UNIVERSITAS
GADJAH MADA
**Evaluasi Akurasi Model Fusi Kamera-LiDAR Berbasis CLOCs Terhadap Sensor Tunggal Untuk Deteksi Objek
3D Pada Dataset Lokal Yogyakarta**
Syauqi Ikhlasun Nadhif, Dr. Dyah Aruming Tyas, S.Si.;Oskar Natan, S.ST., M.Tr.T., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

INTISARI

Evaluasi Akurasi Model Fusi Kamera-LiDAR Berbasis CLOCs terhadap Detektor Sensor Tunggal untuk Deteksi Objek 3D pada Dataset Lokal Yogyakarta

Oleh

Syauqi Ikhlasun Nadhif 22/497956/PA/21469

Deteksi objek 3D merupakan komponen penting dalam sistem persepsi kendaraan otonom, dan fusi kamera-LiDAR umumnya dilaporkan lebih akurat daripada sensor tunggal. Namun, evaluasi tersebut hampir seluruhnya dilakukan pada dataset benchmark luar negeri, sehingga belum diketahui apakah keunggulan fusi tetap teramati pada data jalan di Indonesia maupun ketika model dijalankan pada perangkat edge. Penelitian ini mengevaluasi akurasi pipeline fusi kamera-LiDAR berbasis CLOCs terhadap detektor sensor tunggal LiDAR pada dataset lokal yang direkam di lingkungan Kampus Universitas Gadjah Mada, Yogyakarta.

Dataset yang dibangun terdiri atas 378 frame berlabel kuboid 3D untuk kelas Car dan Cyclist (300 frame latih dan 78 frame validasi) yang dihasilkan dari pra-anotasi model yang dikoreksi secara manual, disertai kalibrasi kamera-LiDAR dengan galat reproyeksi hold-out median 10,8 piksel. Pipeline fusi tersusun atas PointPillars sebagai detektor 3D yang di-fine-tune dari bobot KITTI, YOLO26 sebagai detektor 2D, dan CLOCs sebagai lapisan fusi level akhir. Akurasi diukur dengan metrik AP_R40 tingkat kesulitan Moderate protokol KITTI pada objek dalam jangkauan ≤ 25 m di dalam bidang pandang kamera, menggunakan kandidat 3D yang identik untuk kedua konfigurasi. Pipeline kemudian dijalankan pada NVIDIA Jetson AGX Orin dengan engine TensorRT presisi FP32 dan FP16.

Hasil pengujian menunjukkan bahwa fusi menaikkan AP 3D kelas Car pada IoU 0,70 dari 50,98 menjadi 56,85 (+5,87; interval kepercayaan 95% [+0,93; +10,39]) dan menekan positif palsu dari 47 menjadi 16 pada recall yang setara, sedangkan kelas Cyclist praktis tidak berubah (78,94 menjadi 78,82 pada IoU 0,50) karena keterbatasannya terletak pada presisi geometri kotak 3D. Ablasi menunjukkan bahwa informasi kamera berkontribusi nyata, tetapi tidak seluruh peningkatan berasal dari isi citra yang bersesuaian. Fine-tuning dari bobot KITTI menaikkan AP Car sebesar 5,67 poin dibandingkan pelatihan dari awal. Pada Jetson AGX Orin, engine FP32 mereproduksi akurasi workstation secara identik dan FP16 mempertahankan akurasi Car (56,63) dengan percepatan engine 1,66 kali, namun pipeline fusi penuh baru mencapai sekitar 3,6 FPS sehingga belum memenuhi laju waktu nyata 10 Hz.

Kata kunci: deteksi objek 3D, fusi kamera-LiDAR, CLOCs, PointPillars, Jetson AGX Orin

UNIVERSITAS
GADJAH MADA
**Evaluasi Akurasi Model Fusi Kamera-LiDAR Berbasis CLOCs Terhadap Sensor Tunggal Untuk Deteksi Objek
3D Pada Dataset Lokal Yogyakarta**
Syauqi Ikhlasun Nadhif, Dr. Dyah Aruming Tyas, S.Si.;Oskar Natan, S.ST., M.Tr.T., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

ABSTRACT

Accuracy Evaluation of a CLOCs-Based Camera-LiDAR Fusion Model against Single-Sensor Detectors for 3D Object Detection on a Local Yogyakarta Dataset
Syauqi Ikhlasun Nadhif 22/497956/PA/21469

3D object detection is a key component of the perception system of autonomous vehicles, and camera-LiDAR fusion is generally reported to be more accurate than a single sensor. However, such evaluations have been carried out almost entirely on foreign benchmark datasets, so it remains unknown whether the advantage of fusion persists on Indonesian road data and when the model runs on an edge device. This research evaluates the accuracy of a CLOCs-based camera-LiDAR fusion pipeline against a LiDAR-only single-sensor detector on a local dataset recorded at the Universitas Gadjah Mada campus, Yogyakarta.

The constructed dataset consists of 378 frames annotated with 3D cuboids for the Car and Cyclist classes (300 training and 78 validation frames), produced through manually corrected model pre-annotation, together with a camera-LiDAR calibration with a median hold-out reprojection error of 10.8 pixels. The fusion pipeline combines PointPillars as the 3D detector fine-tuned from KITTI weights, YOLO26 as the 2D detector, and CLOCs as the late-fusion layer. Accuracy was measured using the KITTI AP_R40 metric at the Moderate difficulty level for objects within 25 m inside the camera field of view, using identical 3D candidates for both configurations. The pipeline was then deployed on an NVIDIA Jetson AGX Orin using FP32 and FP16 TensorRT engines.

The results show that fusion increases the Car 3D AP at IoU 0.70 from 50.98 to 56.85 (+5.87; 95% confidence interval [+0.93, +10.39]) and reduces false positives from 47 to 16 at equal recall, whereas the Cyclist class remains practically unchanged (78.94 to 78.82 at IoU 0.50) because its limitation lies in the geometric precision of the 3D boxes. The ablation study shows that camera information contributes substantially, although not all of the improvement originates from the corresponding image content. Fine-tuning from KITTI weights increases Car AP by 5.67 points compared with training from scratch. On the Jetson AGX Orin, the FP32 engine reproduces the workstation accuracy exactly and FP16 preserves Car accuracy (56.63) with a 1.66× engine speed-up, but the full fusion pipeline only reaches about 3.6 FPS and does not yet meet the 10 Hz real-time requirement.

***Keywords**: 3D object detection, camera-LiDAR fusion, CLOCs, PointPillars, Jetson AGX Orin*