
DS 5220 Project Proposal: Classifying ASL Signs using a Neural Network

Abhinav Nippani

Larson Ost

Adwait Patil

1 Problem definition

American Sign Language (ASL) is a visual language used by over 500,000 people in the United States and Canada, predominantly in deaf communities. For those without a reason to learn ASL, communication with their non-verbal peers proves to be a difficult language barrier. Now that video technology and video communication is increasingly more popular, we believe that we can help reduce this barrier by translating ASL speakers for their audience.

2 Why this is interesting

As previously mentioned, video technologies are significantly more used now than in previous years. This is interesting because we can use the video feed as a new form of information gathering to garner results from SML models, and it will be applicable to a wide range of platforms. It also lends itself toward the possibility of increasing accessibility technology for hearing impaired communities. In terms of our project group's personal interest for this project, none of us have ever had to learn or use ASL, so we find that it would be an interesting challenge to work with a new domain for this project.

3 Approach to solve the problem

We propose to tackle the problem using the steps below:

- **Data Augmentation:** We would first like to increase the data to include additional lighting conditions (dim or bright) and also blurring the images using appropriate filters. This will help us train the model on different lighting conditions allowing results to be reproducible in a dynamic live video.
- **Training a Model:** As this is a classification problem which uses image data, we will be training a neural network with convolutional layers. Additionally, we will also try to incorporate transfer learning and evaluate our results using appropriate metrics.
- **Model Testing:** We will test our model on both real time videos and still images. Real time video testing will be completed by taking a set number of frames to classify the signs and decide the optimal number by cross validation.

4 Why this is a good approach compared with other methods

We are working with a dataset comprised of images of ASL signs. Therefore we will be needing to reduce high-dimensional image data without losing information. Convolutional Neural Networks are the clear model of choice for this type of work. Other models such as decision trees would perform very poorly at feature processing and will be computationally demanding.

5 Key components and limitations

- The model will have two different set of results. First on the images of test data from ASL, and the second on the live video.
- We will be building two models: One from scratch, and another model using transfer learning. We will compare the results of each model on the testing data.
- Limitation: we are identifying letters and not words, so words can be expanded on in future projects.

References

[1] Akash. "Asl Alphabet." Kaggle, 22 Apr. 2018, <https://www.kaggle.com/datasets/grassknoted/asl-alphabet>.