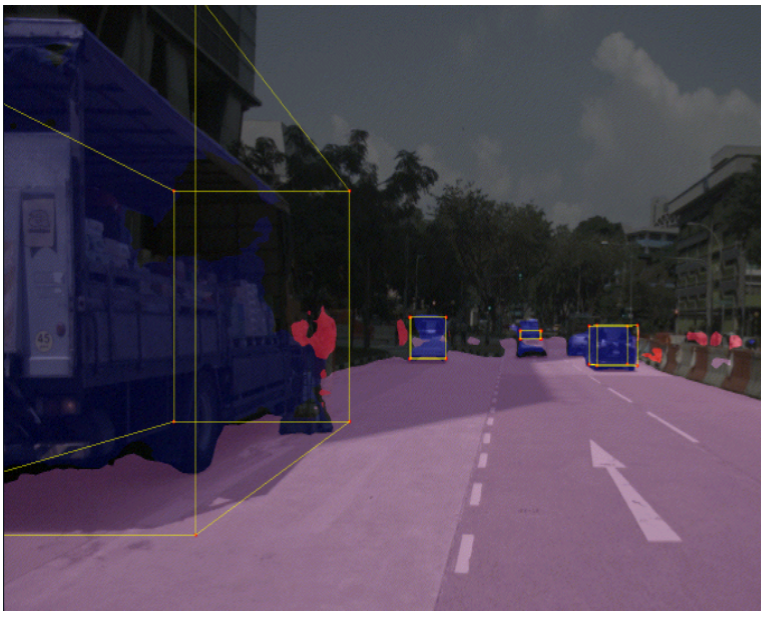
Universidad del País Vasco
Euskal Herriko Unibertsitatea
INFORMATIKA FAKULTATEA
FACULTAD DE INFORMÁTICA
Konputazio Ingeniaritza eta Sistema Adimentsuak Unibertsitate Masterra
Máster Universitario en Ingeniería Computacional y Sistemas Inteligentes

MASTER IN COMPUTATIONAL ENGINEERING AND INTELLIGENT SYSTEMS

MASTER'S THESIS

BEV2SEG_2: ANALYSIS OF SEMANTIC SEGMENTATION OF PLANAR ELEMENTS IN BIRD'S EYE VIEW AND ITS APPLICATIONS IN AUTOMATED SCENE PRE-ANNOTATION



Student: García Justel, Alan

Industrial Consultant: Sánchez Juanola, Martí
Supervisors: Dr. Barrena Orueechebarria, Nagore; Dr. Elordi Hidalgo, Unai

Academic Year: 2024-2025 **Date:** June 4, 2025

1 Introduction

The development of Automated Driving Systems (ADS) has been a hot topic in the automotive industry for the last years. These systems rely on a combination of sensors and algorithms to perform driving tasks, either partially or fully replacing the human driver. A fundamental component of any ADS is its perception system, which is responsible for detecting obstacles and generating a reliable representation of the surrounding environment.

One key element within the perception system is the generation of a local Bird's-Eye-View (BEV) map. A local BEV map provides a top-down, 2D representation of the vehicle's immediate surroundings, typically centered on the vehicle's position. Unlike raw sensor data, which is captured from the perspective of individual cameras or LiDAR units, the BEV map projects this information onto an unified ground-level plane, creating a structured grid in metric space. Each cell in the grid represents a precise location in the real world, enabling the system to interpret spatial relationships between road elements, obstacles, and free space with high accuracy.

Generating a segmented BEV map improves the perception system by providing rich semantic information and precise obstacle localization within a metric space. Semantic BEV representations are useful for a variety of tasks, including scene understanding, map reconstruction, behavior prediction of surrounding agents, and trajectory planning.

To obtain BEV semantic segmentation from cameras, traditional methods first generate semantic masks in image space and then transform them into BEV space using Inverse Perspective Mapping (IPM). Despite its simplicity, it requires accurate camera parameters and assumes a perfectly flat ground surface, which limits its effectiveness. Moreover, while planar or low-height objects such as road curbs, lane markings, and the drivable area retain a meaningful metric representation in BEV space, objects with height appear distorted after the transformation.

With the objective of addressing the afore mentioned limitations, recent methods leverage data-driven techniques for BEV representation [24] [25] [26]. However, to the best of our knowledge, no prior work has investigated the impact of training a standard semantic segmentation model directly on BEV images, with the aim of evaluating whether this approach improves performance on planar elements.

The main objective of this master's thesis is to investigate the hypothesis: *Does training a semantic segmentation model directly on BEV images outperforms the traditional image-space segmentation followed by IPM reprojection?* Additionally, this work explores a technical application of BEV semantic segmentation for anno-

tating vehicular scenes with occupancy, occlusion, and drivable area masks, contributing to the field of monocular 3D object detection given 2D semantic masks.

3 Objectives

This master's thesis aims to address key challenges in BEV semantic segmentation. While initially motivated by the investigation of optimal BEV segmentation strategies, this work also explores a tangible real-world application of these techniques in the ADS context. Accordingly, it focuses on achieving two main objectives: to analyze semantic segmentation in BEV space for planar elements by comparing a traditional strategy with a proposed alternative approach (**O1**), and to implement an annotation system for generating BEV masks for occupancy, occlusion, and drivable areas (**O2**).

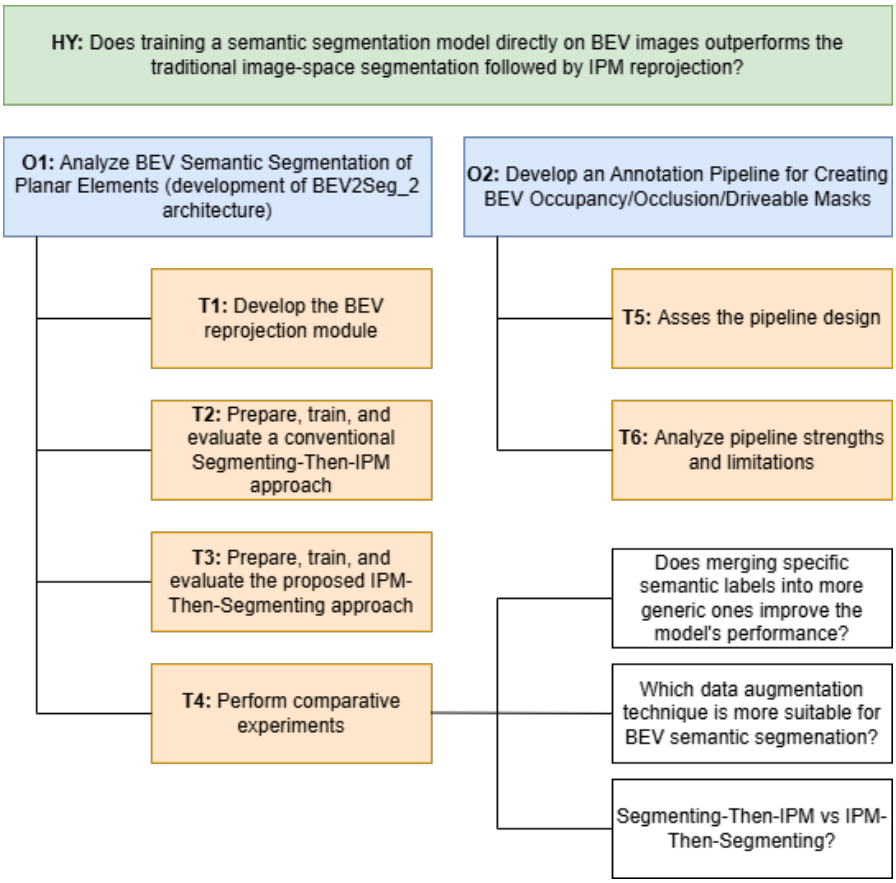


Figure 2: Thesis objectives and tasks

To address the first objective, a methodology named BEV2Seg_2 has been developed, establishing a common comparison framework between the classical image-space segmentation followed by IPM reprojection and the proposed direct BEV segmentation approach. Within this framework, a shared semantic segmentation model must be selected and adapted to both strategies to ensure consistency in

the architecture design. The model will be trained under two different configurations: using conventional perspective images (**T2**), and using BEV reprojected images (**T3**). This dual training approach will produce two versions of the model architecture, each corresponding to one of the compared strategies.

To ensure a fair and valid comparison, the same original dataset will be used as the input source for both training pipelines. From this dataset, BEV reprojected images and their corresponding semantic masks will be generated to serve as training data for the proposed BEV based approach. Consequently, a common BEV reprojection module must be developed (**T1**). This module will apply consistent transformation parameters and serve two key purposes: generating the BEV training dataset and reprojecting inferenced results from the traditional strategy. In doing so, it ensures that both models are evaluated on an identical validation set composed of BEV semantic masks, thereby allowing for a reliable and unbiased performance comparison.

The second objective of this thesis involves developing a practical application of BEV semantic segmentation by implementing an automated annotation pipeline (**T5**). This pipeline aims to generate masks in BEV space that represent occupancy, occlusion, and drivable areas from vehicular scene data. This directly addresses the need for real-world utility of BEV semantic masks for downstream tasks such as motion planning and dynamic obstacle handling. Alongside the implementation, a critical analysis of the solution will be carried out to assess its effectiveness, strengths, and limitations (**T6**).

To achieve the stated objectives, this master's thesis relied on a diverse set of software tools and hardware resources for dataset generation, model training and validation, and annotation pipeline implementation. Python was the main programming language used for most of the custom implementations. Docker was extensively used for software packaging and environment consistency, facilitating interactions with a High-Performance Computing system for model training and enabling the creation of automated systems essential for annotation generation. Furthermore, various visualization tools such as WebLABEL and Open3D were employed to support development and analysis throughout the project.

This project was conducted over a seven-month period at Vicomtech [1]. Vicomtech, is a research center in applied Artificial Intelligence, VisualComputing, and Interaction which provided the necessary hardware infrastructure and technical support for this thesis.

[1] https://www.vicomtech.org/en/