Internship Project Report

Trimas Christos
Sense Space Informatics
Real Time Sign Language Detection using Deep Learning

July 15, 2021

Introduction

Sign languages are languages that use visual-manual modality to convey meaning. Sign languages are expressed through manual articulations in combination with non-manual elements. There are hundreds of sign languages around the world, each one with their own grammar and lexicon. For the purpose of my summer Internship in Sense Space Informatics, I will be working with a custom made dataset of greek words. Since the goal was to implement a real time translator, the best option was to work with neural networks. After using Media Pipe Holistic, a Deep Learning solution provided by google as a feature extractor, the LSTM model was able to classify(or translate) the testing set, with an accuracy of 96%.

Feature Extraction

Google Media Pipe Holistic:

A sign language is a combination of hand and body movement. But in order to add more depth in the language the previous movements are combined with facial expressions. Media Pipe Holistic or MPC, provides a live perception of simultaneous human pose, face landmarks and hand tracking in real time devices. Therefore, it could be used as a Feature Extractor in our model. Details on how MPC works can be found here

Dataset

Data Collection:

A sign language contains a vast amount of words, some of them are represented the same way as others, but the intensity of the gestures changes. In order not to confuse the model with such things, and to keep it simple, 6 words related with the weather were chosen. Each word, was performed 10 times, and each video consisted of 30 frames. In conclusion, 6(images) x 10(videos each image) x 30(frames each video), for a total of 1.800 data points, stored and saved as numpy arrays.

Training/Testing

LSTM Model:

In the next phase of our project, we needed a Neural Network model to feed our dataset and train them. For this task a Long-Short-Term-Memory model was deployed.

LSTM is an artificial recurrent neural network architecture, used in the field of deep learning. Unlike standard feed-forward neural networks, LSTM has feedback connections, which makes it perfect candidate for time series problems that can have lags of unknown duration between important events, such ours.

So after a simple split of the data into training and testing, the training data were fed into the LSTM. As a loss function, the Kullback Leibler Divergence proved to have the best accuracy, between categorical cross-entropy and others. The train was complete after 184 epochs, with a batch size of 1024.

Results

Metrics - Real Time Translation:

To evaluate our model, accuracy was chosen among the various metrics that exist. Since our dataset has no imbalance, the accuracy metric is pretty solid by itself.
In the testing set, the model performed extremely well, achieving 96% accuracy.

The real test was in real time videos. Despite the small amount of data and training, the model performed pretty well, translating the words "vrochi" and "zesth" almost all the time correctly. The other words, since the were a little bit more complicating had an error more often.

Conclusion

Overall it was a fun project to do. I learned how to work with MPC, and tested a deep learning model in a real world time series problem.