DEVELOPMENT OF OFFLINE METHODS FOR LOCALIZATION ON HD MAPS USING DEEP LEARNING BASED NEURAL NETWORKS FOR AUTOMATIC GROUND TRUTH GENERATION

ABSTRACT

INTRODUCTION

CHAPTER 1: AUTONOMOUS DRIVING AND HD MAPS

* Introduction to Autonomous Driving (AD): This section introduces the concept of autonomous driving and covers the SAE levels, which classify different levels of vehicle automation from driver assistance to full automation. This paragraph will emphasize the essential role of maps in autonomous driving, especially as vehicles progress towards higher levels of automation. HD maps enable autonomous systems to better understand their surroundings and make decisions by providing highly detailed spatial and semantic information.
* HD Maps: Definition and Importance: This section will begin with a brief history of digital maps, progressing from standard-definition (SD) maps to high-definition (HD) maps. It will discuss the differences, advantages, and disadvantages of each type, providing a foundation of knowledge on this key technology. HD maps enable autonomous systems to better understand their surroundings and make decisions by providing highly detailed spatial and semantic information. One significant drawback of HD maps is the high cost associated with their maintenance and creation. HD maps often need to be updated frequently to reflect real-world changes, which can be challenging and costly. As roads are constantly evolving, it’s crucial to maintain HD maps to ensure up-to-date accuracy. An alternative approach is to create HD maps in real-time, which could provide up-to-date semantic information without relying on an external provider. Neural networks can be integrated into this process, enabling the real-time generation of semantic information while avoiding the limitations of relying on third-party providers for HD map updates. In the training phase, it is necessary to align the semantic information from external sensors with HD map data, which requires precise localization to perform effectively.
* Localization on HD Maps: Due to the high level of detail in HD maps, precise localization is essential to minimize error. The goal is to leverage HD maps as if they were an additional sensor for environmental perception, enhancing the vehicle's understanding of its surroundings. Precise localization on HD maps allows autonomous vehicles to operate with greater accuracy, providing information that complements onboard sensors and enhances overall environmental awareness.
* Challenges and Limitations of Current Localization Techniques: To achieve accurate alignment, traditional deterministic optimization techniques are often employed to minimize distance errors. However, these methods have their limitations; in some scenarios, minimizing distances can lead to incorrect alignments, such as mistakenly aligning a sidewalk with a bike path border. Neural network-based alignment methods can address some of these issues and are a primary focus of this thesis. By exploring neural network approaches, this research aims to improve upon traditional optimization methods, investigating solutions that might reduce error rates and improve localization reliability in complex environments.
* State of the art: This section will review current methodologies that use neural networks for precise localization on HD maps. It will focus on approaches that address the limitations of traditional positioning methods, such as GPS, by leveraging neural networks to eliminate uncertainties. Highlighting these techniques will help establish the context and motivation for this thesis’s approach, as they represent the cutting-edge developments in localization for autonomous driving.

CHAPTER 2: MAP ALIGN: A DEEP LEARNING BASED APPROACH

* Introduction to Neural Networks: This paragraph will introduce neural networks, focusing on how they operate and how they can be applied to achieve the task at hand. It will cover fundamental aspects of neural network architecture, emphasizing convolutional neural networks (CNNs) and transformers, as both play a key role in processing complex spatial data and enabling the model to learn patterns for localization. CNNs, with their ability to capture spatial hierarchies in data, and transformers, with their advanced attention mechanisms, are particularly suited for this task.
* Architecture Description: This section provides an overview of the neural network architecture used in this project, focusing on shared characteristics across all attempted approaches. The initial concept was to develop a model that takes as input data from the vehicle (such as sensor outputs) and an HD map. Data augmentation techniques were applied to make the model more robust, with the expected output being a transformation (rotation and translation) to align the vehicle’s position with the HD map. A brief introduction to the custom network designed for this purpose will describe the network’s objectives, its target function, and the specific part of the pipeline where it simplifies the localization task.
* Dataset Description and Sensors Suite: The dataset used consists of sequences provided by the company. The data covers three major roads in the city of Parma and includes images, sensor data (such as radar readings), and GPS position information. All data were synchronized and packaged so that each image frame is associated with corresponding sensor and GPS data. This section will also discuss the sensor suite used to record the sequences of data, describing the sensors onboard the vehicle, their placements, and how they contribute to data acquisition. This provides context on the data sources and the level of detail available for training the neural network.
* Methodology: This section will outline the steps taken to solve this task, starting with defining the network’s inputs, targets, and preprocessing steps necessary to prepare the data. Preprocessing includes deserialization, as the data were acquired in a packaged and timestamped format. This ensures temporal alignment and compression of the data. It will also detail data loading steps, covering inputs such as BEV (Bird’s Eye View) representations and other essential features. The target loader will be explained, describing the solution’s objectives: predicting the coordinates (x, y, z) and the heading angle, which represents the rotation required to align the map generated from vehicle sensors with the HD map.

CHAPTER 3: MAP ALIGN: TRAINING AND MODELING APPROACHES

* Approaches and Their Pros and Cons: This section describes the two main approaches explored, although initially only one approach was planned. The original idea was to rely solely on non-image data from the sensors, bypassing the need for a full BEV (Bird’s Eye View) reconstruction, which would shorten both training and inference times and make the process faster overall. However, to achieve better results, BEV reconstruction of the world captured by the six stereo cameras on the vehicle was incorporated. Although this approach increased training times significantly, it got improved accuracy and reliability in localization. The section will discuss the trade-offs between speed and accuracy for each approach.
* First approach: This part will delve into the architecture of the first approach, detailing the layers used, network functions, and the various loss functions tested, along with the reasoning behind each choice. It will provide an in-depth overview of how this approach was structured to rely solely on data inputs without image data, as well as how it aimed to simplify the processing pipeline for faster operation. The section will also outline the pros and cons observed during testing.
* Results: Here, the results of the data-only approach will be presented, accompanied by a comparative table of the different setups tested. Key observations will include how the loss functions behaved as expected, with no signs of overfitting, which can be attributed to the strong data augmentation techniques used in preprocessing. This section will highlight the effectiveness of the first approach in specific scenarios, while also acknowledging its limitations.
* Second approach: In this approach, the architecture incorporated BEV reconstruction using images from the vehicle’s six stereo cameras, adding significant detail to the environment representation. This section will provide a detailed description of the BEV reconstruction architecture and its integration with the main network. The pros and cons of using both data and image inputs will be discussed, focusing on how the added image data enhanced accuracy, although with increased training time and computational demands.
* Results: The results of the second approach will be presented here, with another comparative table outlining the performance of the tested setups. With insights gained from the first phase, the less promising configurations were discarded early on to conserve resources. This section will summarize the improvements in accuracy observed with this approach and how these gains justify the increased resource consumption in scenarios that demand higher precision.

CHAPTER 4: ANALYSIS OF RESULTS AND FUTURE IMPROVEMENTS

* Results comparison: This section will compare the results from the two approaches, highlighting key differences. Remarkably, during inference, similar issues occurred as with traditional optimizers; specifically, without reliable reference points, lane alignment was occasionally inaccurate. However, the second approach, which incorporated a more detailed reconstruction of the external environment, largely mitigated these alignment issues, demonstrating the value of integrating BEV reconstruction and richer spatial information. This section will analyze how and why the second approach offered improvements in accuracy and stability.
* Future Developments and Real-World Applications: To be integrated effectively into the dataset creation pipeline for an RTMG (Real-Time Map Generation) network, this approach shows strong potential. However, it requires training on a significantly larger and more diverse dataset, with sequences covering a variety of scenarios, including complex environments and edge cases. Testing on datasets with missing modalities (i.e., gaps in sensor data) is also essential to ensure the approach’s robustness and generalizability, making it adaptable to various sensor suites.
* Improvements: This section will suggest further potential refinements to the network architecture, training process, and deployment strategy. Examples might include optimizing network layers for faster inference, implementing new data augmentation techniques to further reduce overfitting, or testing additional fusion strategies for combining data from various sensors.

CONCLUSION

REFERENCES/BIBLIOGRAPHY