Vehicle's Driver Distraction Detection using CNN

Snehil Vishwakarma, Shivani Gupta

Computer Vision - CSCI B657

Introduction

- In this project we are developing a system to detect driver's motion in an image taken from inside the car using camera on the dashboard.
- This project uses Convolutional Neural Networks to train the model to classify the images into 10 different categories :
  c0 - safe driving
  c1 - texting (right)
  c2 - talking on the phone (right)
  c3 - texting (left)
  c4 - talking on the phone (left)
  c5 - operating the radio
  c6 - drinking
  c7 - reaching behind
  c8 - hair and makeup
  c9 - talking to passenger

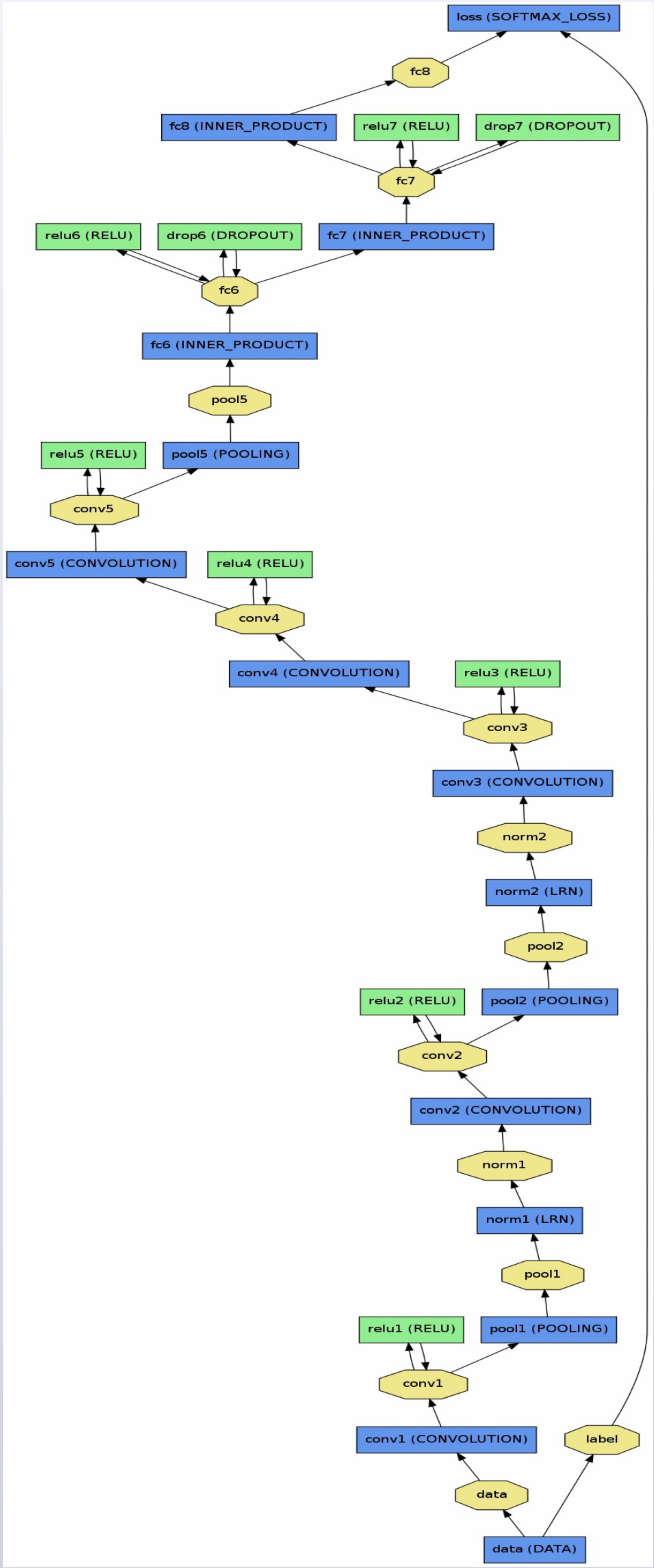
Motivation and Application

According to the CDC motor vehicle safety division, one in five car accidents is caused by a distracted driver. Sadly, this translates to 425,000 people injured and 3,000 people killed by distracted driving every year.
Last year State Farm launched a kaggle competition to improve these alarming statistics and better insure their customers, by testing whether dashboard cameras can automatically detect drivers engaging in distracted behaviors. This problem was directly motivated from the competition.

Dataset

- 2D dashboard camera images where the driver's actions are captured from passanger seat's POV.
- Training set - 15698 images
- Validation set -3363 images
- Test set - 3364 images

Model



Approach

- Software Packages used - **Caffe**
- Pre-trained model - **Alexnet**
- Trained the model after fine-tuning on our train-set and validation-set for 100,000 iterations
- Final model - 90000 iteration

Working of Different Layers:

1. *Convolution Layer: convolves the input image with a gaussian filter*
2. *Pooling Layer: max pooling*
3. *Dropout Layer: Drops with probability 0.5*
4. *ReLU Layer: computes the output as x if x > 0 and negative_slope * x if x <= 0*
5. *Normalization Layer: normalize over local inputs. Each input is divided by* $(1+(\alpha/n)\sum_i x_i^2)^\beta$, *where alpha = 0.0001 and beta = 0.75*

Result

- **Accuracy: 99.43 %**
- **Negative prediction - 19 out of 3364**

Confusion Matrix

		Predicted Label									
		c0	c1	c2	c3	c4	c5	c6	c7	c8	c9
True Label	c0	388	0	0	0	0	0	0	0	1	0
	c1	0	324	0	0	0	0	0	0	0	0
	c2	0	0	357	0	0	0	0	0	2	0
	c3	1	0	0	347	2	0	0	0	0	0
	c4	0	0	0	3	368	0	0	0	0	0
	c5	3	0	0	0	0	346	0	0	0	0
	c6	0	0	0	0	0	0	332	0	0	0
	c7	0	0	0	0	0	0	0	297	0	0
	c8	0	0	0	0	0	0	0	0	286	6
	c9	0	0	0	0	0	0	0	0	1	300

Observations

True Label:
c4 - **Talking on the phone (left)**

Predicted:
c3 - **Texting (left)**

True Label:
c8 - **hair and makeup**

Predicted:
c9 - **talk to the passenger**

True Label:
c3 - **Texting (left)**

Predicted:
c0 - **Safe driving**

True Label:
c0- **safe driving**

Predicted:
c8 - **hair and makeup**

References

- Yan, Chao; Coenen, Frans; Zhang, Bailing: 'Driving posture recognition by convolutional neural networks', IET Computer Vision, 2016, 10, (2), p. 103-114, DOI: 10.1049/iet-cvi.2015.0175
  IET Digital Library, http://digital-library.theiet.org/content/journals/10.1049/iet-cvi.2015.0175
- Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. 2014. Caffe: Convolutional Architecture for Fast Feature Embedding. In *Proceedings of the 22nd ACM international conference on Multimedia* (MM '14). ACM, New York, NY, USA, 675-678. DOI=http://dx.doi.org/10.1145/2647868.2654889
- https://www.kaggle.com/c/state-farm-distracted-driver-detection/data
- https://kb.iu.edu/d/bcqt
- caffe.berkeleyvision.org
- https://jeremykarnowski.wordpress.com/2015/07/15/alexnet-visualization/