Distracted Driver Detection

Wrik Bhadra
MT18027

Suraj Pandey
MT18025

Arjun Tyagi
MT18023

IIIT Delhi, New Delhi 110020, India

Problem Statement

Distracted driving is a global public health concern that is largely preventable. According to a study [1] conducted by NCBI-NIH, Govt. of India, as much as 44.7% of motor vehicle collisions were due to the use of mobile phones alone. Other forms of distraction include adjusting one's hair, looking in the rearview mirror for a prolonged time, peeking outside windows and others.

In this project, we aim to detect such distracted drivers from among a set of actions portrayed in each sample of the image dataset provided as part of an already-concluded Kaggle competition [2].

Literature Review

Distraction can be broadly categorized into threetypes: visual (eg. taking one's eyes off the road), cognitive (eg. taking one's mind off driving) and manual (eg. taking one's hands off the steering wheel). The focus of this project is on manual distractions. In order to process images corresponding to this specific problem, CNNs (Convolutional Neural Networks) have been tested [2] as the most state-of-art technique and has led to efficient performance on visual recognition.

Another popular model that has been used is VGGNet, which is a pre-trained deep CNN. In 2017, Abouelnaga et al. [3] created a new dataset similar to StateFarm's dataset for distracted driver detection. Authors preprocessed the images by applying skin, face and hand segmentation and proposed the solution using weighted ensemble of five different CNNs.

For the specific distracted driver detection task, most of the top teams on Kaggle Leaderboard have tried [2] ResNet and Inception models; then ensembled different models.

Dataset Description

The dataset consists of 100k images among which 79k (unlabelled) are for testing and 22k labelled images are for training. For our project, we shall divide the labelled images as 70% for training and 30% for testing. *Table 1* gives a concise description of the dataset.

Class Symbol	Class Name	Number of Samples	Sample Class Image
C0	Normal Driving	2490	
C1	Texting (right)	2268	
C2	Talking on the phone (right)	2318	
C3	Texting (left)	2347	
C4	Talking on the phone (left)	2327	
C5	Operating the radio	2313	
C6	Drinking	2326	
C7	Reaching behind	2003	
C8	Hair and makeup	1912	
C9	Talking to passenger	2130	

Table 1: Class labels alongwith their description, count and sample image

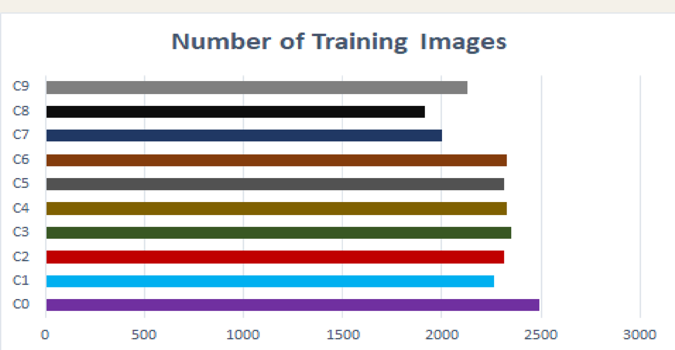


Figure 1: Class-wise data visualization

Activity Workflow

The primary steps followed in order to achieve the classification task is given below:

Pre-processing → Feature Extraction → Training → Validation → Prediction

Baseline Model

The existing baseline models are based on classic machine learning techniques such as Support Vector Machine (SVM), Logistic Regression (LR) etc.

Proposed Algorithm

Pre-processing
Balance dataset
Histogram Equalization

Feature Extraction

Feature dimension:
4096(AlexNet), 4096 (VGG-16)

These features are combined and projected to 1024 dimension using PCA (Principal Component Analysis)

Training
For SVM, use RBF (Radial Basis Function) kernel
For CNNs, use FC6 layer of AlexNet and FC7 layer of VGG-16 models
Additionally, an ensemble of AlexNet and VGG-16 was done.

Validation

Prediction

Work Done

A. Pre-processing
The training dataset consists of imbalanced class samples and images with different illumination. This is addressed by manually removing some of the images from training data. The illumination is handled by using histogram equalization.

Additionally, number of images in each class varies. Thus we select the least count among all and use this number to select the number of datapoints for each class.

B. Feature Engineering

- Intensity value
- Histogram value
- Haar wavelets
- Histogram of Oriented Gradient (HOG)
- Local Binary Pattern (LBP)
- Features derived from pre-trained AlexNet model
- Features derived from pre-trained VGG-16 model

We used these features individually and in conjunction to train and test our data using different models.

C. Classifiers
For classification, we use the following classifiers alongwith the features mentioned in the previous section.

- Gaussian Naïve Bayes (GNB)
- Logistic Regression (LR)
- Support Vector Machine (SVM)
- Multi-class Adaboost
- Bagging (Bootstrap Aggregating)

Application

With the aim of reducing road crashes, detection and prevention of distracted driving can have a major impact in enhancing road safety.

To improve upon the system, an alarm may be sounded when the system detects the driver being detected. This might also alert individuals in the vicinity to be aware.

Tools and Technologies

- Language: Python
- Libraries
  - Scikit-learn
  - OpenCV
  - PyTorch
  - Scikit-image
  - Seaborn

Evaluation metric

Standard accuracy and multi-class log-loss value

$$logloss = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M} y_{ij} log(pij)$$

where,
N: number of images in test set
M: number of image class labels
p_{ij}: predicted probability that observation i belongs to class j
y_{ij} = 1 if observation i belongs to class j; 0 otherwise

Results & Inferences

Following are the set of test accuracies obtained for different classical machine learning classifiers:

Classifier	Accuracy (%)				
	Intensity	Histogram	HOG	LBP	Haar
Naïve Bayes (NB)	49.69	19.34	75.4	19.15	54.11
Logistic Regression (LR)	71.39	21.77	73.46	19.01	77.48
Support Vector Machine (SVM) [RBF kernel]	61.75	61.41	75.92	21.31	**79.39**
Adaboost	45.20	21.56	57.12	44.54	48.29
Bagging	43.10	18.56	65.12	33.45	37.49

Following are the set of test accuracies obtained for deep learning classifiers:

Classifier	Accuracy (%)
VGG-16	84.29
AlexNet (RBF Kernel +SVM + PCA)	95.19
Ensemble (VGG-16 + AlexNet)	**98.22**

The results for some classifiers using different features were uploaded on Kaggle and the following multi-class logarithmic loss scores were obtained as summarized below:

Classifier	Feature	Kaggle score
LR	**HOG**	**2.01986**
SVM	Wavelet	2.59980
LR	HOG + LBP	5.57286

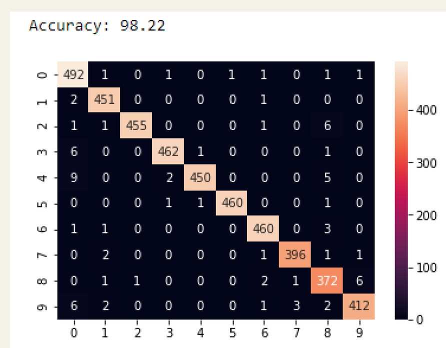


Figure 2: Confusion Matrix for Ensemble model

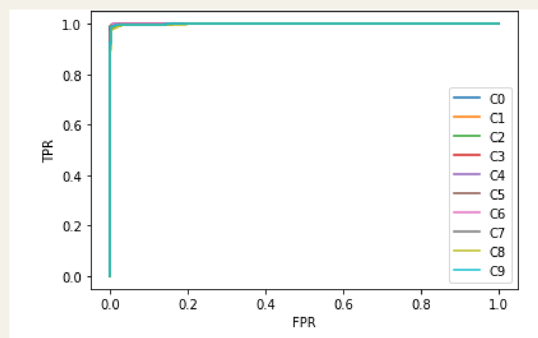


Figure 3: ROC curve for Ensemble model

Observations & Inference

- It was observed that for classification using classical machine learning models, SVM using Haar wavelets gave the most accuracy for the given dataset.
- For deep learning models, ensemble of VGG-16 and AlexNet performed best.
- Also, for the submissions made on Kaggle, Logistic Regression model using HOG features gave the lowest multi-class log-loss value.

Conclusion

Using classical machine learning models like SVM, LR, NB, we found the combination of LR along with HOG features to give the best log-loss value of 2.01986.

However, when we switched to CNN models like AlexNet, VGGNet etc, we found VGGNet and AlexNet to perform best, i.e. they give higher accuracy.

The accuracy can perhaps be further improved by segmenting the person and his/her action before extracting features and feeding to the classifier during training. Further, accuracy might improve if we add more layers to our deep learning models.

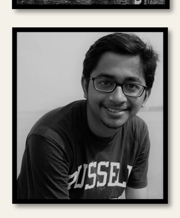
References

[1] Natasa Zatezalo, Mete Erdogan, and Robert S Green. Road traffic injuries and fatalities among drivers distracted by mobile devices. Journal of emergencies, trauma, and shock, 11(3):175, 2018.
[2] State Farm Distracted Driver Detection challenge on Kaggle. Internet: https://www.kaggle.com/c/state-farm-distracted-driver-detection, April 2016 [April 2019]
[3] Yehya Abouelnaga, Hesham M Eraqi, and Mohamed N Moustafa. Real-time distracted driver posture classification. arXiv preprint arXiv:1706.09498, 2017.

Team members

ARJUN TYAGI is pursuing Master of Technology in Computer Science & Engineering at IIIT Delhi (2018 - 2020). His interests are in Computer Vision and Visual Tracking.
Email: arjun18023@iiitd.ac.in

SURAJ PANDEY is pursuing Master of Technology in Computer Science & Engineering at IIIT Delhi (2018 - 2020). His interests are in Computer Vision and Data Augmentation.
Email: suraj18025@iiitd.ac.in



WRIK BHADRA is pursuing Master of Technology in Computer Science & Engineering at IIIT Delhi (2018 - 2020). His interests are in Computer Vision and Artificial intelligence.
Email: wrik18027@iiitd.ac.in