END TO END LEARNING FOR VISUAL NAVIGATION

Ute Schiehlen, Natalie Reppekus, Raymond Chua

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

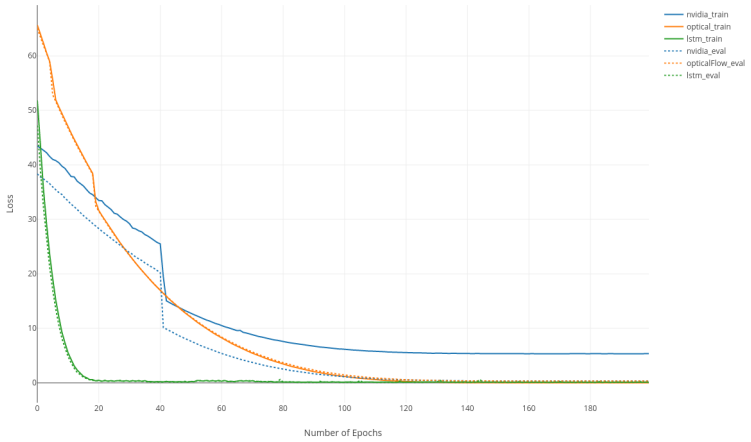
Images

RGB Image → Randomized Horizontal Flip → Normalization → Convert to YUV → Zero-Mean

Labels

Angle in ° → Angle in Radians

Training



Test Results

Nvidia	CNN + Optical Flow	Optical Flow + LSTM
0.19311	0.1910	0.072

This values are in radians. Compared in Degrees: 0.19 rad ~ 11°, 0.072 ~ 4°

Loss Functions

CNN + Optical Flow Model

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

CNN + Optical Flow + LSTM

$$loss = \sum(y - labels)^2$$

References

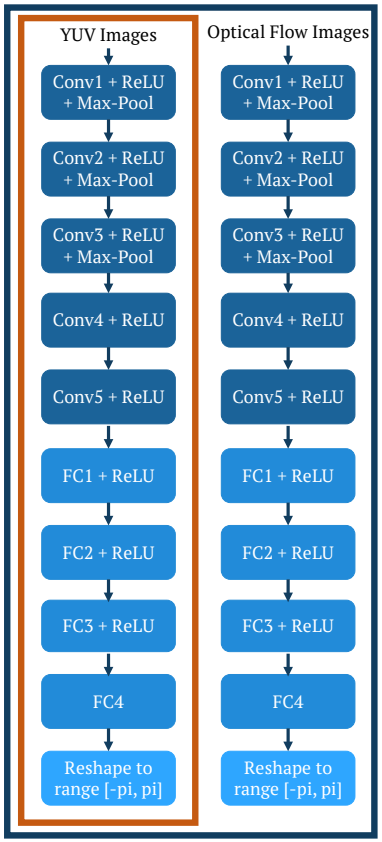
[1] M. Bojarski, D. D. Testa, D. Dworakowski, B. Firner, B. Flepp, P. Goyal, L. D. Jackel, M. Monfort, U. Muller, J. Zhang, X. Zhang, J. Zhao, and K. Zieba. End to end learn- ing for self-driving cars. *CoRR*, abs/1604.07316, 2016. 1, 2

[2] S.HochreiterandJ.Schmidhuber.Longshort-termmemory. *Neural Computation*, 9(8):1735–1780, 1997. 1

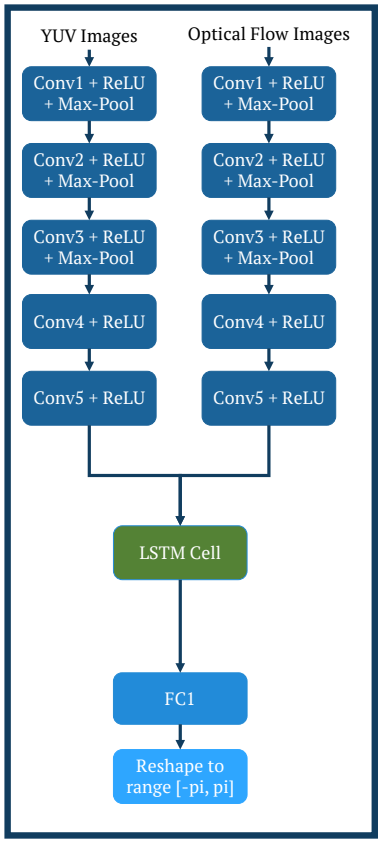
[3] K. Simonyan and A. Zisserman. Two-stream convolutional networks for action recognition in videos. In *Advances in Neural Information Processing Systems 27: Annual Confer- ence on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada*, pages 568– 576, 2014. 1

Architectures

CNN + Optical Flow Model



CNN + Optical Flow + LSTM Model



Nvidia Model