Final project topic

CV for American Sign Language

EE 541: Fall 2023

Authors: Aditya Anulekh Mantri and Dr. Brandon Franzke

Introduction and Background

American Sign Language is a natural language that serves as a predominant sign language for the Deaf people in the United States of America. The goal of this problem is to build a deep learning model to read the American Sign Language. You will be using the dataset found on Kaggle to build your model.

Dataset

The dataset linked above 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.

Dataset Structure

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.

Suggested Approach

1. Take a look at the dataset and decide on the kind preprocessing required before training the model. The images in the dataset are manually captured and not computer generated. Think about any methods that can be employed to enhance the image quality if required.

2. Conduct a short literature survey about the work already done in this domain and understand the network architectures used in those models.

3. Experiment with different architectures and hyper-parameters. The results of the different experiments you conduct have to be included in your final report.

4. Given that the test set is very small for this problem, you must also include results of your own testing.

5. You are allowed to finetune (transfer learning) existing models that perform classification on RGB images. You must include details of the source model in your report.

   NOTE: You are not allowed to finetune existing models trained on the same dataset or similar datasets. We recommend checking with the teaching team before you decide on a model for finetuning.

Possible Extensions

- You can extend the project to build a small application to capture video and give a live transcription of the feed. This transcription can then be used to produce speech. Take a look at Google Text to Speech to add this functionality to your project.

References

[1] A Recipe for Training Neural Networks - A blog post by Andrej Karpathy. It is a fairly old blog, but definitely a must read before you start training neural networks.

[2] Kaggle Dataset