Proposal

Pooja Lodhi
Machine Learning Nanodegree

19th Jan 2019

Domain Background

Communication is one important requirement for living in society [1]. Similarly, for people who are deaf and mute have sign language as their mode of communication. It's tough for normal people to understand it. So it important to have application to understanding the sign and gesture then convert them back to language normal people can understand.

Gesture recognition has much application in improving human-computer interaction. There is various research done in this domain [1,5,6,7]. According to the literature review, there are many approaches to sign language recognition. It is mainly categorized into as sensor-based and vision based. Our main focus was on vision based. Further exploring vision based we came to know there are many different techniques or algorithm that can be used in vision-based approach. For example Feedforward ANN, BackPropagation, Convexity defect, Convex Hull, KNN, HMM, Adaboost and Haar, PCA, SVM, Kalman filter and K-Curvature [5].

Problem Statement

American sign language (ASL) is one of the natural language used for communication. This language has a different gesture for various alphabets. Thus design an application to understand the sign and gesture then convert them back to language normal people can understand is a good problem to solve.

The goal of this project would be to design a model which accurately predict the alphabet corresponding to ASL Gesture. Then to use the corresponding model to real-life data and find its accuracy for new and unknown data.

Datasets and Inputs

ASL dataset from kaggle is used for this project [3]. The size of the data is 1GB which has 8700 images. It is well structured in different folders. The image size is 200x200 pixels. There are 29 classes, of which 26 are for the letters A-Z and 3 classes for SPACE, DELETE and NOTHING. The Dataset is well balanced with 3000 images per class.

These images well structure to train a model to classify hand gesture to appropriate alphabet. It contains a train dataset folder and test dataset folder.

Dataset Size	1GB
Dataset Type	Dataset has image of jpeg format

Total Image Count	8700
Image Size	200X200 pixels
Classes	29
Images/Class	3000 images/class

Table 1: Dataset Summary

For testing, the dataset already has a test folder. This folder has one image that belongs to each class label. Other than that we would prefer images from real life and other sources for testing the model. For the validation set, we might use the 10% images from the test dataset. Which will contain the equal number of images from each class label.

Solution Statement

In this project we will be going to design a model which will classify the ASL Alphabets. The transfer knowledge will be used to design the model. There are various pre-trained models available [2][4]. But according to our dataset we don't know which pre-trained model will best suit our problem statement. Thus going to apply different model and decide on the basis of the accuracy which model to use.

Benchmark Model

Various results exist in the domain of sign recognition with different types of dataset. In [9], a Gabor Filter based method was used to identify 24 static ASL alphabet signs. It had an accuracy of 75% with a high confusion rate between r and u. In [8], researcher has proposed system proposed a system that uses depth images with a Kinect device. They used Random Forest learning model and got an accuracy of 90%. There are many other models exist with different dataset. Below is a table with small summary of models encountered by me. The dataset used in this model are not same as the one used by me. But they all tries to recognize ASL Alphabets.

Model used	Accuracy
Feed Forward ANN [5]	94%
Gabor Filter based method [9]	75%
Random Forest learning model [8]	90%
GMM-HMM [7]	90.8%

There are various kernels associated to this dataset on kaggle. I am going to run one of them and use its accuracy as benchmark model for my dataset. The links to some of them I am planning to use are provided in references [12].

Evaluation Metrics

As the dataset is a balanced dataset thus accuracy is one of the good evaluation metric. Accuracy in classification problems is the number of correct predictions made by the model over all kinds predictions made[10].

Classification accuracy = number of correct predictions / all kinds predictions made

We can also use Log Loss metric to train our model well. Log Loss takes into account the uncertainty of your prediction based on how much it varies from the actual label[11]. This gives better view about the performance of given model. It is good for multi-class classification.

K-fold cross validation is a process that can be used to estimate the quality of a neural network. It is widely used to check whether a model is an overfit or not. If the performance metrics at each of the k times modelling are close to each other and the mean of metric is highest.

There are many kernels linked to the dataset on the kaggel website. So we can compare our model with them.

Project Design

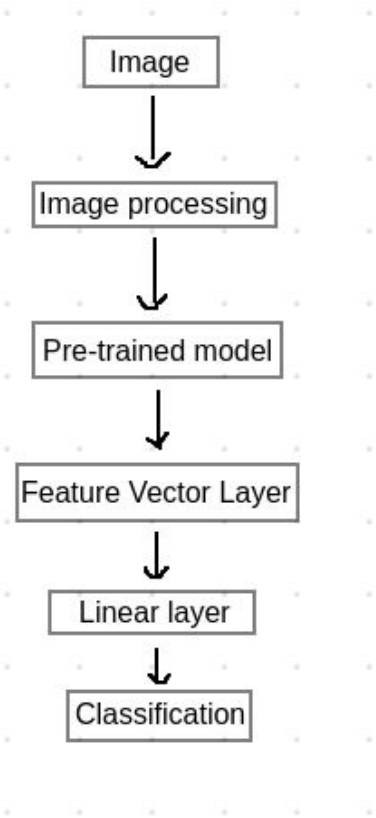
The workflow for the project would be discussed in this section. But to get better results would prefer to train the model with data augmentation also.

Workflow:

- Exploring the data
- Data preprocessing
    - Resize the image
    - Data augmentation
- Evaluate various algorithms
    - Build a different model : The pretrained model thinking to use are VGG, Xception, Inception
    - Select best model
    - Make prediction on validation set
- Make prediction on testing set.
- Conclusion

The flowchart below provides an overview of how the model is going to train. The image dataset is well structure so we will first mix the dataset so that different images are fed to the network while training and network get well trained. With validation set in between.

After the model gets trained we will find its accuracy with help of test dataset. Then, at last, select the best model and test it with some new real-life images to identify how it performs in real time.



References

[1] Sanil Jain, K.V.Sameer Raja “Indian Sign Language Character Recognition”

[2]https://www.analyticsvidhya.com/blog/2018/07/top-10-pretrained-models-get-started-deep-learning-part-1-computer-vision/

[3] https://www.kaggle.com/grassknoted/asl-alphabet

[4] https://keras.io/applications/

[5] Shivashankara S, Srinath S “A comparative Study of Various Techniques and Outcomes of Recognizing American Sign Language: A Review”, International Journal of Scientific Research Engineering & Technology, September 2017

[6] Nachamai. M, “Alphabet recognition of American Sign Language: Hand gesture recognition using SIFT Algorithm”, International Journal of Artificial Intelligence & Applications (IJAIA), Vol.4, No.1, January 2013

[7] Brandon Garcia, Sigberto Alarcon Viesca, “Real-time American Sign Language Recognition with Convolutional Neural Networks”

[8] Cao D., Ming C.L., Zhaozheng Y. American Sign Language alphabet recognition using microsoft kinect; Proceedings of the 2015 IEEE Conference on CVPRW; Boston, MA, USA. 7–12 June 2015

[9] N. Pugeault, and R. Bowden. Spelling It Out: Real-Time ASL Fingerspelling Recognition. 2011 IEEE Workshop on Consumer Depth Cameras for Computer Vision, pp. 1114-1119, 2011

[10]https://medium.com/greyatom/performance-metrics-for-classification-problems-in-machine-learning-part-i-b085d432082b

[11] http://wiki.fast.ai/index.php/Log_Loss
[12] Benchmark Models Thinking to use:
https://www.kaggle.com/danrasband/classifying-images-of-the-asl-alphabet-using-keras
https://www.kaggle.com/dansbecker/running-kaggle-kernels-with-a-gpu
https://www.kaggle.com/kairess/99-9-asl-alphabet-classification-with-slimcnn