SAGIVTECH IMVC 2026
15th Israel Machine Vision Conference (IMVC) 2026
Monday | April 27, 2026 | Pavilion 10, EXPO Tel Aviv

RAD: Retrieval-Augmented Monocular Metric Depth Estimation for Underrepresented Classes

Michael Baltaxe, General Motors

Monocular Metric Depth Estimation (MMDE) is essential for physically intelligent systems, yet accurate depth estimation for underrepresented classes in complex scenes remains a persistent challenge. To address this, we propose RAD, a retrieval-augmented framework that approximates the benefits of multi-view stereo by utilizing retrieved neighbors as structural geometric proxies. Our method first employs an uncertainty-aware retrieval mechanism to identify low-confidence regions in the input and retrieve RGB-D context samples containing semantically similar content. We then process both the input and retrieved context via a dual-stream network and fuse them using a matched cross-attention module, which transfers geometric information only at reliable point correspondences. Evaluations on NYU Depth v2, KITTI, and Cityscapes demonstrate that RAD significantly outperforms state-of-the-art baselines on underrepresented classes, reducing relative absolute error by 29.2% on NYU Depth v2, 13.3% on KITTI, and 7.2% on Cityscapes, while maintaining competitive performance on standard in-domain benchmarks.