KI DELTA LEARNING
Scalable AI for Automated Driving

Active Learning On Dynamic Scenes Using Multi View Consistency

Daniel Derkacz-Bogner, Norman Müller

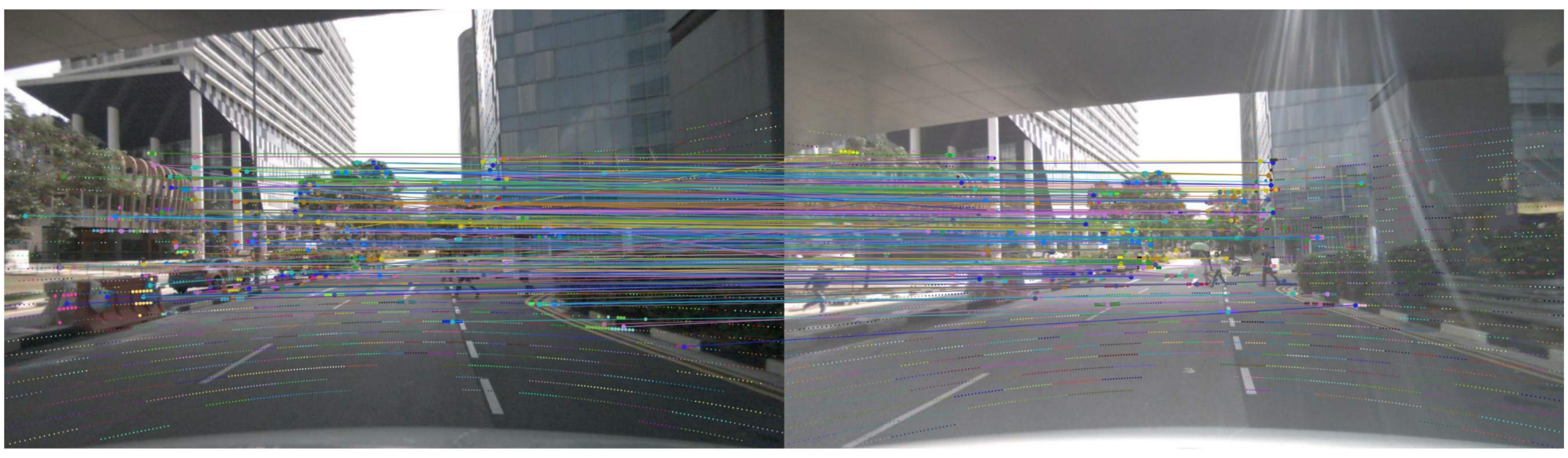

Figure 1: Qualitative evaluation of the SIFT-LIDAR matching (© TUM)

Method	5%	10%	15%
Random	20.89	23.15	24.55
Monte Carlo Dropout	25.99	27.24	25.91
Ours	24.35	28.08	27.71

Table 1: Evaluation of our proposed method using 5%, 10%,15% of the nuScenes dataset.

Less can be more

Deep learning models need very large datasets to reach high accuracy. While true, high amounts of data do not always lead to better results, if the data is not well sampled. Thus, sampling data in a clever way can keep the data size low while increasing performance. Active learning investigates how the model being trained can decide which data sample would lead to an increase in results.

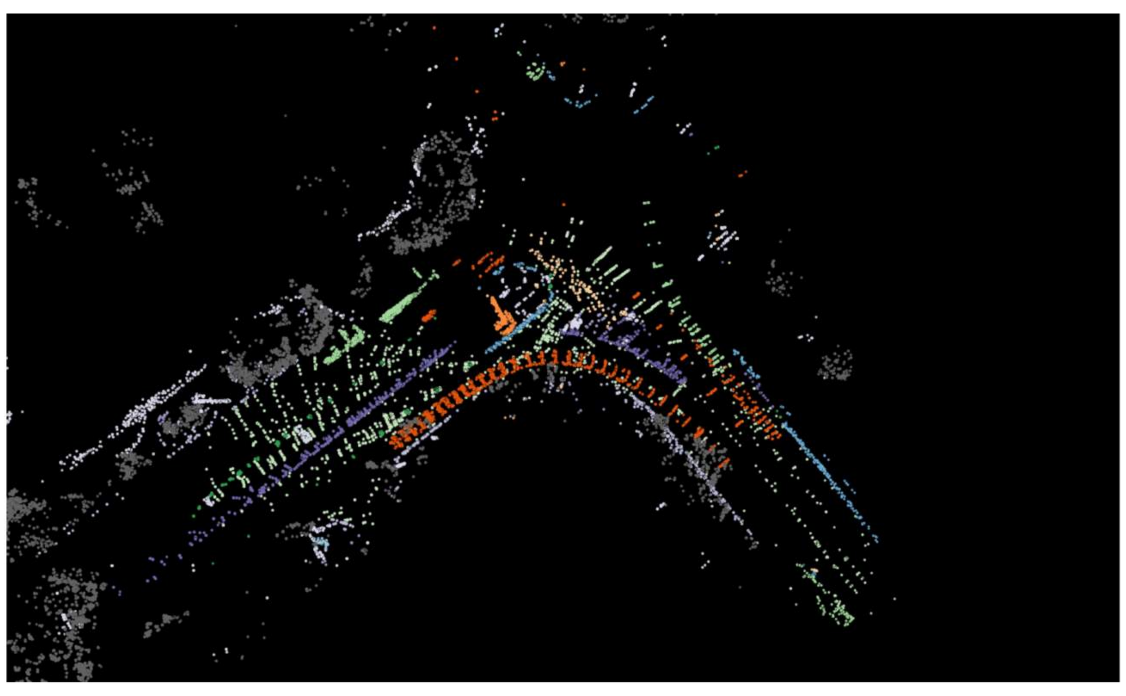

Figure 2: A reconstructed scene using only SIFT-LIDAR correspondences. Our proposed matching can capture the static and dynamic objects in a scene. (© 1. TUM)

See it from another angle

Enforcing consistent predictions through a multi-view setting showed promising results on static scenes [1], which we modified to work on dynamic outdoor scenes [2]. We enforce consistent predictions over time using the predictive uncertainty of each point $p_{i,t}$ at time step t compared to $p_{i,t+n}$.

To generate correspondences ($p_{i,t}$, $p_{i,t+n}$), we investigated two methods, (1) a nearest neighbor matching which only works on static scenes and (2) a keypoint matching approach, using 2D SIFT descriptors and a distance based SIFT-LIDAR matching.

Evaluation And Future Work

We evaluate our approach on the nuScenes [2] dataset using Cylinder3D [3] as our segmentation backbone.

As seen in Table 1, our method is able to increase the segmentation performance compared to our baselines, with the drawback of only working on static scenes.

While the SIFT-LIDAR matching is able to generate good point correspondences (see Figure 1 and Figure 2), the amount of such is not enough for an active learning approach.

Future directions may include the use a learned point encoders which may need to be learned for each new dataset, reducing the practicallity of active learning. A more promising approach would be the use of contrastive learning [4], optimized simultaneously with the activie learning method.

References:

[1] Siddiqui, Y., Valentin J., Nießner M. "Viewal: Active learning with viewpoint entropy for semantic segmentation." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020

[2] Caesar, H., et al. "nuscenes: A multimodal dataset for autonomous driving." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020

[3] Xinge, Z. et al. "Cylindrical and asymmetrical 3d convolution networks for lidar segmentation." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021

[4] Chen, Y., Nießner, M., & Dai, A. (2022). 4dcontrast: Contrastive learning with dynamic correspondences for 3d scene understanding. In European Conference on Computer Vision(pp. 543-560). Springer, Cham.

Sensors / Active Learning

Partners

External partners

For more information contact:
daniel.derkacz@tum.de
norman.mueller@tum.de

KI Delta Learning is a project of the KI Familie. It was initiated and developed by the VDA Leitinitiative autonomous and connected driving and is funded by the Federal Ministry for Economic Affairs and Climate Action.

KI FAMILIE

Supported by:
Federal Ministry for Economic Affairs and Climate Action
on the basis of a decision by the German Bundestag