Distracted Driver Behavior Classification

Elmer Atienza, Yining Feng, Narmada Gomatam

The AI Consulting Group

What is distracted driving?

Distracted driving is any activity that could divert attention from the primary task of driving.

Problem Statement

Develop a computer vision model that distinguishes distracted driving behaviors from safe driving posture

Business Applications:

- Facilitate faster claims processing and assessing the liability of claims
- Minimize insurance disputes with more accurate damage inspection
- Usage-Based Auto Insurance (UBI)
- Inference-based Driving Risk Prediction

Image Classification

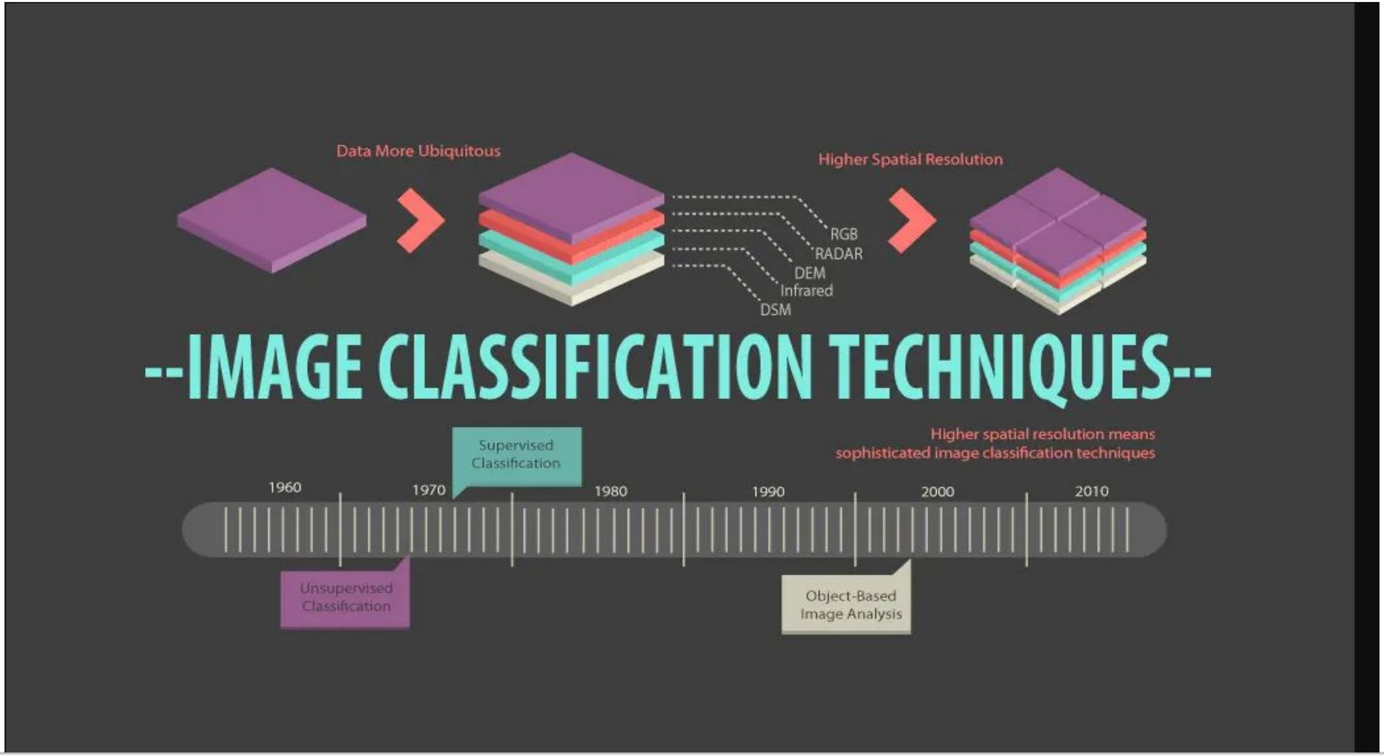


Image classification models

Modeling techniques:

- Convolutional Neural Network (CNN)
- Transfer Learning Models
    - VGG-16
    - ResNet
    - Xception
    - MobileNet
    - K-NN ensemble
- AutoML

Baseline CNN & Transfer Learning Models

Baseline CNN A 2-layer CNN neural network with 100 neurons in the hidden layer ● Hidden Layer: Rectified linear unit (ReLU) ● Output Layer: Softmax	VGG-16 A 16-layer CNN neural network with 3 × 3 filters used for all convolutional layers ● Multilayer perceptron (MLP) classifier + 3 fully connected (FC) layers	ResNet50 A 34-layer CNN neural network with mostly 3 × 3 filters used for all convolutional layers ● A shortcut connection is added to each building block
Xception A 36-layer CNN neural network with 60,000 parameters ● 4 residual depth-wise separable convolution layers with residual connections	MobileNet A 28-layer CNN neural network with 3 × 3 depthwise convolutional layers ● Depth wise separable convolutions	K-NN Ensemble Model Combining (averaging) the results of baseline CNN, VGG-16, ResNet, Xception, and MobileNet models

Baseline CNN & Transfer Learning Models

Baseline CNN ● Overfitting: training 90% validation 65% ● Misclassifications between the class ‘talking to passengers’ and other classes	VGG-16 ● Misclassifications between the class ‘talking on the phone – left’ and ‘hair and makeup’	ResNet50 ● Misclassifications between the class ‘texting right’ and ‘hair and makeup’
Xception ● Misclassifications between the class ‘texting left’ and ‘safe driving’ & between ‘talking on the phone – left’ and ‘texting left’ or ‘hair and makeup’	MobileNet ● Misclassifications between the class ‘texting left’ and ‘safe driving’ & between ‘talking on the phone – left’ and ‘hair and makeup’ ● Request the least amount of training time	K-NN Ensemble Model ● Mitigate misclassification problems ● Outperform all the other stand-alone models ● K-NN helps smoothen predicted probabilities for each class

Model Type	Extra Dense Layers	Filters/ Units	Batch Normalization	Dropout (0.5 rate)	Classification Accuracy			Output Loss			Training Time (second /epoch)
					Train	Valid	Test	Train	Valid	Test	
Baseline CNN	No	32	Yes	No	0.91	0.63	0.62	0.34	1.32	1.35	224
VGG-16	No	64	Yes	Yes	0.78	0.85	0.82	0.62	0.50	0.56	181
VGG-16	Yes	64	Yes	Yes	0.86	0.82	0.79	0.42	0.57	0.69	178
ResNet	No	64	Yes	Yes	0.88	0.85	0.84	0.34	0.55	0.67	201
ResNet	Yes	64	Yes	Yes	0.89	0.86	0.85	0.32	0.47	0.56	214
Xception	No	64	Yes	Yes	0.80	0.83	0.82	0.61	0.55	0.58	241
Xception	Yes	64	Yes	Yes	0.86	0.84	0.82	0.39	0.52	0.57	236
MobileNet	No	64	Yes	Yes	0.91	0.86	0.85	0.28	0.39	0.40	167
MobileNet	Yes	64	Yes	Yes	0.89	0.83	0.80	0.32	0.63	0.65	171
K-NN Ensemble	No	64	Yes	Yes	0.94	0.92	0.89	0.18	0.24	0.30	797

Performance Comparison

- K-NN Ensemble Model
  - The highest classification accuracy (92% validation accuracy) and the lowest output loss (0.24)
- MobileNet
  - The best stand-alone transfer learning model
  - Requires the least amount of training time (167 seconds per epoch)
  - 86% validation accuracy

Table 1: *Comparison of baseline CNN's & transfer learning models' performance and running speed.*

Performance Comparison

	VGG-16	MobileNet	Xception	ResNet	K-NN Ensemble
Safe Driving	54.4%	92.5%	78.9%	78.2%	95.3%
Texting-Right	96.2%	92.3%	93.2%	89.1%	96.6%
Talking-Right	95.1%	94.3%	99.4%	97.4%	96.5%
Texting-Left	98.5%	98.2%	98.2%	98.5%	96.2%
Talking-Left	87.9%	75.8%	76.5%	77.3%	94.8%
Operating the Radio	96.8%	99.0%	94.8%	97.5%	95.1%
Drinking	90.3%	86.2%	80.4%	79.3%	98.0%
Reaching Behind	81.1%	99.4%	99.1%	99.7%	95.4%
Hair and Makeup	63.6%	53.3%	40.1%	59.6%	92.8%
Talking to passenger	83.1%	68.5%	58.8%	79.5%	97.7%

- K-NN Ensemble Model
  - > 90% classification accuracy for each of the 10 classes
- MobileNet
  - The best stand-alone transfer learning model
  - Highest validation accuracy across most classes
  - 99.4% validation accuracy for the ‘reaching behind’ class

Table 2: *Comparison of transfer learning models’ prediction accuracy per class.*

AutoML Model Performance Summary

DISTRACTION CLASS	Google Cloud Platform (GCP) AutoML Vision						Apple CreateML					
	TRAIN		VALIDATION		TEST		TRAIN		VALIDATION		TEST	
	PRECISION	RECALL	PRECISION	RECALL	PRECISION	RECALL	PRECISION	RECALL	PRECISION	RECALL	PRECISION	RECALL
	WITH IMAGE EXPOSURE AND BLUR DATA AUGMENTATION (200 ITERATIONS)											
	TRAINING ACCURACY: 99.4% **VALIDATION ACCURACY: 95%**						**TRAINING ACCURACY: 99.6%** **VALIDATION ACCURACY: 95.3%**					
	TRAINING TIME: 8 hours						**TRAINING TIME: 8 hours**					
Safe Driving	0.99	0.98	0.92	0.9	0.7	0.61	0.99	0.99	0.92	0.89	0.71	0.51
Texting - right	1	1	0.98	0.95	0.89	0.55	1	1	0.98	0.92	0.89	0.53
Talking on the phone - right	0.99	1	0.98	0.98	0.84	0.78	1	1	0.98	0.98	0.83	0.76
Texting - left	1	1	0.97	94	0.73	0.73	1	1	0.98	0.92	0.71	0.71
Talking on the phone - left	1	1	0.98	0.98	0.81	0.72	1	1	0.97	0.97	0.8	0.7
Operating the radio	0.99	0.98	0.97	0.96	0.77	0.7	1	1	0.95	0.98	0.75	0.69
Drinking	1	1	0.97	0.97	0.63	0.58	1	1	0.96	0.96	0.62	0.57
Reaching behind	1	1	0.98	0.97	0.88	0.84	1	1	0.96	0.98	0.87	0.85
Hair and makeup	0.99	0.99	0.95	0.94	0.48	0.81	0.99	0.99	0.92	0.96	0.44	0.79
Talking to passenger	0.99	0.99	0.9	0.98	0.46	0.68	0.99	0.99	0.9	0.97	0.45	0.67

AutoML Model Performance Summary

True Label \ Predicted Label	talking left	texting left	drinking	reaching behind	texting right	operate radio	talking right	talking passenger	hair makeup	safe driving
talking left	100%	-	-	-	-	-	-	-	-	-
texting left	-	100%	-	-	-	-	-	-	-	-
drinking	-	-	100%	-	-	-	-	-	-	-
reaching behind	-	-	-	100%	-	-	-	-	-	-
texting right	-	-	-	-	100%	-	-	-	-	-
operate radio	-	-	-	-	-	99%	-	-	0%	1%
talking right	-	-	0%	-	-	-	100%	-	-	-
talking passenger	-	-	-	-	-	-	-	99%	1%	-
hair makeup	-	-	0%	-	-	-	0%	-	99%	-
safe driving	-	-	-	-	-	-	-	-	0%	100%

True positives

Your model correctly predicted **texting left** on these images

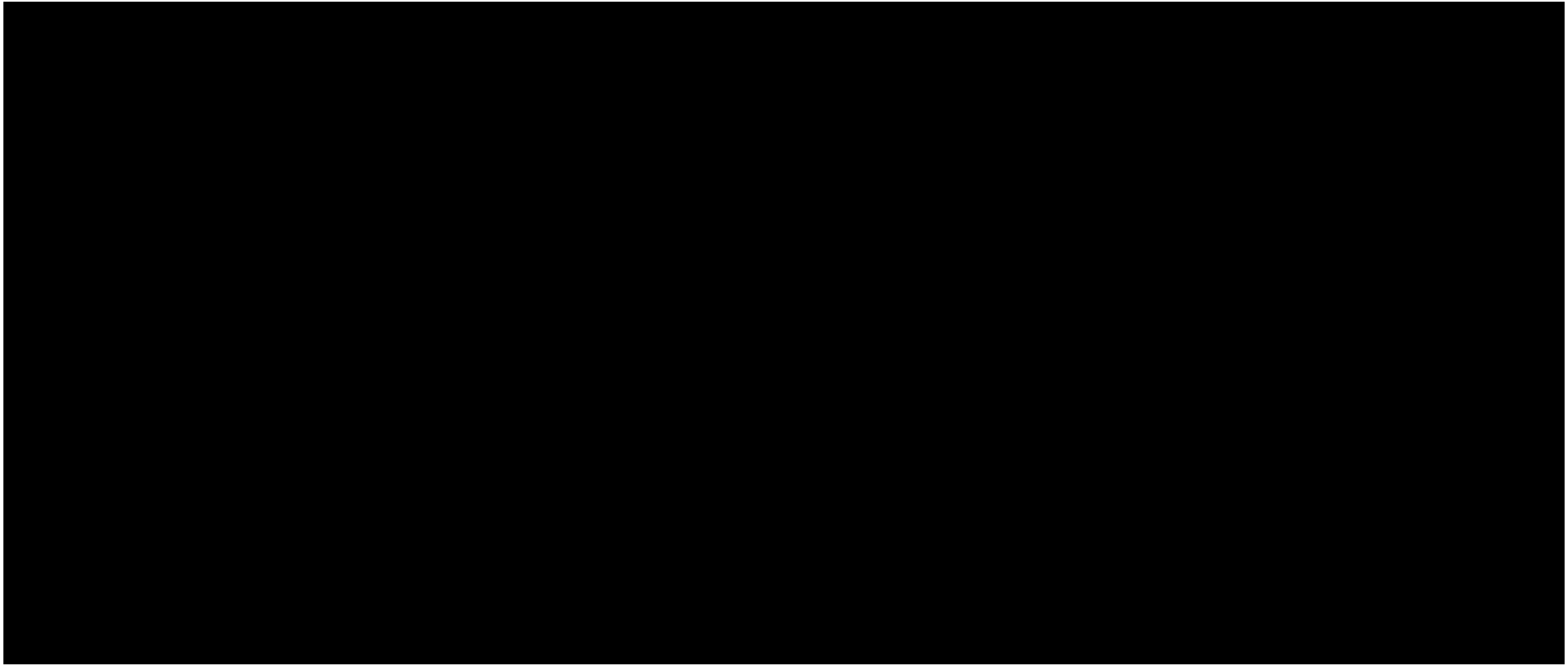
Score: 0.67843133 Score: 0.73333335 Score: 0.7372549 Score: 0.77254903 Score: 0.7764706

False negatives

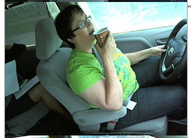
Your model should have predicted **hair makeup** on these images

Score(s): 0.29803923 Score(s): 0.33333334 Score(s): 0.3529412 Score(s): 0.44313726

DDBC Solution Prototypes



Inferences on New/Unseen Data

CLASS	**Google Cloud Platform (GCP) AutoML Vision**	**Apple CreateML**	CLASS	**Google Cloud Platform (GCP) AutoML Vision**	**Apple CreateML**
	Texting - left (0.63)	Texting - left (0.88)		Operating the radio (0.74)	Operating the radio (0.85)
	Texting - right (0.72)	Texting - right (0.99)		Drinking (0.30)	Drinking (0.92)
	Talking on the phone - right (0.81)	Talking on the phone - right (0.99)		Reaching behind (0.73)	Talking to passenger (0.70)
	Texting - left (0.85)	Safe driving (0.79)		Hair and makeup (0.77)	Hair and makeup (0.98)
	Talking on the phone - left (0.83)	Talking on the phone - left (0.99)		Operating the radio (0.31)	Operating the radio (0.51)

* **Green** means **correct** classification, **red** means **incorrect** classification

Predictive Risks of DDBC

CLASS	PREDICTION	RISK
	Safe driving (0.79)	• Prediction as driving safely even if the driver is not • False sense of security • Safe driver record despite behavior violations
	Texting - left (0.88)	• Prediction of distracted driver even if the driver is in safe driving position • Annoying and irritating incorrect distracted behavior alerts • Unjustifiable unfavorable driving record

Recommendations for future study

Our project provided a baseline performance and accuracy to benchmark against for future research.

Future recommendations include:

- Data augmentation using color shifting and rotation of images for improved accuracy and to prevent overfitting
- A better methodology to improve detection of face, hands, and skin
- Training using a more powerful computing platform
- Interfacing with Amazon assistant Alexa and all the Internet of Things (IoT) devices that Alexa interfaces with such as the Amazon Echo's suite of personal assistant products.

Questions?