Capstone Proposal

CNN based detection of distracted drivers

Pranav Vaidik Dhulipala

June 24th, 2018

1 Domain Background:

As drivers, we all know that distracted driving is a dangerous maneuver, not only to you and your fellow passengers, but also the other vehicles around you. We have seen several cases of distracted driving everyday: a vehicle on a freeway suddenly slows down and swerves from side-to-side, affecting the entire traffic flow. Or a car in front doesn't budge even when the lights turn green.

On passing these vehicles, we often spot the drivers seemingly distracted by their mobile devices, social media or fiddling with the radio. Though unsurprising, this behavior on road is not only inconvenient to other drivers but also very dangerous.

More than 3,000 people are killed and 425,000 people are injured by distracted driving every year according to CDC motor vehicle safety division.

According to National Highway Traffic Safety Administration (NHTSA), distracted drivers were involved in the motor vehicle crashes which resulted in death of 3450 people in 2016 [2]. Also, 9% of the total fatal crashes occurred in the year 2016 were reported to be caused by distracted driving [2].

Many states in the US prohibit drivers from using hand held devices during driving, in order to prevent distracted driving. State Farm insurance company has hosted a competition on Kaggle website in hopes to improve these alarming statistics by testing if distraction behaviors of drivers can be automatically detected using dashboard cameras.

2 Problem Statement:

For this project we aim to train a machine learning model that can predict the likelihood of what the driver is doing from an image. The model is trained using a set of manually labeled images of drivers doing something while driving in a vehicle. The activities are classified to 10 different classes such as texting, eating, talking on the phone etc., where the machine learning model is expected to output the likelihood of the driver doing one of the following 10 pre-defined activities:

- c0: safe driving
- c1: texting - right
- c2: talking on the phone - right
- c3: texting - left
- 4: talking on the phone - left
- 5: operating the radio
- c6: drinking
- c7: reaching behind
- c8: hair and makeup
- c9: talking to passenger

The model should output the predicted activity with highest predicted probability and output the probability value as the confidence score, expressed in percentage.

3 Datasets and Inputs:

The datasets are provided by Statefarm in the kaggle competition website [1]. The dataset consists of color images which are already divided to train and test sets.A CSV file is also available, containing the labels for all the 22,434 images available in the train set.

The train set consists of 23,434 images of same size, with 26 different drivers. The distribution of the number of images in each class is described by the histogram in Figure 1. The distribution of images for drivers is also not uniform, observed from the histogram in figure 2.

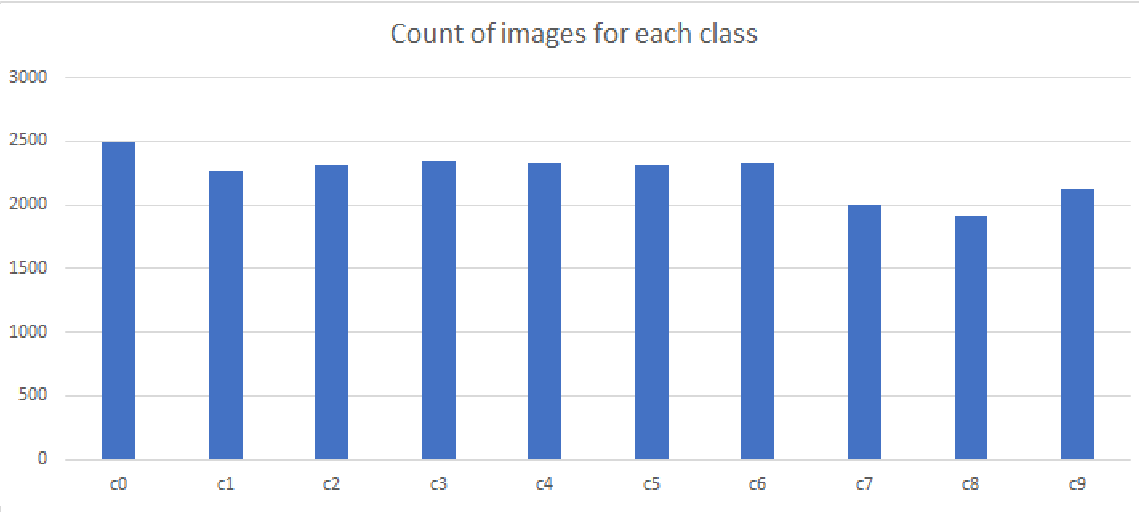

Figure 1: Distribution of classes in train set

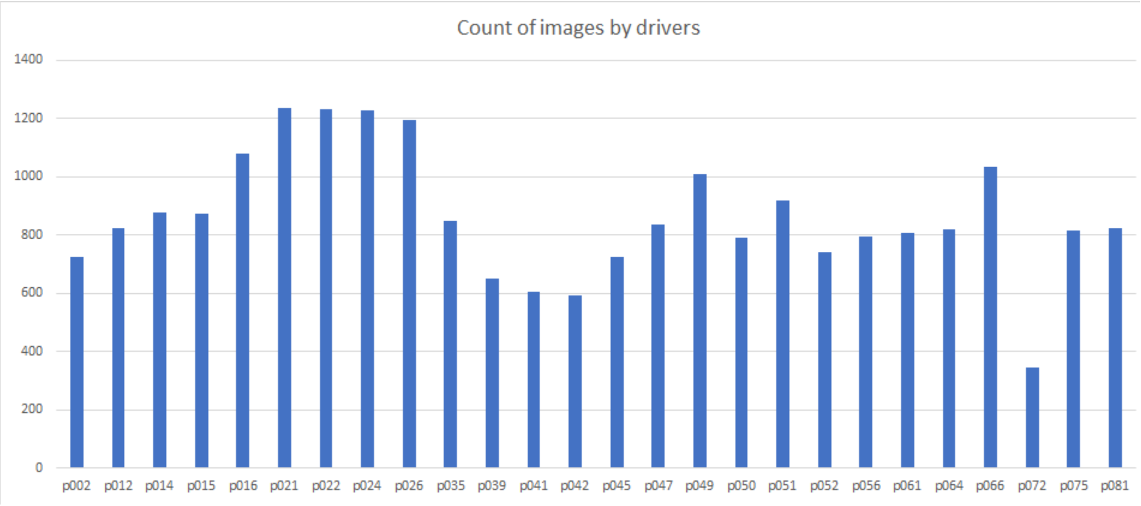

Figure 2: Distribution of drivers in train set

The labels for the test set were not provided, so we would be splitting the train set to obtain validation and test sets for this project to evaluate our machine learning model.

4 Solution Statement:

The proposed solution for the problem is to use a Convolutional Neural Network model to classify the images. The model should take a color image as an input and output the predicted activity of the driver. The probability of the predicted activity expressed in percentage is also obtained as output, and shown as confidence score.

For training and validation purposes, the dataset is split into training, validation and testing sets, consisting of 80%, 10% and 10% of the data respectively.

CNN is constructed using a transfer learning approach, where we start with a *resnet*50 model and fine tune it with our train dataset. Fine tuning is a necessary step here as our dataset is fairly different from the ImageNet dataset.

Since the driver seats are adjustable (seat distance from steering and back rest positions) and the drivers can be in different sizes, scale, translation and rotation invariance are valid concerns in the training process. Hence, to ensure scale and translation invariance in the model data augmentation is used for training.

5 Evaluation Metric:

Submissions are evaluated using the multi-class logarithmic loss. Each image has been labeled with one true class. For each image, you must submit a set of predicted probabilities (one for every image). The formula is then,

A multi-class logarithmic loss is used to evaluate the model, defined by equation 1.

$$logloss = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M} y_{ij}\log p_{ij} \tag{1}$$

where N and M are the number of images in test set, image class labels respectively, log is the natural logarithm, y_{ij} is 1 if i^{th} image in the test set belongs to class j and 0 otherwise, and p_{ij} is the predicted probability that i^{th} belongs to class j.

In order to avoid the extremes of log function in equation 1 (i.e. $p_{ij} = 0$ or 1), we replace the predicted probabilities with $\max(\min(1 - 10^{-15}, p_{ij}), 10^{-15})$.

In addition to the *logloss*, accuracy score is also used to evaluate the method to show how accurate the model is.

6 Benchmark Model:

I am using a vanilla CNN model as the benchmark model for this project. The model will consist of a CNN layer followed a *maxpool* layer and a *softmax* activation layer.

The sample benchmark log-loss provided by the competition was 2.302, which is very high. Therefore, for this project, I choose to use the submission from the top 10 percentile in the leader board as the benchmark result, which is 0.256.

7 Project Design:

For this project, I aim to construct and train a CNN model that can classify the activity of the driver from a picture into a set of predefined activities, with a confidence score. The model takes a color image as input and the predicted activity as output with an prediction confidence value. Preprocessing steps would consist normalizing the color images. We also add a resizing step in order to make sure that the model would work for images with different sizes as well.

In order to build the CNN model, we use transfer learning procedure starting with *resnet*50 model, which we fine-tune using the train set. For training, we augment the training data in order to make sure the model works for people of different sizes and seat positions, as people can be of different sizes and the driver seats are usually adjustable in the vehicles.

A multiclass *logloss* score is used to evaluate the model. An accuracy score is also additionally used to evaluate the model.

References

[1] *State Farm Distracted Driver Detection — Kaggle*. 2016 (accessed May 26, 2018). URL: https://www.kaggle.com/c/state-farm-distracted-driver-detection.

[2] National Center for Statistics and Analysis. "Distracted driving 2016". In: *Traffic Safety Facts Research Note. Report No. DOT*. Vol. HS 812 517. Washington, DC: National Highway Traffic Safety Administration, April 2018.