Predicting Distracted Driving

Tracy Cardwell

Distracted Driving Statistics

National Highway Traffic Safety Administration Distracted Driving Estimates (2018)

- 8% of all fatal crashes
- 15% of all non-fatal crashes
- 8 deaths every day
- 1000 injuries every day

Estimates may be low:

- Self-reporting
- Difficult to determine distraction in driver fatality cases

What is Distracted Driving?

Driving while doing any of the following:

- Looking at something not related to driving
- Taking one or both hands off the steering wheel
- Taking your mind off driving

Hand held cell phones are the most dangerous as they combine all three.

A five second text message at 55 mph is like driving the length of a football field with your eyes closed!

What Can Be Done?

Distracted Driving Laws in 2020

- 21 states ban hand-held cell phone use
- 48 states ban texting for all drivers
- 39 states ban all cell phone use by novice drivers
- 20 states ban all cell phone use by school bus drivers

In a AAA Foundation for Traffic Safety survey in 2013, many drivers admitted using phones while driving despite acknowledging the danger and laws.

What Can Be Done?

What if:

- Drivers could get a discount on their auto insurance for agreeing to use a distraction warning system?
- Or avoid a traffic violation by agreeing to use a distraction warning system?

The first step to such a system is detecting distractions. This project uses dashcam images to train deep learning models to predict one of ten categories:

- Nine types of driver distractions, or
- Safe driving

Driver Image Data

State Farm simulated distracted driving and created a dataset with ten categories.

- c0: safe driving
- c1: texting – right hand
- c2: talking on the phone – right hand
- c3: texting – left hand
- c4: talking on the phone – left hand
- c5: operating the radio
- c6: drinking
- c7: reaching behind
- c8: hair and makeup
- c9: talking to passenger

Data Exploration

- 22,424 labeled training images
- Ignored unlabeled test images
  - Need labels for model evaluation
- 640 x 480 RGB
- 26 drivers
- Several different cars
- Different camera locations
- Extraneous data unrelated to distractions
  - Backseat, passengers, dash, console

Data Analysis

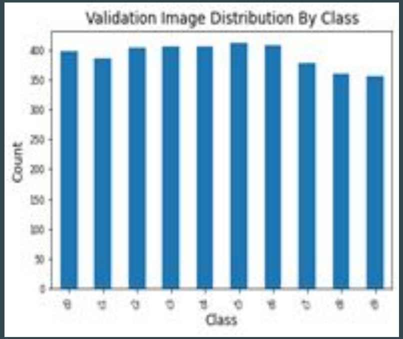
- Class distribution fairly even
- Range from 8.5% to 11% of total dataset
- All drivers represented in each class
- Driver images per class uneven
  - 0.6% - 19.8%
  - Most have less variation
  - Few driver/class combinations with high variance
- Total images per driver: 1.5% - 5.5%

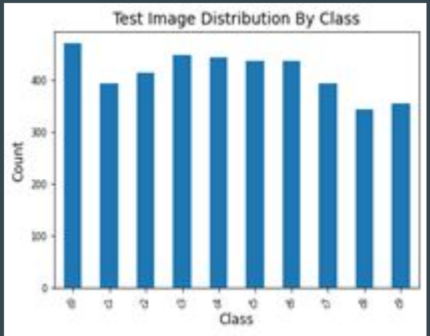
Data Wrangling

- Cropped images to reduce unneeded data
- Original: 640x480; Cropped: 470x400

Data Wrangling

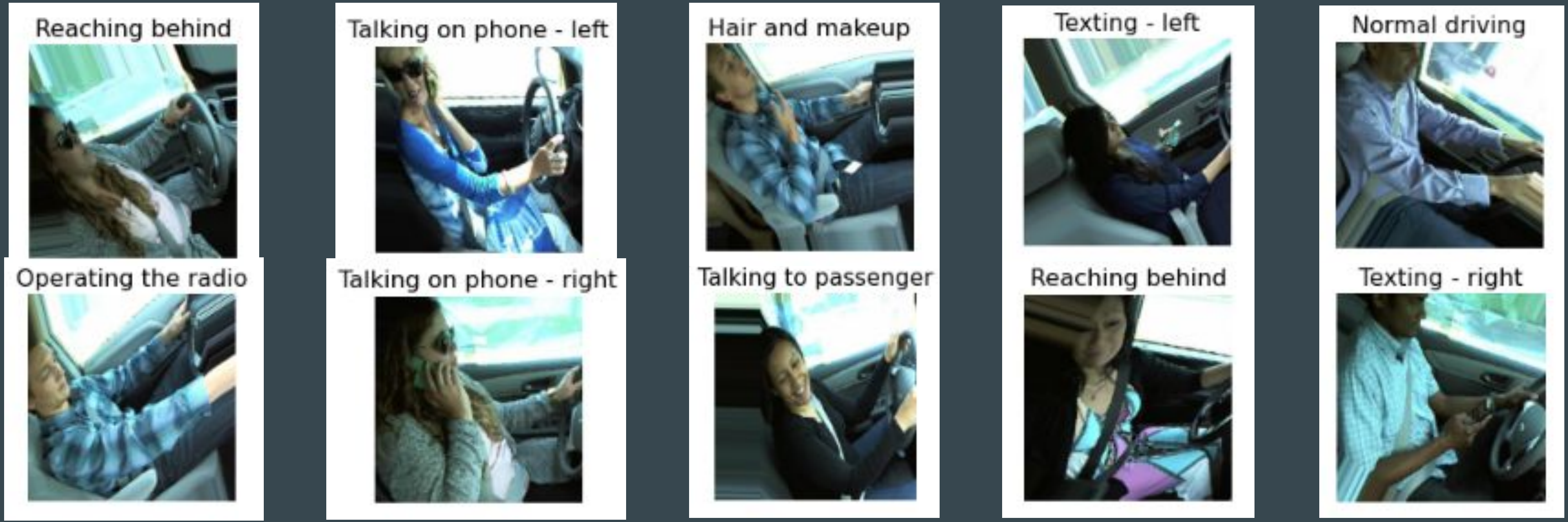
- Train/Validation/Test Split by Driver
- Ensure models generalize well to new drivers
- Goal was 60/20/20
- Actual was 64.2/17.4/18.4 due to splitting by driver, uneven driver representation in data





Data Augmentation

- ImageDataGenerator augmentation used due to small training dataset
- More training data and avoid overfitting



Deep Learning

Initial Plan

1. Start simple, increase complexity as needed for performance
2. Simple Convolutional Neural Network (CNN) with no image augmentation
3. More Complex CNN with no augmentation
4. Complex CNN with augmentation
5. Transfer Learning
    a. ResNet50
    b. MobileNet V2
    c. Inception V3
    d. VGG19

Training Parameters

- Batch size 32 for all training
  - Larger batch sizes yielded larger loss
- 10 - 30 epochs
  - Resource and time constraints
  - Early Stopping
- Stochastic Gradient Descent with learning rate decay
  - More volatile loss with other optimizers, particularly Adam
  - Lower learning rates generally performed better, more stable training

Models

Simple Model

Accuracy	Validation Loss	Range of Class F1-scores	Epochs	Learning Rate (Adam)
.43	3.5	.18 - .57	10	.0005

More Complex Model, No Augmentation

Accuracy	Validation Loss	Range of Class F1-scores	Epochs	Learning Rate (SGD)
.48	2.3	.27 - .63	18	.1 - .004

Complex Model With Augmentation

Accuracy	Validation Loss	Range of Class F1-scores	Epochs	Learning Rate (SGD)
.53	3.4	.33 - .87	12	.05 - .01

Transfer Learning

- Transfer learning was originally done by training only new output.
- Results were poor, even after unfreezing and fine tuning portions or all of pre-trained base model.
- Much better results from unfreezing all layers and fine tuning entire model from beginning.
- Two output blocks were tested.
  - Simple: global average pooling and softmax prediction
  - Complex: global average pooling, fully connected layers, dropout, batch normalization, softmax prediction
    - Complex output performed better on scratch-built CNN models, so tried it on pre-trained models as well.

Models

ResNet50

	Accuracy	Validation Loss	Range of Class F1-scores	Epochs	Learning Rate (SGD)
Complex	.87	.57	.57 - .96	21	.001 - .00025
Simple	.84	.44	.58 - .99	30	.001 - .00025

MobileNet V2

	Accuracy	Validation Loss	Range of Class F1-scores	Epochs	Learning Rate (SGD)
Complex	.82	.59	.43 - .95	27	.001 - .00025
Simple	.78	.58	.43 - .96	28	.001 - .00025

Inception V3

	Accuracy	Validation Loss	Range of Class F1-scores	Epochs	Learning Rate (SGD)
Complex	.82	.72	.50 - .95	21	.001 - .00025
Simple	.83	.57	.36 - .94	29	.001 - .0005

VGG19

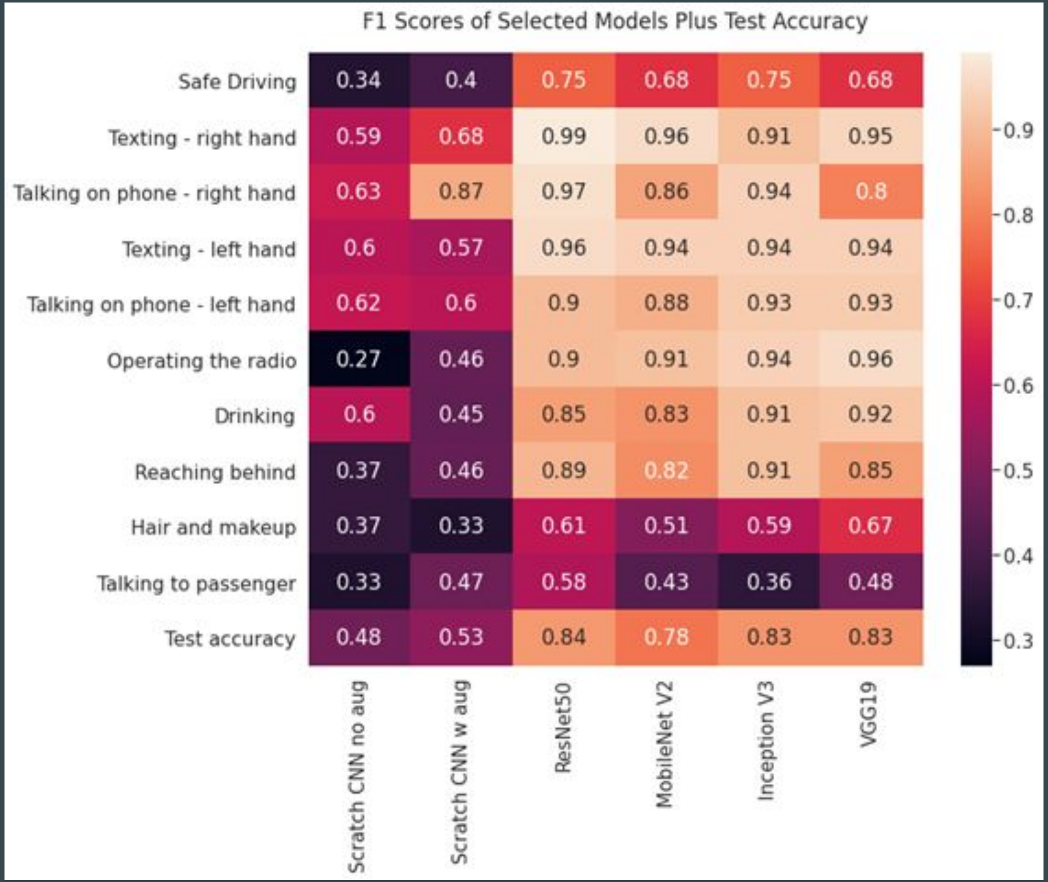
	Accuracy	Validation Loss	Range of Class F1-scores	Epochs	Learning Rate (SGD)
Complex	.80	.58	.52 - .96	20	.001 - .0005
Simple	.83	.83	.48 - .96	20	.001 - .0005

Findings

All models had trouble with classes c8: Hair and makeup, c9: Talking to passenger, and c0: Safe Driving.

Pre-trained models did much better than scratch-built models.

ResNet50 performed best overall.

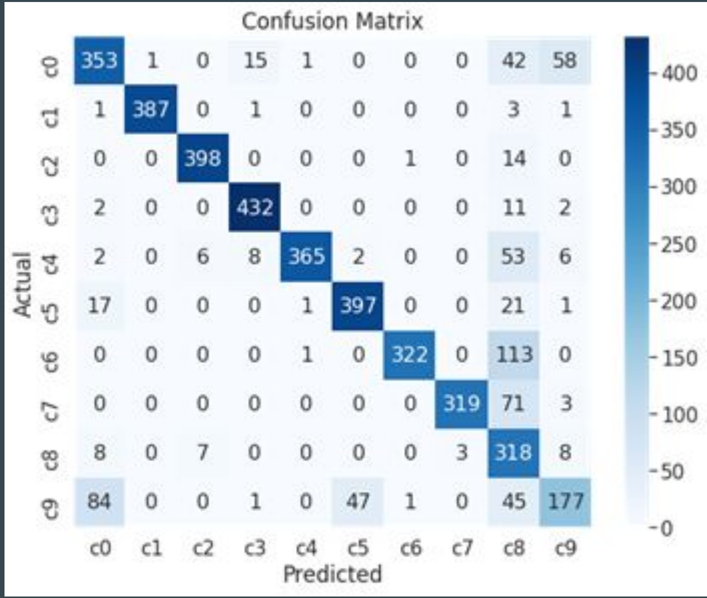


ResNet50 Model

ResNet50 had trouble with precision with class c8, with many false positives.

It had trouble mostly with recall with class c9, with many false negatives.

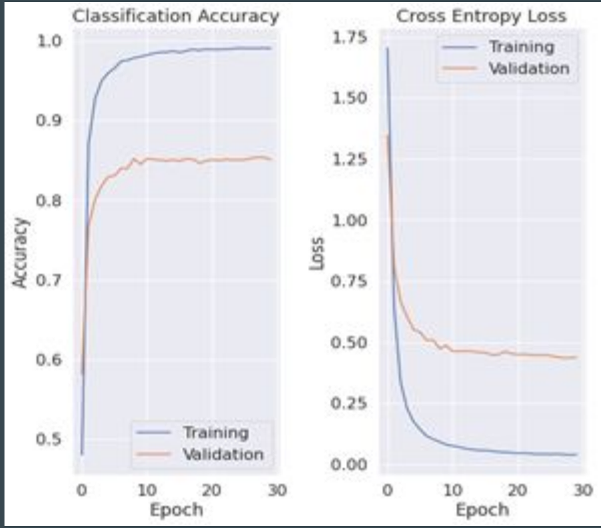
It had trouble with both with class c0 to a lesser extent.



ResNet50 Model

ResNet50 model could have improved with more training.

- Validation loss still slowly decreasing after 30 epochs
- Validation accuracy fairly flat for the last 10 epochs



Further Work

- Segmentation of head and hands
- Refinement of learning rate decay
- Adjustments to augmentation parameters
- Longer training for ResNet 50 model
- Ensemble models