End to End Learning for Visual Navigation

Ute Schiehlen[1], Natalie Reppekus[1], and Raymond Chua[1]

[1]Technical University of Munich

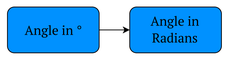
Abstract

An important component in autonomous vehicles is visual navigation. We used a sequence of frames recorded at the front of the car to predict the steering angle using an end to end approach. Our network is based on Bojarski et al.[1] which consists of five convolutional and three fully connected layers using the human steering angle as the training signal. As this model did not fully exploit the temporal information of image sequences, we extended this architecture by training an additional convolutional network with *optical flow* computed from the original images as input and optimizing over the combined loss. As this approach yielded good results, we trained a third network consisting of a *Long-Short term memory* (LSTM) cell to further exploit temporal information. We achieved an improvement of 0.12 in the Mean Squared Error in our approach compared to Nvidia?s model.

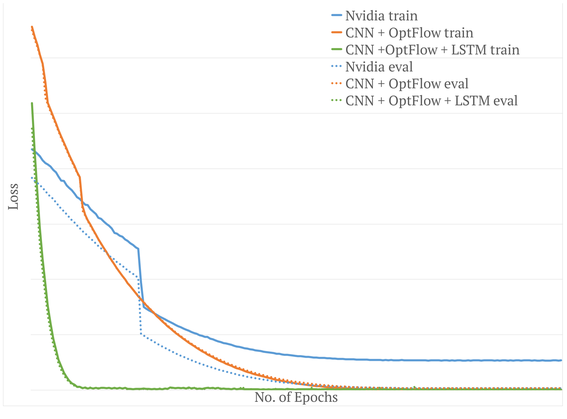
Data Preprocessing

Images

Labels



Training and Evaluation Loss



Test Results

Model	NVIDIA	CNN+OptF	CNN+OptF+LSTM
MSE	0.1911	0.1910	**0.072**

The values above are in radians. Compared in Degrees: 0.19 rad ≈ 11°C, 0.072 rad ≈ 4°C.

Loss Functions

$$loss_{opt} = \sum (y_{orig} - labels)^2 + (y_{opt} - labels)^2$$
$$loss_{lstm} = \sum (y - labels)^2$$

Our Approach

We converted RGB images into YUV color space, which encodes the human perception.We trained all networks using mini-batch Stochastic Gradient Descent with a learning rate of 0.001 over 200 epochs. For the Nvidia and CNN+Optical Flow network we consider a balanced subset of the training data in order to reduce the number of samples where the car is driving straight. However this was not applied to the LSTM model due to the requirement of having sequential images. In our experiments we have shown that the LSTM model converges faster during training and achieves a better performance for testing.

References

[1] M. Bojarski, D. D. Testa, D. Dworakowski, B. Firner, B. Flepp, P. Goyal, L. D. Jackel, M. Monfort, U. Muller, J. Zhang, X. Zhang, J. Zhao, and K. Zieba. End to end learning for self-driving cars. *CoRR*, abs/1604.07316, 2016.

[2] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural Computation*, 9(8):1735–1780, 1997.

[3] K. Simonyan and A. Zisserman. Two-stream convolutional networks for action recognition in videos. In *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada*, pages 568–576, 2014.

Architectures