Project proposal
Advanced Deep Learning for Computer Vision

Alexander Ziller, Leonhard Feiner

1 Introduction

Part of this course is to practically use the aquired knowledge of lectures. This is done via implementing and at best improving state of the art approaches in Deep Learning for Computer Vision. Our focus hereby is to specifically apply approaches for localizing a camera in the world using deep neural networks.

A crucial problem for mobile robotics is to estimate their position and orientation in a real world scenario. It has been shown Deep Neural Nets outperform classical approaches. We want to implement and combine two existing methods and evaluate their shared performance.

2 Related work

Estimation of camera position in a real world scenario is a long known problem. Therefore there exist classical approaches that have been used in the last decades.

Simultaneous localization and mapping (SLAM) is a traditional solution to the problem of appearance based relocalization. Variations are metric and appearance based SLAM.

PoseNet [2] is a Deep Learning based approach on camera localization in the world. It is training a convolutional neural network end-to-end to regress six degrees of freedom of camera position and orientation. It uses transfer learning from classification tasks and therefore could overcome past problems. For new scenes it is very sample efficient.

VLocNet++ [3] is also tackling this approach of camera localization and showed that including semantic segmentation is a possible and effective way to improve performance.

MapNet [1] is a very recent method with good performance in camera localization. It includes available sensory inputs such as GPS and visual odometry.

3 Current problems

PoseNet and MapNet use only image and sensory input but do not exploit semantic information encoded into these images. VLocNet++ showed that this is an effective way to improve performance of absolute camera localization. However, its architecture is quite complex.

4 Goals

Our final goal is to improve accuracy of absolute camera localization by using semantic information. To do this we aim to include semantic information learning into a pose regression network.

So first we set up an running implementation of a Pose Regression Network (such as MapNet) (2 weeks) and make sure we can reproduce results from the paper (1 week). Secondly we approach to include semantic information by adapting network architecture and loss function (6 weeks). If that is feasible we show changes in results especially in performance and accuracy (2 weeks). As a last step we show if semantic information changes learning of the network e.g. visualizing the focus of the network and compare it to classical MapNet (2 weeks).

Datasets are available online with existing papers (e.g. DeepLoc dataset). Evaluation will be done by standard techniques of other papers to ensure comparability.

Literatur

[1] Samarth Brahmbhatt u. a. „Geometry-Aware Learning of Maps for Camera Localization“. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2018.

[2] Alex Kendall, Matthew Grimes und Roberto Cipolla. „Posenet: A convolutional network for realtime 6-dof camera relocalization“. In: *Proceedings of the IEEE international conference on computer vision*. 2015, S. 2938–2946.

[3] Noha Radwan, Abhinav Valada und Wolfram Burgard. „VLocNet++: Deep Multitask Learning for Semantic Visual Localization and Odometry“. In: *IEEE Robotics and Automation Letters (RA-L)* 3.4 (2018), S. 4407–4414. DOI: 10.1109/LRA.2018.2869640.