Predicting a vehicles velocity using dashcam footage and dense optical flow

Florian Wolf
Department of Mathematics and Statistics, University Konstanz
Konstanz, Germany
florian.2.wolf@uni-konstanz.de

Franz Herbst
Department of Physics, University Konstanz
Konstanz, Germany
franz.herbst@uni-konstanz.de

***Abstract*—In this report bla bla bla**

***Index Terms*—deep learning, computer vision, velocity prediction, dense optical flow**

I. Introduction

Here are some motivational words need

A. Aim of the project

Here we need to clarify the aim of the project

II. Data collection, analysis and preprocessing

For our data set, we used the comma ai speedchallenge[1] data base. This data set provides two dashcam videos: a training video, (20400 frames, shoot at 20 frames per second) including ground truths and a testing video (10798 frames, shoot at 20 frames per second) without labels, which they use for applications to check how well a submitted model is able to generalize. As we only have access to the labels of the test video frames, we decided to split the provided train data by the 80/20 principle into training and testing subsets. Here we did not shuffle the data randomly, as we needed to always have two consecutive frames to calculate the optical flow. We initially used a hard cut off after 80% of the frames, as we wanted to test our model on unseen data, to measure how good our model is able to generalize. We will analyse will analyse this naive approach later, when we take a closer look at the results.

A. Data analysis

To analyse the velocity distribution in the two subsets, we plotted the velocity per time distribution in REFERENCE MISSING

Write down, what we expect from this distribution and mention how important brightness is (to possibly explain bad result). Keep the point, that in the end the car is in the city, will in the first part on the highway (reason why our model performed so poorly).

[1]https://github.com/commaai/speedchallenge

B. Preprocessing

Each of the provided frames has a size of $(640, 480, 3)$ pixels. Due to computational limitations, we decided to cut off the last 60 pixels from the lower border, to remove a black frame inside the car, which did not have any effect on the optical flow. Furthermore, we cut the frame size in half and calculated the optical flow using the Farneback pyramid REFERENCE TO THE PAPER MISSING method with the following parameters (WHY DID WE CHOOSE THEM LIKE THAT)

$$\begin{aligned} \text{pyramid levels} &:= 3 \\ \text{pyramid scaling} &:= 0.5 \\ \text{window size} &:= 6 \\ \text{pixel neighborhood size} &:= 5 \\ \text{SD of the gaussian filter} &:= 1.1 \end{aligned}$$

To decrease the training duration, we halved the size of the optical flow image again, resulting in a resolution of $(160, 105, 3)$ pixels per frame. As we used a window size of 6, a comparison between the original optical flow and the down sampled one lead to the result, that we do not loose a lot information. The preprocessing pipeline is shown in figure REFERENCE MISSING. As we later on wanted to see if the model performs better using the dashcam frames as additional material, we did the same down sampling with the frames.

Explain, how we display the optical flow.

III. Method selection and architecture

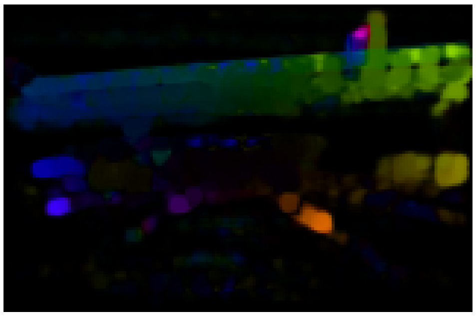
The prediction of the vehicles speed is a non-linear regression task, the choice of a neural network is reasonable. Recent architectures we discussed in the lecture (RESNET, GOOGLENET, etc.) have shown than using multiple stacked convolution layers combined with stacked dense layers, perform well on image classification tasks. Therefore the choice of a convolutional neural network is justified.

As a base architecture we decided to give the model of https://arxiv.org/pdf/1604.07316v1.pdf a try. As the paper used it for self-driving cars, the model has enough complexity to handle a task like ours and with the amount of layers, we had a lot of possibilities, to fine-tune and improve the model.

In a first approach, we tried the raw model, using different activation functions and we embedded residual layer and batch

(a) Original frame

(b) Frame after down sampling



(c) Optical flow field, already down sampled

Fig. 1: Preprocessing of the video frames.

normalization as improvements we discussed in the lecture, to speed up training and improve the performance. We tried the following activation functions

$$\mathrm{ReLu} : \mathbb{R} \to \mathbb{R}, x \mapsto \max\{0, x\}$$

relatively low number of epochs, as the model always seemed to overfit, because we have a relatively complex model for not a lot of data frames

IV. Results and comparison

V. Further work

Acknowledgment

The preferred spelling of the word "acknowledgment" in America is without an "e" after the "g". Avoid the stilted expression "one of us (R. B. G.) thanks . . .". Instead, try "R. B. G. thanks. . .". Put sponsor acknowledgments in the unnumbered footnote on the first page.

References

[1] G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955.

[2] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[3] I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[4] K. Elissa, "Title of paper if known," unpublished.

[5] R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.

[6] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[7] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.