

Titre : Segmentation sémantique multi-capteurs par apprentissage profond : application à l'analyse de l'état des infrastructures routières

Mots clés : Segmentation Sémantique, Adaptation de Domaine, Prédiction d'occupation, Multi-capteur

Cette thèse étudie la compréhension sémantique de scènes routières sous des contraintes industrielles : rareté et faible fiabilité des annotations 3D pour les classes d'infrastructure, fort décalage de domaine entre jeux de données, et parfois absence de certains capteurs au déploiement. Elle propose trois contributions successives. Seg3Dby2D transfère une supervision 2D issue des images vers le LiDAR grâce au rendu multi-vues, à la fusion 3D, et à la génération de pseudo-annotations, ce qui évite le besoin d'annotations 3D tout en conservant une inférence basée seulement sur le LiDAR. Seg3DbyPC2D améliore ce principe pour l'adaptation de domaine non supervisée (ADNS) en apprenant directement dans le domaine des vues rendues ; la méthode produit de meilleures pseudo-annotations, surpasse l'état de l'art de l'ADNS en réel-vers-réel, et peut-être utilisée comme pré-entraînement pour la première méthode.

Une tâche proche de la segmentation sémantique, la prédiction d'occupation sémantique, est ensuite explorée afin de prévoir le cas où un capteur LiDAR n'est pas disponible. La méthode FreeOcc est proposée, prédisant l'occupation sémantique et panoptique de scènes 3D uniquement à partir de caméras, et à « vocabulaire ouvert », en combinant prédiction sémantique et reconstruction géométrique à partir de modèles de fondation, identification des instances et raffinement des grilles d'occupation 3D. L'ensemble montre que les données non annotées, le post-traitement hors ligne et la complémentarité capteurs peuvent compenser le manque de supervision 3D, tout en mettant en évidence les principaux verrous restants pour un déploiement industriel robuste.

Title : Multi-sensor semantic segmentation through deep learning: application to the analysis of the state of road infrastructures

Keywords : Semantic Segmentation, Domain Adaptation, Occupancy Prediction, Multi-sensor

Abstract : This thesis studies the semantic understanding of road scenes under industrial constraints: scarce and low-reliability 3D annotations for infrastructure classes, strong domain shift across datasets, and, in some cases, the absence of certain sensors at deployment. It proposes three successive contributions. Seg3Dby2D transfers 2D supervision from images to LiDAR through multi-view rendering, 3D fusion, and pseudo-label generation, thereby avoiding the need for 3D annotations while preserving LiDAR-only inference. Seg3DbyPC2D improves this principle for unsupervised domain adaptation (UDA) by learning directly in the rendered-view domain; the method produces better pseudo-labels, outperforms the state of the art in real-to-real UDA,

and can be used as pretraining for the first method. A task closely related to semantic segmentation, semantic occupancy prediction, is then explored to address the case where a LiDAR sensor is not available. The proposed method, FreeOcc, predicts semantic and panoptic occupancy of 3D scenes from cameras only, in an open-vocabulary setting, by combining semantic prediction and geometric reconstruction from foundation models, instance identification, and 3D occupancy-grid refinement. Overall, the thesis shows that unlabeled data, offline post-processing, and sensor complementarity can compensate for the lack of 3D supervision, while also highlighting the main remaining bottlenecks for robust industrial deployment.