Self-Supervised Traversability Estimation via LiDAR-Camera Fusion on Construction Sites

Jinhee Yu[1], Monika Jayakumar[1], Yilong Chen[2], Yong Cho[2], Jingdao Chen[1]

(1) Computer Science and Engineering, Mississippi State University (2) Civil and Environmental Engineering, Georgia Tech

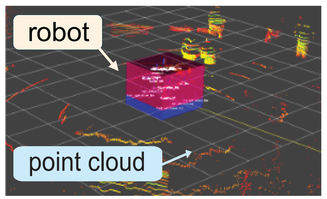
Background

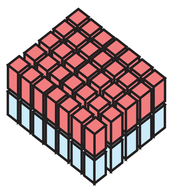
- Autonomous robots are increasingly used for a variety of tasks on construction sites such as earthmoving, concrete printing, and drilling.
- To ensure safe navigation, it is crucial for ground robots to understand which areas are traversable.
- Collecting labeled data in construction environments is challenging due to their unstructured and dynamic nature.

Goal

- Use self-supervised learning with demonstration trajectories and robot geometry as pseudo-labels.
- Combine camera and LiDAR data to obtain both visual and geometry features.
- Compare early vs. late fusion strategies.

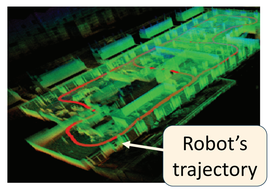
Pseudo Labeling



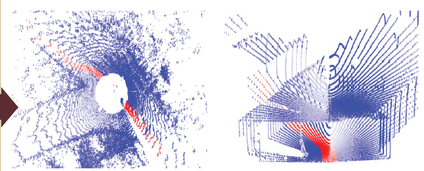


Traversable: $(p \in \mathrm{OBB}_{\text{Ground to Top}}) \wedge (p \notin \mathrm{OBB}_{\text{Chassis to Top}})$

else, Unlabled



Reconstructed trajectory from SLAM



Pseudo-labels

Datasets

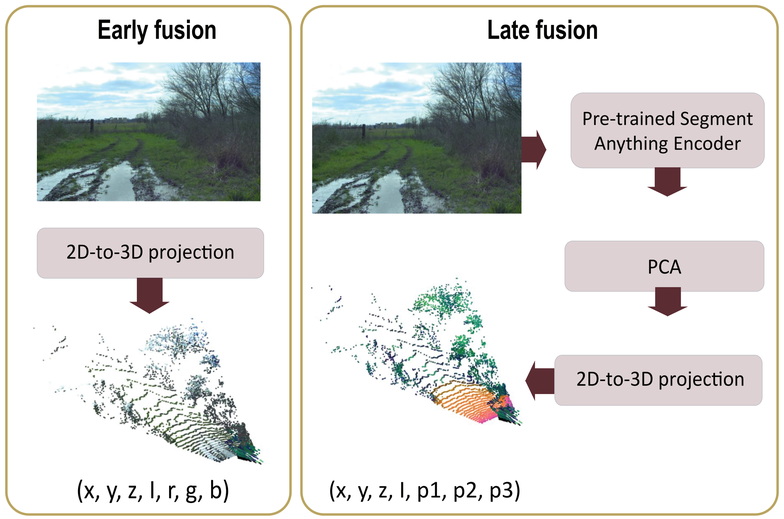
RELLIS-3D (Jiang et al. 2020): multimodal off-road driving dataset

- Semantic segmentation ground truth labels are provided, enabling quantitative evaluation.
- To ensure safe navigation, it is crucial for ground robots to understand which areas are traversable.

ConSLAM (Trzeciak et al. 2023): construction site dataset

- No ground truth segmentation labels; only quantitative evaluation performed.
- Per-scan poses obtained using SLAM (FAST-LIO) (Xu et al. 2021).

Early fusion & Late fusion



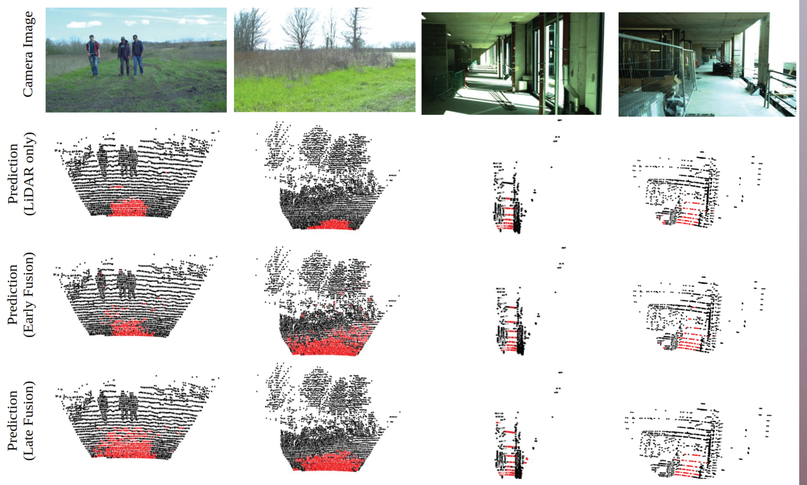
- Feature embeddings are fed into Sphereformer (Lai et al. 2023), a transformer-based model for 3D point cloud segmentation.
- Non-negative Positive-Unlabeled (nnPU) loss (Kiryo et al. 2017) is used to train with only positive and unlabeled data.

Results

RELLIS-3D: quantitative

Method	Threshold	Accuracy	Precision	Recall	F1 Score	AUROC
LiDAR only	0.5	0.26	0.90	0.37	0.53	0.70
Early Fusion	0.5	0.36	0.88	0.65	0.75	0.68
Late Fusion	0.5	0.30	0.90	0.47	0.62	0.69
LiDAR only	0.7	0.20	0.90	0.22	0.35	0.70
Early Fusion	0.7	0.24	0.92	0.31	0.47	0.68
Late Fusion	0.7	0.22	0.90	0.27	0.41	0.69
LiDAR only	0.9	0.15	0.89	0.10	0.17	0.70
Early Fusion	0.9	0.16	0.93	0.10	0.18	0.68
Late Fusion	0.9	0.16	0.90	0.12	0.21	0.69

RELLIS-3D: qualitative **ConSLAM: qualitative**



Conclusion

- Our model produced reasonable traversability estimation results by training only on demonstration trajectories without any manual labels.
- Both early fusion and late fusion strategies outperformed the LiDAR-only method.
- Future work includes exploring alternative embedding strategies for visual features and addressing the loss of FoV from 2D-to-3D projection.

Contact: jy603@msstate.edu