Distracted Driving Recognition: Classifying Safe and Unsafe Driving With Deep Learning

James Carroll and Aakarshan Dhakal

Introduction

Distracted driving is an epidemic in the United States; *every day* about 1,000 people are injured and nine are killed from car accidents involving distracted driving[1]. Our project aims to identify distracted driving via dashboard cameras. With State Farm's distracted driver dataset, we used three convolutional neural network models to distinguish between safe and distracted driving. Our models were not able to recognize some distractions such as drinking or applying makeup, but they were very good at identifying texting and talking on the phone.

Data

Our dataset (published by State Farm) comes from Kaggle and contains 22,424 labeled images of safe and distracted drivers. There were 26 drivers in total, with each driver appearing multiple times in each of the 10 image classes. The 10 classes were: safe driving, texting (left/right hand), talking on the phone (left/right hand), operating the radio, drinking, reaching behind, applying makeup, and talking with a passenger. Some example images are shown below.


Safe driving


Texting


Reaching behind

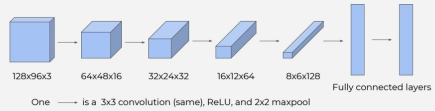

Talking on phone

Models

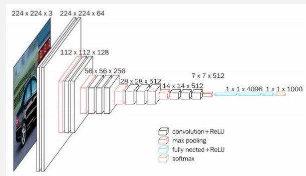
We used three models in total: a 4 layer CNN, a pretrained VGG-16 model[2], and a pre-trained InceptionV3 model[3]. We modified the basic CNN from the class repo to fit our image resolution and also added L2 regularization, dropout, and data augmentation code to prevent overfitting the training data. For both the VGG-16 and InceptionV3 models, we retrained the last fully connected layer and final softmax activation layer to better fit our dataset. For all of our models, we calculated the loss using cross entropy with softmax.

Model Architectures

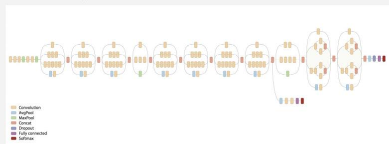
Basic 4-layer CNN model architecture



Pre-trained VGG-16 model architecture



Pre-trained InceptionV3 model architecture



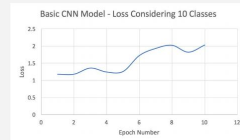
Results

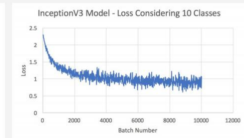
Due to computational and time constraints, we were only able to run the VGG-16 model on three classes (safe driving, texting/talking with right hand). Below are comparisons of the three models' performances on these three classes.

Model	Train Accuracy	Dev Accuracy	Loss	Epochs Run
Basic CNN	99.9%	97.3%	0.074	10
VGG-16	98.9%	83.6%	-	20
InceptionV3	97.5%	92.1%	0.427523	500

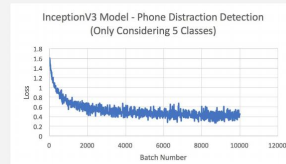
Results (cont'd)

Our results when considering all 10 classes were not ideal. Our best dev accuracy for the basic CNN was 78.6% and for the InceptionV3 model it was 73.5%.





We shifted our focus to identifying just 5 classes (safe driving, talking/texting with either hand) to build a phone distraction detector. We were able to achieve 87.4% dev accuracy.



Discussion

We performed error analysis on our InceptionV3 model's mislabeled images. The most common mistake was classifying talking to passengers as safe driving. We believe this error, and several other errors, were due to the fact that it is difficult to distinguish between some of the classes. While our models were not able to achieve high accuracy on all 10 classes, we were able to build a good phone distraction detector. In the future, we look forward to accessing better computational resources to test the VGG model on all 10 classes, as well as run all of our models for much longer.


Talking to a passenger (classified as safe driving)

References

[1] Centers for Disease Control and Prevention (2017). *Distracted Driving*. [online] Cdc.gov. Available at: https://www.cdc.gov/motorvehiclesafety/distracted_driving/ [Accessed 7 Jun. 2018].
[2] Moindrot, Oliver (2017). Example TensorFlow script for fine-tuning a VGG model (uses tf.contrib.data), Available at: https://gist.github.com/omoindrot/dedc857cdc0e680dfb1be99762990c9c/ [Accessed 5/27/2018]
[3] d'Almeida, Wisdom (2017). Image-classification-transfer-learning, Available at: https://github.com/wisdal/Image-classification-transfer-learning/blob/master/retrain.py/ [Accessed 5/29/2018]