Appended Form 2

Internship Implementation Plan

Name	Anindya Kundu
Research title	Semantic Segmentation and Lane Detection in Urban Traffic Scene

Research Proposal

Semantic Segmentation and Lane Detection in Urban Traffic Scene

Anindya Kundu
anindyaak007@gmail.com

Advisor: Dr. Toru Tamaki
tamaki@hiroshima-u.ac.jp
Department of Information Engineering
Hiroshima University

Abstract

The project aims to develop a deep learning model for semantic segmentation of different objects and lane detection in multiple urban traffic scenes, using state-of-the-art convolutional neural networks, providing a real-time scene understanding for autonomous driving. The model will enable a self-driving vehicle to adjust its course by following the lane and avoiding objects in its path. The model will be trained and tested on image datasets created by extracting frames from a camera feed and curated for such specific scenes, and we would be using the PyTorch library to implement it; the computation will be carried out on an NVIDIA graphics card.

Keywords: *semantic segmentation; lane detection; scene understanding; deep learning*

1 Introduction

The project addresses two issues — semantic segmentation of objects and lane detection, both of which are implemented using only image frames extracted from a video feed provided by a camera mounted on a mobile vehicle [1].

"Lane detection plays a pivotal role in autonomous driving as lanes could serve as significant cues for constraining the maneuver of vehicles on roads. Detecting lanes in-the-wild is challenging due to poor lighting conditions, occlusions caused by other vehicles, irrelevant road markings, and the inherent long and thin property of lanes" [2]. Semantic segmentation helps in providing very detailed positions of surrounding objects like barriers, humans, vehicles, etc. This project is directed towards the development of a model suited for such applications, by comparing the performance of the recently developed architectures and fine-tuning them for this purpose.

2 Rationale

"Nearly 1.25 million people die in road crashes each year, on average 3,287 deaths a day" [4]. Investigating on the critical reason, i.e. the last event in a "crash causal chain", finds human error responsible in 94 percent (±2.2%) of the cases [5]. Self-driving cars are purely analytical, relying on sensors to navigate; computers in a smart car can simply react quicker than our minds can and aren't susceptible to the many potential mistakes.

Much work has gone into solving the daunting tasks of generating an understanding of the traffic scene which is often harsh due to weather, lighting, presence of strong structure with less appearance cues, etc. [3] Some extensive datasets like the *CULane* dataset [6], and the *Berkeley DeepDrive* dataset [7], have been curated for this purpose. Quite a handful number of deep learning architectures have been recently proposed based on *VPGNet*

[8], *Spacial CNN* [3], *LaneNet* [9], *ENet-SAD* [2], etc. for lane detection. For semantic segmentation, there have been developments such as *SegNet* [10], a *mask R-CNN* model [11] based on *Feature Pyramid Network* (*FPN*) on *Microsoft COCO* dataset [12], etc. This project proposes to study and compare such existing developments to develop our model.

3 Approach

This project takes up a comparative study of recent works followed by an understanding of urban traffic scene features, to develop a model based on existing backbones and fine-tuning them. Initially, datasets would be collected and curated, followed by training models on some existing architectures and comparing statistics of their performance, for both, semantic segmentation and lane detection. A combined model would then be developed in the *Python* programming language using *PyTorch et al*, based on our understanding of the studied proposals. The computation would be carried out on an *NVIDIA* graphics card using the *CUDA* parallel computing API. Finally, the model would be tested and evaluated in comparison with the ground truth.

4 Timeline

This section establishes a basic timeline for the project scheduled from May 18 to July 3, 2020, the total duration of the research stay being slightly less than 7 weeks in duration. The tenure would be divided into 1-week slots, initially for hardware setup and interfacing, and then beginning with the study, which is to follow the collection of datasets, and finally starting work on the development of the proposed model. This timeline is subject to changes prompted by different scenarios after work commences. Lastly, it takes into account my understanding of the related domain, and that I am currently working on honing my skills further in this domain.

- **Week 1** - Initial meetings with project advisor to discuss the modus operandi and other related aspects of the project. Setting up the hardware and familiarizing with the interfaces.
- **Week 2** - Collection and (optional) curation of datasets, and preprocessing them (if required).
- **Week 3-4** - Reading up on earlier developments, building/collecting models based on said works, training/testing and collecting performance statistics.
- **Week 5-6** - Development of the proposed model by combining approaches. Alternating periods of testing and optimization to be carried out, to lower redundancy between semantic segmentation and lane detection.
- **Week 7** - Documentation is completed for the project. The final presentation of the project is also carried out during this time.

References

[1] Massimo Bertozzi, and Alberto Broggi *GOLD: A Parallel Real-Time Stereo Vision System for Generic Obstacle and Lane Detection*. IEEE Transactions on Image Processing, vol. 7, no. 1, January 1998.

[2] Yuenan Hou, Zheng Ma, Chunxiao Liu, and Chen Change Loy. *Learning Lightweight Lane Detection CNNs by Self Attention Distillation*. International Conference on Computer Vision, August 2019.

[3] Xingang Pan, Jianping Shi, Ping Luo, Xiaogang Wang, and Xiaoou Tang. *Spatial As Deep: Spatial CNN for Traffic Scene Understanding*. December 2017.

[4] *Road Safety Facts*. Association for Safe International Road Travel.

[5] *Critical Reasons for Crashes Investigated in the National Motor Vehicle Crash Causation Survey*. National Highway Traffic Safety Administration, U.S. Department of Transportation, February 2015.

[6] Xingang Pan, Jianping Shi, Ping Luo, Xiaogang Wang, Xiaoou Tang. *CULane Dataset*. Multimedia Laboratory, The Chinese University of Hong Kong.

[7] Fisher Yu, Wenqi Xian, Yingying Chen, Fangchen Liu, Mike Liao, Vashisht Madhavan, Trevor Darrell. *BDD100K: A Diverse Driving Video Database with Scalable Annotation Tooling*. May 2018.

[8] Seokju Lee, Junsik Kim, Jae Shin Yoon, Seunghak Shin, Oleksandr Bailo, Namil Kim, Tae-Hee Lee, Hyun Seok Hong, Seung-Hoon Han, In So Kweon. *VPGNet: Vanishing Point Guided Network for Lane and Road Marking Detection and Recognition*. The IEEE International Conference on Computer Vision (ICCV), 2017.

[9] Ze Wang, Weiqiang Ren, Qiang Qiu. *LaneNet: Real-Time Lane Detection Networks for Autonomous Driving*. July 2018.

[10] Alex Kendall, Vijay Badrinarayanan, Roberto Cipolla. *SegNet*. University of Cambridge.

[11] *Mask R-CNN for Object Detection and Segmentation*. GitHub Repository

[12] Tsung-Yi Lin, Michael Maire, Serge Belongie, Lubomir Bourdev, Ross Girshick, James Hays, Pietro Perona, Deva Ramanan, C. Lawrence Zitnick, Piotr Dollár. *Microsoft COCO: Common Objects in Context*. February 2015.