Rich Gude — Sam Cohen — Luis Ahumada

TABLE OF CONTENTS

TESLA

WAYMO

TOYOTA

Audi

cruise

Uber

and many more...

comma.ai

Welcome to the comma.ai Programming Challenge!

Your goal is to predict the speed of a car from a video.

- data/train.mp4 is a video of driving containing 20400 frames. Video is shot at 20 fps.
- data/train.txt contains the speed of the car at each frame, one speed on each line.
- data/test.mp4 is a different driving video containing 10798 frames. Video is shot at 20 fps.

Deliverable

Your deliverable is test.txt. E-mail it to givemeajob@comma.ai, or if you think you did particularly well, e-mail it to George.

Evaluation

We will evaluate your test.txt using mean squared error. <10 is good. <5 is better. <3 is heart.

Twitter

Follow us!

What you are given...

- A video of driving containing 20,400 frames (20 fps)
- A train.txt containing the speed of the car at each frame (mph)

OBJECTIVE

Predict the speed of a car at each frame of a different video

METRICS

Mean Squared Error (MSE)

DATA PREPROCESSING

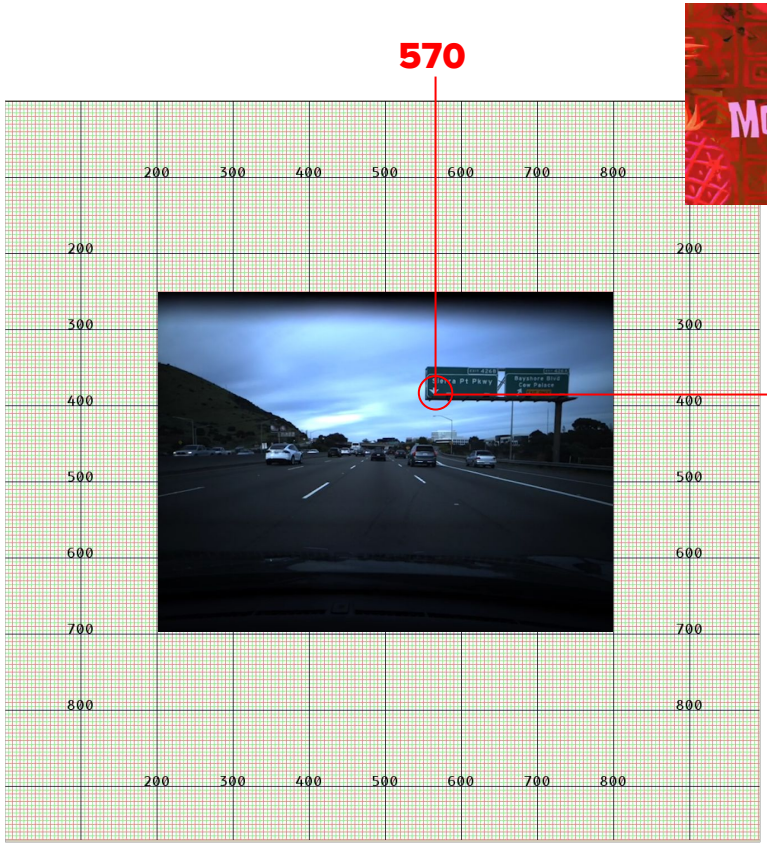
DATA LOAD

- Video frames to images using cv2.VideoCapture
- Image resize from 640x480 to 320x60
- Saved to NumPy array

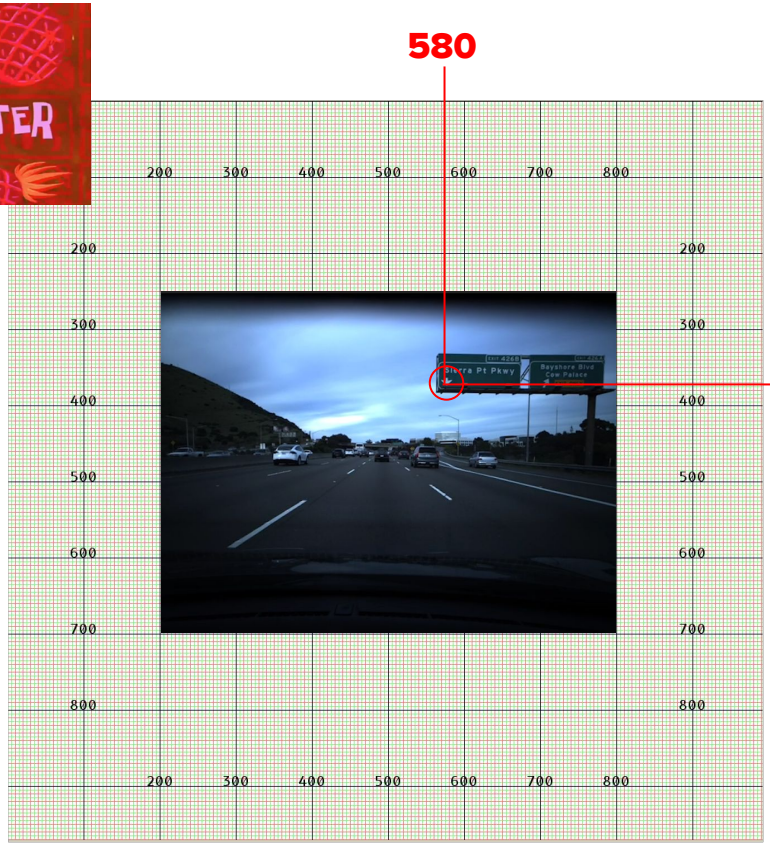
Before

After

HOW DO WE ESTIMATE SPEED?



A FEW MOMENTS LATER

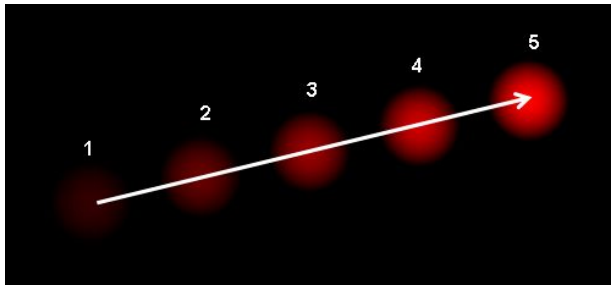


Frame 1

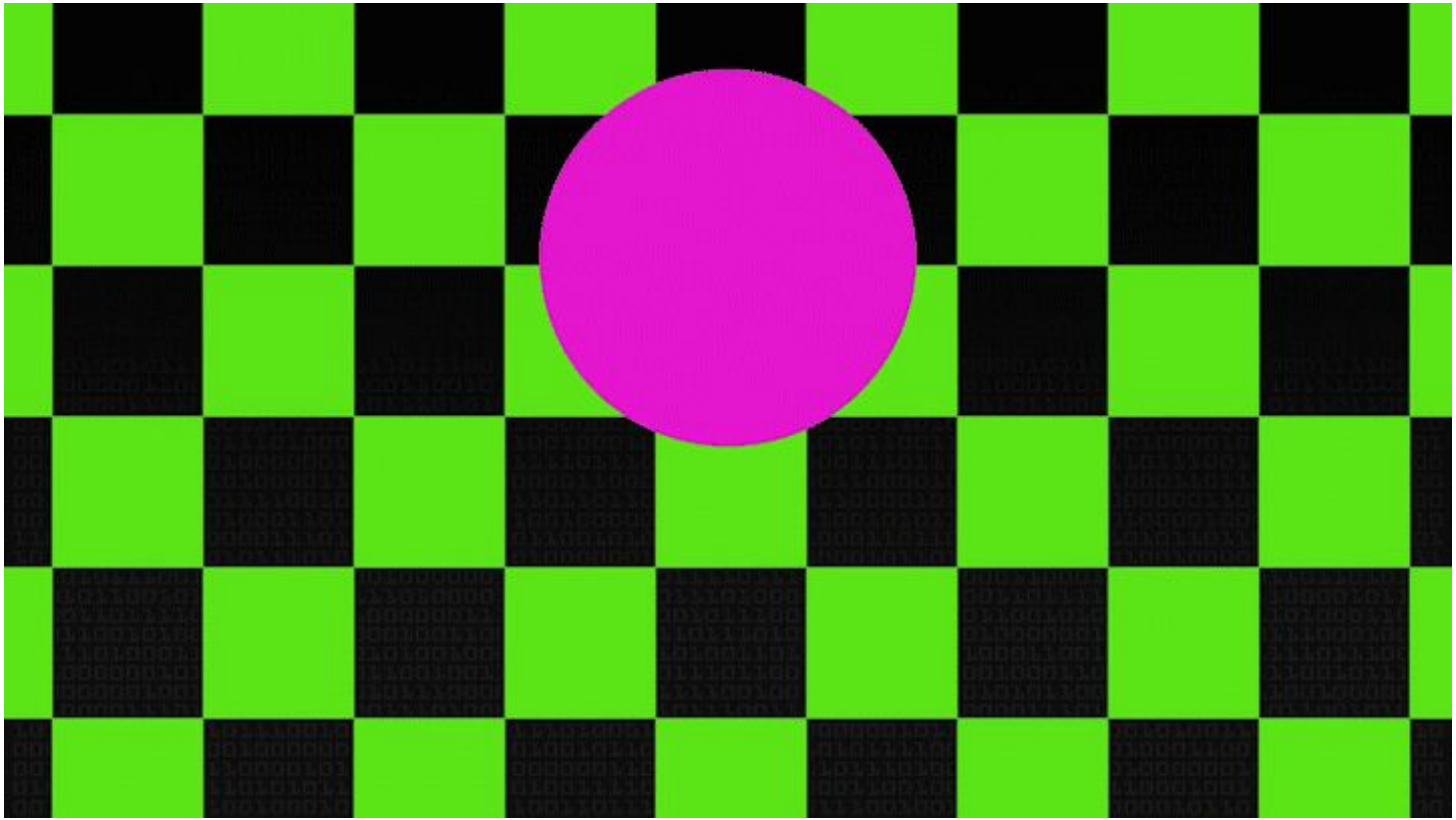
Frame 5

OPTICAL FLOW

- Pattern of movement between images at a pixel level.
- cv2. calcOpticalFlowFarneback: Class computing a dense optical flow using the Gunnar Farneback's algorithm.
- Assumes the intensity of pixels does not change.
- We used dense optical flow.

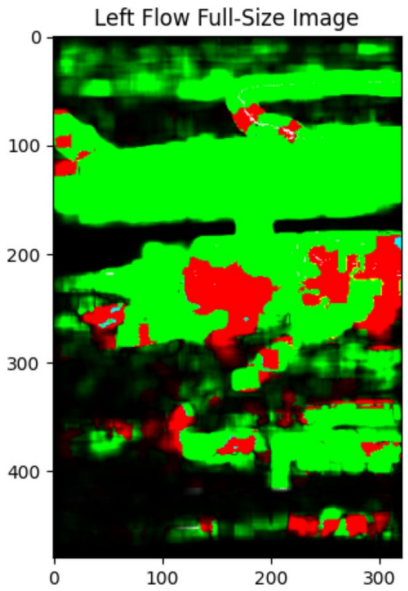


The optical flow vector of a moving object in a video sequence.

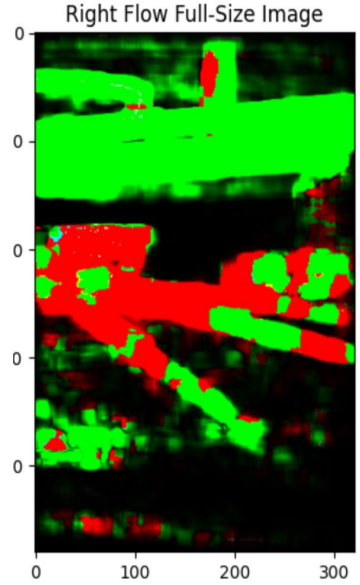


Example


Left Side Full-Size Image 1
0
100
200
300
400
0
100
200
300


Left Flow Full-Size Image
0
100
200
300
400
0
100
200
300


Right Side Full-Size Image 1
0
100
200
300
400
0
100
200
300


Right Flow Full-Size Image
0
0
0
0
0
0
100
200
300

BATCH SHUFFLING

- Randomly shuffle pairs of rows in the array inputs, separating each pair into train, test, and validation data.
- Step size is used to determine the time difference between pairs.
    - A large step size means more significant changes between the images

Example -> StepSize: 4

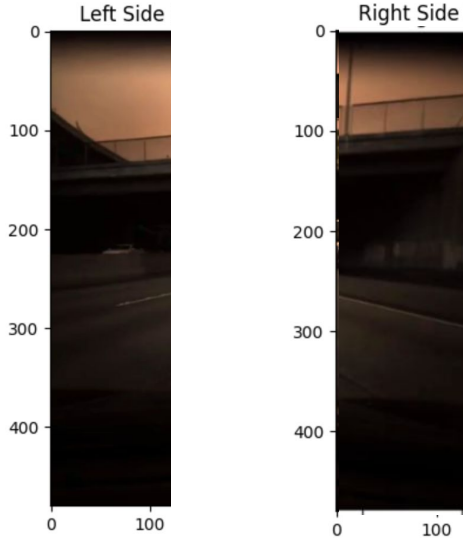
Frame 1 Frame 2 Frame 3 Frame 4 Frame 5

IMAGE SLICING

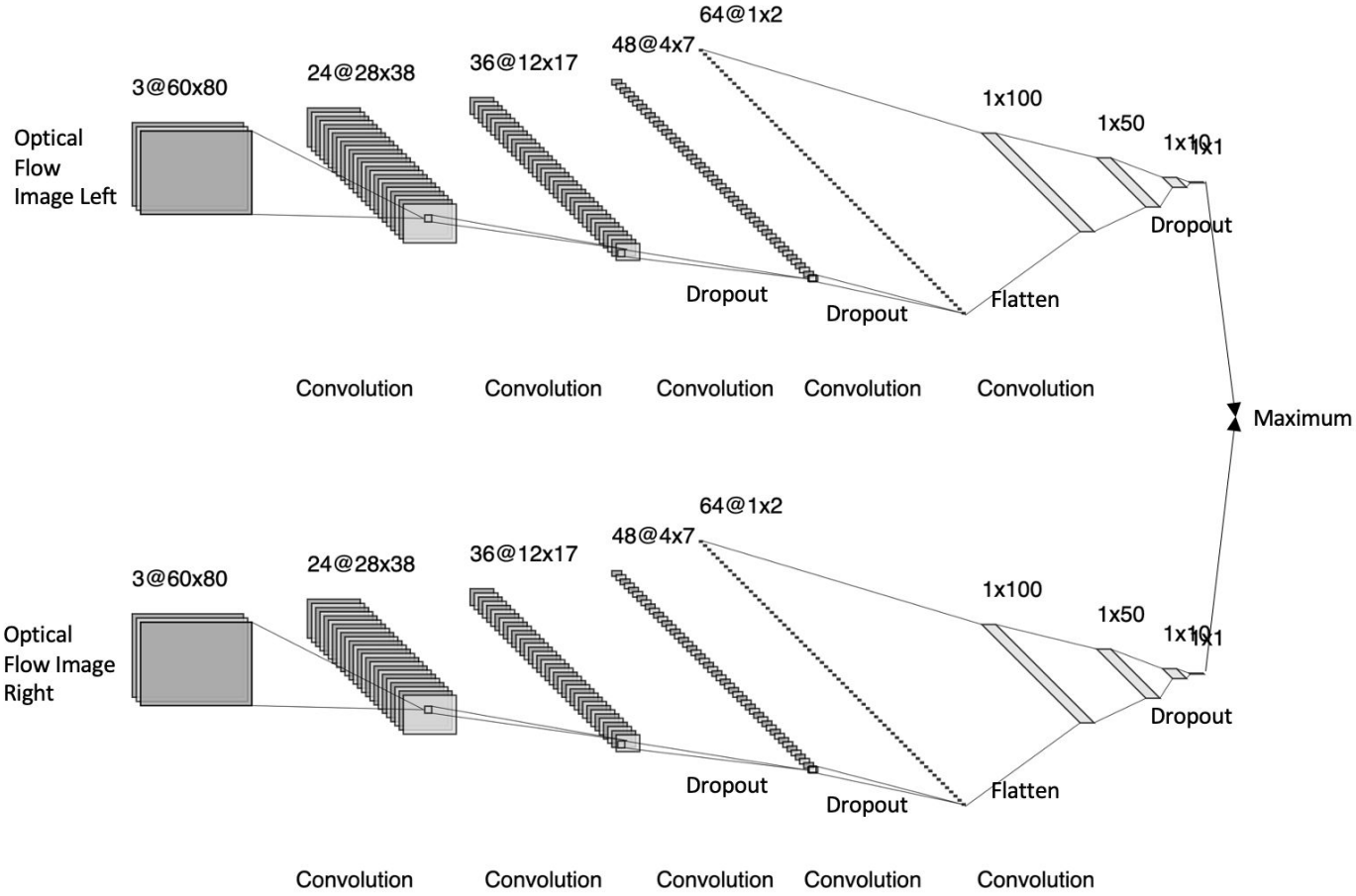
- Focus of the model should be on the outer edges of the image.
- Cut out the middle half of the image.
- Apply changes to the brightness.

Before

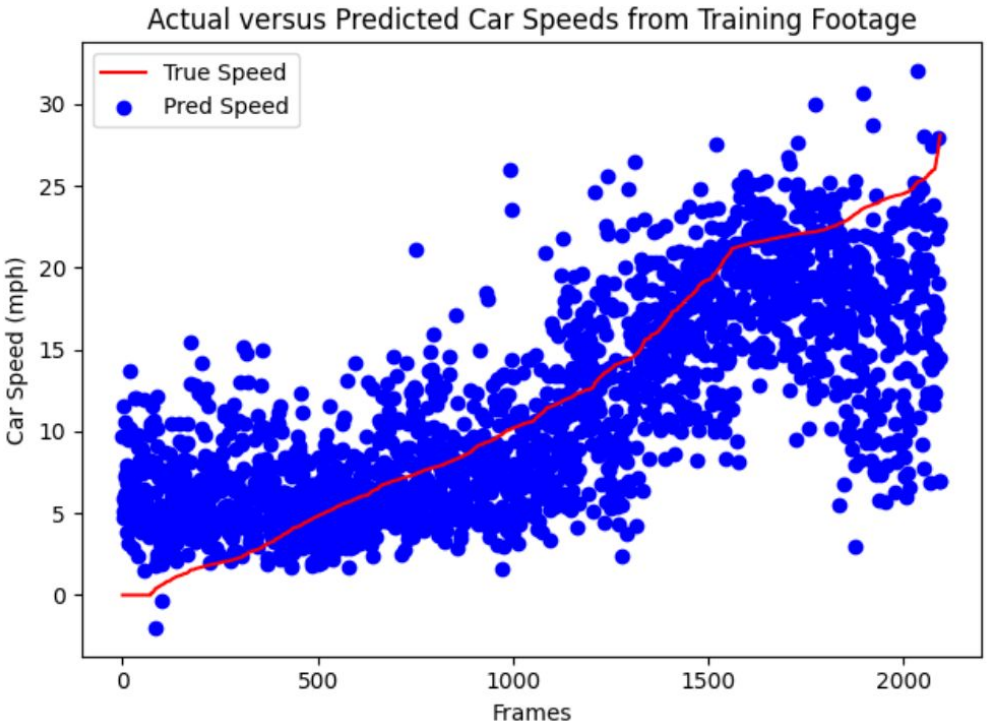
After



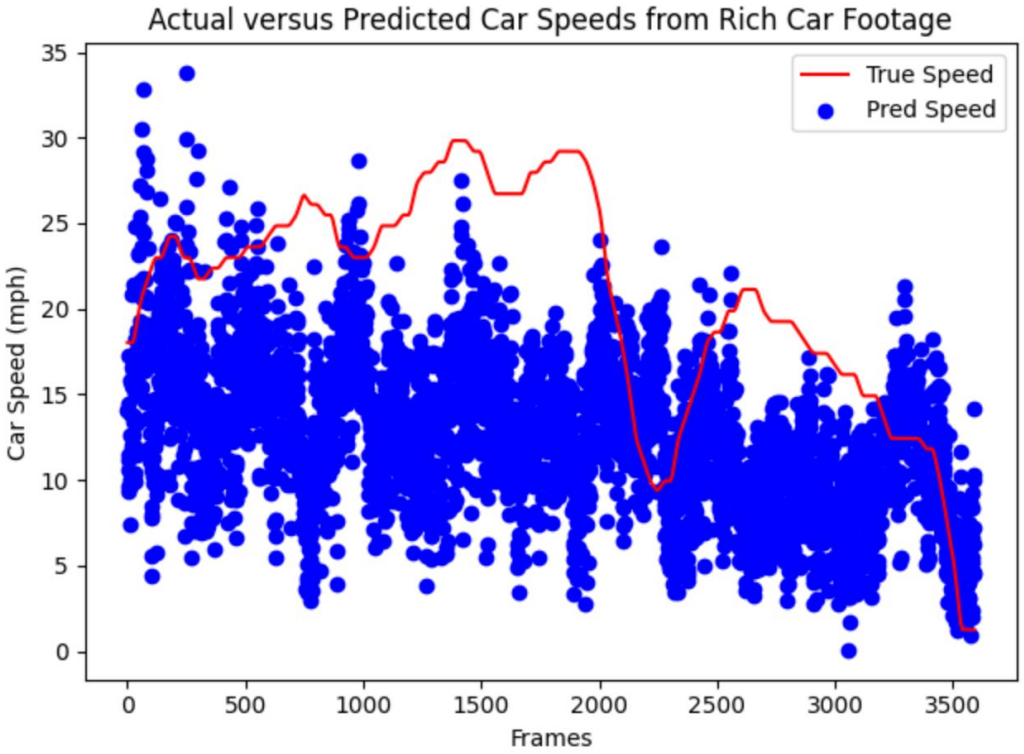
THE MODEL


Optical
Flow
Image Left
3@60x80
24@28x38
36@12x17
48@4x7
64@1x2
1x100
1x50
1x10
1x1
Dropout
Dropout
Dropout
Flatten
Convolution
Convolution
Convolution
Convolution
Convolution
Maximum
Optical
Flow Image
Right
3@60x80
24@28x38
36@12x17
48@4x7
64@1x2
1x100
1x50
1x10
1x1
Dropout
Dropout
Dropout
Flatten
Convolution
Convolution
Convolution
Convolution
Convolution

RESULTS


Actual versus Predicted Car Speeds from Training Footage
True Speed
Pred Speed
Car Speed (mph)
Frames
0
5
10
15
20
25
30
0
500
1000
1500
2000

MSE: 27.25


Actual versus Predicted Car Speeds from Rich Car Footage
True Speed
Pred Speed
Car Speed (mph)
Frames
0
5
10
15
20
25
30
35
0
500
1000
1500
2000
2500
3000
3500

MSE: 110.99

CONCLUSION

- Model performance is inferior to other participants.

Future work

- Experiment with current parameters (eg. image size, crop size, step size, etc.)
- Experiment with different network architectures.
- Improve optical flow methods.
    - FlowNet 2.0 (Nvidia)