Monocular 3D Object Detection in Autonomous Driving

Alina Griesel (StudID)
Elisa Hagensieker (991850)
Madleen Uecker (StudID)

TO BE DELETED: TASK: ensure reproducibility: write a lab diary with the research question, experimental setup, methodology, and the results Topic: 3D monocular object detection in autonomous driving.

Introduction

Research Question: How can 3D monocular object detection performance be enhanced in autonomous driving systems, and what role can synthetic data play in improving the training pipeline?

The rapid advancement of autonomous driving technology has led to increased focus on developing reliable and accurate object detection systems. These systems are essential for the safe navigation of autonomous vehicles. They allow the vehicle to perceive and interpret its surroundings by identifying obstacles such as pedestrians, other vehicles, and road infrastructure in real time. Among the different techniques used for object detection, 3D monocular object detection stands out due to its cost-effectiveness and simplicity. This technique relies on a single camera rather than complex multi-sensor setups.

One potential way to improve the performance of 3D monocular object detection is by integrating synthetic data into the training process. Synthetic data, created using computer simulations or GANs, offers certain advantages. It enables the creation of large, diverse, and highly controlled datasets that can be customized for specific training requirements. This approach reduces the costs and time associated with annotating the datasets.

This project aims to test the benefits of using synthetic data to improve the overall performance and robustness of the detection model. By exposing the model to a wide range of simulated environments and conditions, including varying lighting and weather, we expect the model to be better prepared to handle unseen real-world situations. Additionally, using synthetic data will accelerate the development and testing cycles of autonomous driving systems, as it is not limited by the complex process of collecting and annotating real-world data.

Previous Work

Experimental Setup

To investigate the impact of synthetic data on 3D monocular object detection for autonomous driving, we have chosen to use data from a single-camera setup. Rather than dealing with sparse point clouds from LiDAR sensors, we need to infer depth from 2D images. This involves estimating the 3D position and depth of objects, which requires accurate interpretation of visual cues and prediction of 3D object properties.

We implement a 3D object detection algorithm, based on a pretrained model. To customize the pretrained model, we add additional layers for fine-tuning. These layers are used to refine the model's predictions taking into account the distinct characteristics of the combined dataset.

We use two datasets: real-world KITTI and synthetic Virtual KITTI dataset. The KITTI dataset provides annotated real-world images, containing various object classes. In contrast, virtual KITTI, generated using the simulation platform Unity, offers annotated images in a controlled environment, focusing on cars only. Virtual KITTI is built on basis of KITTI, using 5 KITTI scenes with 10 variations each. It simulates different weather conditions such as fog or rain, as well as various lightning conditions.

We trained two versions of the deep learning model to evaluate the impact of synthetic data on detection performance. To simplify, we focused on detecting cars only. The first model was trained using the KITTI dataset and serves as baseline model for real-world data performance. The second model was trained on the combined dataset including synthetic data, aiming to harness advantages of synthetic data.

Methodology

Datasets

Preprocessing steps To ensure vompatibility between KITTI and Virtual KITTI datasets, the following preprocessing steps were taken:

- Resizing and Normalizing Images: Images from both datasets were resized to a consistent resolution and normalized to ensure uniformity in pixel value distribution.
- Filtering Data: We filter out non-car objects from the KITTI dataset to match the labeling schema of Virtual KITTI, which only includes annotated cars. This step ensures consistency in the training data and makes the evaluation of synthetic data feasible.

Training We begin with a pretrained model. (short explanation of the architecture) To adapt the model for our specific task, w e add two convolutional layers and a ddense layer to refine the feature extraction process and obtain the

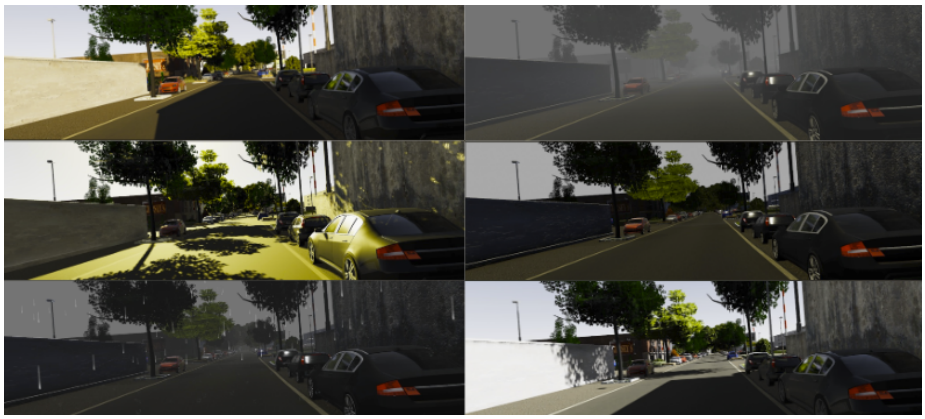

Figure 1: Samples from the same scene with different variations.
Upper left: sunset **Upper right:** fog
Middle left: RGB **Middle right:** original
Lower left: rain **Lower right:** morning

object detection output. We split the dataset into training and testing sets using an m:n ratio. Model A is exclusively trained on KITTI dataset, while Model B is trained on a combination of KITTI and Virtual KITTI.

Evaluation To evaluate the performances of the model, we use Average Precision (AP). (explanation of AP) By comparing the AP scores of Model A and Model B, we aim to quantify the impact of synthetic data on 3D monocular object detection performance.

Results

Conclusion

References