American Sign Language Recognition

Hao Wang
hwang779@wisc.edu

Tianchang Li
tli289@wisc.edu

Jing Wen
jwen29@wisc.edu

1. Introduction

Sign language is a type of language important to certain type of people. Because of limited usage of sign language, the communication using sign language should be paid more attention. American Sign Language uses hand gestures to represent 26 letters. People's names, places, titles, brands, new foods, and uncommon animals or plants all fall broadly under finger spelling alphabet, and this list is by no means exhaustive. Due to this reason, the recognition process for each individual letter plays quite a crucial role in its interpretation [2]. Researches on ASL were widely conducted with deep learning. Human body pose estimation and hand detection are two important tasks for systems that perform computer vision-based sign language recognition(SLR)[3]. Oyebade K. Oyedotun and Adnan Khashman proposed applying deep learning to the problem of hand gesture recognition for the whole 24 hand gestures obtained from the Thomas Moeslund's gesture recognition database [4]. Their experiment showed that more biologically inspired and deep neural networks such as convolutional neural network and stacked denoising autoencoder are capable of learning the complex hand gesture classification task with lower error rates [3].

Our project will focus on recognizing images of the 26 letters in American Sign Language and 3 classes for SPACE, DELETE and NOTHING. Our group will convert sign pictures into numerical data set and obtain bias and weights by constructing a multiple-layer Convolutional Neural Network. After achieving sufficient accuracy on this step, we will then try to build a text analysis model which takes in a sequence of classifications from the first step and identify multiple-letter words and even sentences.

2. Motivation

Nowadays, American Sign Language (ASL) is used predominantly in the United States and in many parts of Canada by deaf communities. Even though ASL facilitates the communication among deaf people, the communication between deaf community and normal people is still limited by rare usage of ASL among ordinary people.

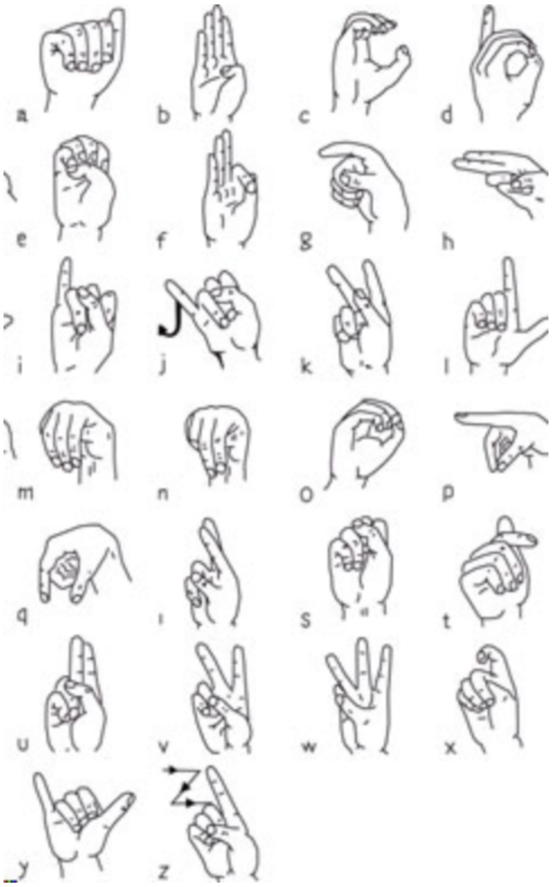

Figure 1. American Sign Language Alphabet

By figuring out algorithms that performs computer version ASL recognition, we hope that ASL of deaf community can be directly translated to English letters or classes for SPACE, DELETE and NOTHING. This will help people who cannot recognize American Sign Language to smoothly communicate with people with disability through computer version ASL recognition. The translation of ASL will also help beginners of this language to learn in the same way as subtitles of foreign movies do. We hope this algorithm to break the communication barrier between people with speaking disability and the rest.

Moreover, our group hope to apply knowledge from class to real life problems. Examining ASL with deep learning will also prove the capacity of convolutional neural network recognizing complicated gesture images with low error rates. except knowledge learnt from class, our group hope to learn more by optimizing our model with out class knowledge.

3. Method

There are two phases of this project. In the first phase, we will firstly read and convert the image data into RGB-channel matrices in Python. Then we plan on constructing a multiple-layer Convolutional Neural Network with convolutional layers, pooling layers, and fully connected layers. The graph displayed below shows a general idea of this model. There will be a few convolutional and max pooling layer alternate. Each layer will be followed with an activation function such as ReLU or softmax which will play an important role of adding non-linearity. The last one or two layers will be fully connected to maximize accuracy. The input layer with be a vector containing all pixels in one image. The output layer will consist of 29 classes. The weights and biases will be obtained from model training. We will split the dataset into three sets following the ratio of 6:2:2 and perform 3-way holdout evaluation method.

Once achieving a good accuracy, we will move on to the second phase where we will start building our text analysis model implementing Bayesian rule. More details will be discussed later but this model is aiming at auto-correcting words when a few letters were falsely classified in the first phase. This step is similar to auto-correction on phones keyboards.

Eventually if time permits, we will try to recognize sentences with similar Bayesian model. The expected high accuracy of first two models should make this step easier.

In the end, we will compare the performance of our models with the existing, well-developed image classification models and text analysis models in terms of accuracy and runtime. The difference will be reported.

4. Evaluation

The goal of this project is to recognize the letter or the class represented by the picture of American Sign Language. The performance of our models will be evaluated based on the predicting accuracy and run time. We trained with a categorical cross entropy loss function for both our data sets. Our goal is to reach 90% on letter and word classification and 85% on sentence recognition.

To generalize our model to real-life environment, we will also test by feeding in the pictures taken on our own hand gestures.

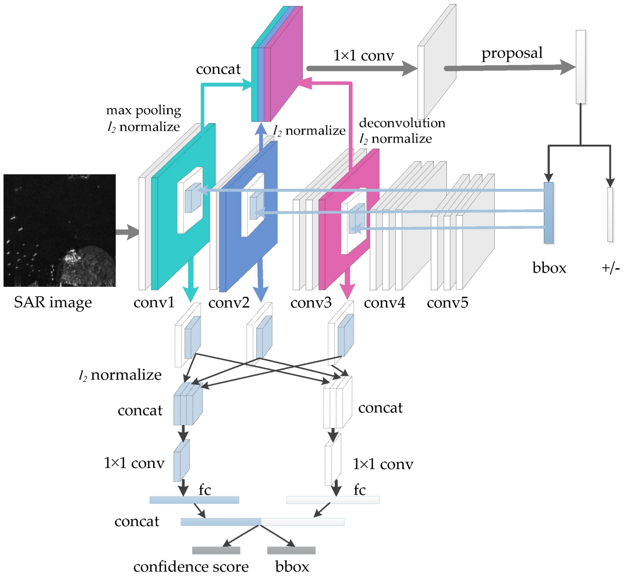

Figure 2. CNN model structure

5. Resources

Our group decided to use image data set for alphabets in the American Sign Language from Kaggle [1]. The data set includes 87,000 images which are 200x200 pixels. Images are evenly distributed into 29 classes, of which 26 are for the letters A-Z and 3 classes for SPACE, DELETE and NOTHING. These 3 classes are very helpful in real time applications, and classification. The images were not taken in the same place and time. The light, position, and background will be different for each pictures.

We will use the following computer hardware and computational tool:

MacBook Pro (2.7 GHz Intel Core i5, Early 2015)
MacBook Pro (2.3 GHz Intel Core i5, 2017)
Python 3.7, PyTorch 1.4.0

6. Contributions

The experiment and writing tasks were assigned to group members evenly. All of the group members will be involved with experiment and writing tasks. For Proposal writing, Jing Wen is focusing on motivation, resources part and the format of the proposal. Meanwhile, Hao Wang and Tianchang Li take charge to write introduction and evaluation while browsing related scholar articles online. Moreover, for experiment, Jing Wen and Tianchang Li will go ahead to import the picture data set with multiple-layer Convolutional Neural Network. Hao Wang will take fully charge on building our text analysis model implementing Bayesian rule. While we each doing our assigned parts of the projects, we will need to communicate with each other's part. In this case, mistakes can be more easily avoided and the project efficiency will be enhanced.

References

[1] Akash. Asl alphabet, Apr 2018.

[2] V. Bheda and D. Radpour. Using deep convolutional networks for gesture recognition in american sign language, 2017.

[3] S. Gattupalli, A. Ghaderi, and V. Athitsos. Evaluation of deep learning based pose estimation for sign language recognition. In *Proceedings of the 9th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, pages 1–7, 2016.

[4] O. K. Oyedotun and A. Khashman. Deep learning in vision-based static hand gesture recognition. *Neural Computing and Applications*, 28(12):3941–3951, 2017.