "Sign language recognition using convolutional neural networks". In: *Lecture Notes in Computer Science* 8651, pp. 572–578.