

Metadata Matters: Fusion of Vehicle and Road Metadata for Semantic Segmentation in Autonomous Driving

Lars De Pauw

Ph.D, Researcher at EAVISE, Faculty of Engineering Technology, Sint-Katelijne-Waver, Belgium

Abstract:

We investigate whether (and how) multimodal fusion of internal vehicle signals (e.g., speed, yaw rate, temperature) and external contextual information (e.g., GPS-derived road type, traffic light presence) can enhance semantic image segmentation in autonomous driving. Building on a U-Net backbone, we implement and compare five fusion strategies at multiple encoder–decoder stages. We use the well-known Cityscapes dataset, which contains real-life driving scenes with pixel-level annotations for 19 relevant classes. This dataset is extended with GPS-derived metadata, for which we propose a two-stage fusion approach based on skip-initialized training to enable stable integration of heterogeneous inputs. Experimental results are obtained in two settings. First, fusion with ground truth class–presence vectors provides insight into optimal fusion configurations, showing that fusion at the bottleneck or early decoder stages is most effective, yielding up to +5.8% mIoU improvement. Second, we evaluate the use of real-world GPS-derived metadata. While these features provide limited overall gains, they consistently improve performance for underrepresented classes. An ablation study further analyzes the contribution of individual metadata parameters. These findings highlight the practical value of lightweight metadata fusion, demonstrating that contextual information can enhance segmentation performance when combined with an appropriate training strategy.

Keywords:

Metadata fusion, Autonomous driving, Cityscapes, Skip Initialization, Training.