Capstone Project Proposal

Bahi Ali

Machine Learning Engineer Nanodegree

I. Domain Background

In our modern society, data shapes our reality, our understanding of the world and our everyday life. We use it to do and understand things that we can't usually do easily. One of those areas of interest is the visual recognition and understanding the content of the media and get useful insights and information out of them, this can help us in many different fields like Health-care as we can diagnose diseases and self-driving cars to detect and recognize objects and so much more and visual recognition has proved to be very important in those fields. Also, one of the fields we can use visual recognition it is the Sign languages, - One of the ways we convey and deliver messages without explicitly speaking, by moving hands and making gestures, movements. As the orientation of the fingers arm and body convey a different message.

After doing some research, I found this project on GitHub which also, deals with sign language.

This paper this paper to try solving the same problem but using edge detection ,this paper also solves the same problem using CNN as got a very good accuracy 98% so I will be very helpful.

II. Problem Statement

In signed languages, we can express numbers in a certain way by giving certain gestures that represent a corresponding digit. For this project, I'll be using this dataset to train a model to recognize the number represented in an image by sign language into its respective digit from 0 to 9 using CNN architecture with Keras.
The purpose of this project is to create a classification model that's capable of taking an image as an input and predict its respective digit in the sign language.

III. Dataset and Input

For this project, I'll be using this dataset from kaggle, The dataset contains 2180 images of hand gestures that represent a certain digit. This will be the main catalyst for this capstone. As for the inspiration behind the data, this data was collected and preprocessed in Turkey Ankara Ayrancı Anadolu High School by taking pictures of the student's hands.

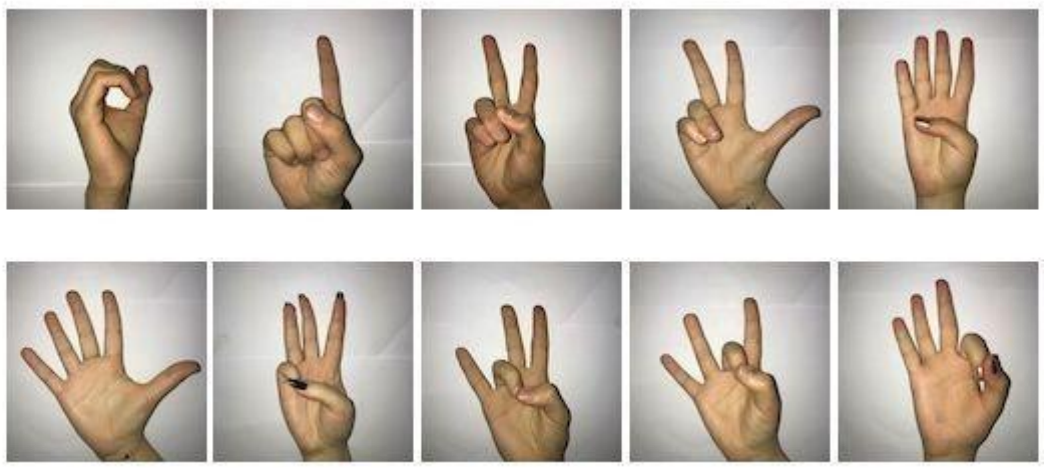
As for the content of the dataset:

- Image size: 64x64
- Color space: Grayscale
- File format: npy
- Number of classes: 10 (Digits: 0-9)
- Number of participant students: 218
- Number of samples per student: 10

Class distribution:

- Class 0 (0 digit) : 204
- Class 1 (1 digit) : 205
- Class 2 (2 digit) : 206
- Class 3 (3 digit) : 207
- Class 4 (4 digit) : 206
- Class 5 (5 digit) : 208
- Class 6 (6 digit) : 207
- Class 7 (7 digit) : 206
- Class 8 (8 digit) : 206
- Class 9 (9 digit) : 207

As we can see, the dataset is balanced, which helps a lot with the metrics and we can use the accuracy metric. The images look like this:



As for Input, it'll be an array of squared pixels (number of rows = number of columns. The dataset consists of 2 NumPy arrays, one for the X-axis and one for the Y-axis.

IV. Solution Statement

What I have in mind is to build a model and train a machine learning model that can recognize the content of an image and classify it to its corresponding digits from zero to nine. When I thought about which algorithm or model I'll be using, so I thought I might use a variety of algorithms that I saw in the nanodegree and learned on my own, like SVM, Logistic Regression and Finally CNN, so that I can compare between many algorithms and use keras and SciKit-Learn.

As for the CNN, I'll try to tweak the number of hidden layers to get the highest possible accuracy from it.

V. Benchmark Model

I'll be using different models in this project, so I'll be comparing the models implemented in the project and the one with the highest accuracy will be benchmarked against other models in the kaggle leaderboard to check whether my model got higher or lower accuracy.

Since kaggle isn't an option in this project I will use SVM as benchmark just like you stated.

Running the SVM model got the accuracy : 43.5%

VI. Evaluation Metrics

I'll be evaluating the model using different metrics like the F1-Score, training time, prediction time and accuracy to calculate the performance of my model.

VII. Project Design

I’ll be starting with doing some exploratory data analysis and visualization to have a better understanding of the data and get more insights about it.
I don’t think I’ll be needing to do any preprocessing to the dataset because it has the same size and type (numpy arrays, in grayscale with a size of 64*64). But I’ll normalize the pixel values by dividing on 255 and then I’ll do some augmentation like rotation, skewing and adding noise to the images to enhance the model and learn the features.

After that, I’ll be splitting the data into training, testing and, validation sets. Then I’ll start building the model, starting with Support Vector Machine (SVM), then the logistic regression and finally the CNN model.

Having all the models built, I’ll compare their performance as I mentioned in the evaluation metrics section using accuracy metrics and other algorithms mentioned above.

Review Link