

INTISARI

Kombinasi LiDAR dan Kamera untuk Deteksi Objek Tiga Dimensi menggunakan AVOD-FPN pada Kendaraan Otonom

Oleh

Muhammad Rayyan Dzaki Santosa

22/499065/PA/21534

Kendaraan otonom membutuhkan sistem persepsi yang mampu mendeteksi objek secara andal dalam ruang tiga dimensi. Penelitian ini mengimplementasikan arsitektur AVOD-FPN dengan fusi data LiDAR yang direpresentasikan dalam bentuk *Bird's Eye View* (BEV) dan citra kamera RGB untuk deteksi objek tiga dimensi kelas Car.

Evaluasi dilakukan pada dua skenario pelatihan, yaitu skenario *in-dataset* menggunakan *dataset* KITTI serta skenario *cross-dataset* dengan *pretraining* pada data simulasi CARLA melalui CADET yang diikuti *fine-tuning* pada KITTI. Kinerja dievaluasi menggunakan Average Precision 3D (AP3D) dan Average Precision Bird's Eye View (APBEV) dengan ambang *Intersection over Union* (IoU) sebesar 0,7 pada *validation set* KITTI.

Untuk tingkat kesulitan Easy–Moderate–Hard, nilai AP3D pada skenario *in-dataset* masing-masing mencapai 89,21%, 81,12%, dan 74,49%, sedangkan pada skenario *cross-dataset* diperoleh nilai 89,62%, 75,34%, dan 74,03%. Dibandingkan dengan *baseline* deteksi objek tiga dimensi berbasis LiDAR-*only* terbaik (SECOND) dengan AP3D 87,43%, 76,48%, dan 69,10%, pendekatan sensor *fusion* AVOD-FPN menunjukkan performa yang lebih tinggi pada seluruh tingkat kesulitan apabila dibandingkan dengan skenario *in-dataset*. Dari sisi efisiensi komputasi, rata-rata waktu pelatihan per-*mini-batch* berada pada kisaran ~29,9 ms dan waktu inferensi per-*frame* pada KITTI berada pada kisaran ~205–208 ms untuk kedua skenario, yang menunjukkan tidak adanya *overhead* komputasi yang signifikan.

Kata kunci: *deteksi objek 3D, AVOD-FPN, LiDAR, kamera RGB, sensor fusion, CARLA, KITTI, cross-dataset.*

ABSTRACT

LiDAR and Camera Fusion for Three-Dimensional Object Detection Using AVOD-FPN in Autonomous Vehicles

by

Muhammad Rayyan Dzaki Santosa

22/499065/PA/21534

Autonomous vehicles require a perception system capable of reliably detecting objects in three-dimensional space. This study implements the AVOD-FPN architecture with LiDAR–RGB sensor fusion, where LiDAR data are represented in Bird’s Eye View (BEV), for 3D object detection of the Car class.

The evaluation is conducted under two training scenarios: an in-dataset scenario using the KITTI dataset and a cross-dataset scenario with pretraining on simulated CARLA data via CADET followed by fine-tuning on KITTI. Performance is evaluated using Average Precision 3D (AP3D) and Average Precision Bird’s Eye View (APBEV) with an Intersection over Union (IoU) threshold of 0.7 on the KITTI validation set.

For the Easy–Moderate–Hard difficulty levels, the in-dataset scenario achieves AP3D values of 89.21%, 81.12%, and 74.49%, while the cross-dataset scenario attains 89.62%, 75.34%, and 74.03%. Compared to the best LiDAR-only baseline (SECOND), which reports AP3D values of 87.43%, 76.48%, and 69.10%, the AVOD-FPN sensor fusion approach demonstrates consistently higher performance across all difficulty levels in the in-dataset setting. In terms of computational efficiency, the average training time per mini-batch is approximately 29.9 ms, and the inference time per frame on KITTI ranges from 205 to 208 ms for both scenarios, indicating no significant computational overhead.

Keywords: *3D object detection, AVOD-FPN, LiDAR, RGB camera, CARLA, KITTI, sensor fusion, cross-dataset.*