Supplementary Materials for "Learning to Detect 3D Lanes by Shape Matching and Embedding"

Ruixin Liu, Zhihao Guan, Zejian Yuan
Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, China
{sweetylrx, duduguan}@stu.xjtu.edu.cn yuan.ze.jian@xjtu.edu.cn

Ao Liu, Tong Zhou, Tang Kun, Erlong Li, Chao Zheng, Shuqi Mei
T Lab, Tencent Map, Tencent, China
{allliu, kyriezhou, kunntang, erlongli, chrisczheng, shawnmei}@tencent.com

In this supplemental document, we provide:

§A 4-channel BEV pseudo-images for BEV-3DLanes dataset.

§B More visualization results of 3D predictions in various scenarios.

§C Visualizations of the intermediate results.

A. 4-channel BEV pseudo-images for BEV-3DLanes dataset.

3D scenes for experiments are collected by LiDAR sensors. In contrast to the previous point cloud datasets, our raw data are aggregated into high-density ones through specific processing, which overcomes the semantic information loss caused by sparse point cloud distributions. The raw point cloud is shown on the left of Figure 1. Since lane detection requires less attention to the surroundings, a fixed-size 3D window is set to filter out the noise points. The rest points are first rasterized and then projected onto a bird's-eye view, generating a 4-channel pseudo-BEV image. The 4 channels include mean intensity, density, elevation difference, and minimum elevation, which are demonstrated on the right of Figure 1.

The intensity channel preserves the mean intensity information for each pixel, distinguishing the lane markings from roads and other surroundings. The density channel records each pixel with a normalized value that encodes the number of point clouds inside each spatial cubic. The point density far from the trajectory is sparse, while higher density indicates the existence of objects and the area near the driving trajectory. The elevation difference channel measures the relative difference between the maximum and minimum elevation, which can help distinguish objects with a slender shape similar to the lanes, such as fences and curbs. The minimum elevation stores the absolute elevation information provided by LiDAR.

B. More visualization results of 3D predictions in various scenarios.

Since our DSANet provides lane detection results in 3D space, the predictions projected on pseudo-BEV images are inversely calculated to the 3D coordinates. Figure 2 respectively visualizes the prediction results in the pseudo-BEV images and the local 3D coordinate systems. Specifically, complex scenarios including curves, forks, dense lanes and gradients are demonstrated to confirm the robustness of our DSANet. We find that although our method achieves a noticeable precision improvement, there is still a lack of smoothness constraints among the grids, resulting in unsmooth outputs, especially in the elevation axis. Our further research will take such deficiencies into consideration.

C. Visualizations of the intermediate results.

Figure 3 shows the visualizations of intermediate results of our DSANet in various complex scenarios, including curves, dense lanes and forks. The first row shows a set of rough instance shapes predicted by the instance shape branch. Although the predictions are of low precision, it is sufficient to *distinguish the different instances*. In the experiments, the start and the end of the lane instance are connected to obtain a straight line segment, denoted as instance shape. The center coordinates, length and radian of the straight line segment make up the 2D Gaussian distributions of each ground truth $\bar{\mathcal{N}}_l$, and the predicted 2D Gaussian distribution $\mathcal{N}_l$ is output by the instance shape branch.

Besides, the segment shape branch predicts grid-level segment shapes with precise locations, lengths and yaw angles. The visualizations are shown in the second row, where the points localize the centers of the local segments, the long axes and the angles of the ellipses represent their lengths and radians. In reality, *the predicted ellipses are indiscriminate*, we assign them with different colors for better visual distinction. The ellipses are the visualizations of predicted 2D Gaussian distribution, which are denoted as the prediction $\mathcal{N}_s$ in the main text. The ground truth $\bar{\mathcal{N}}_s$ of segment shape is represented by the properties of segments inside

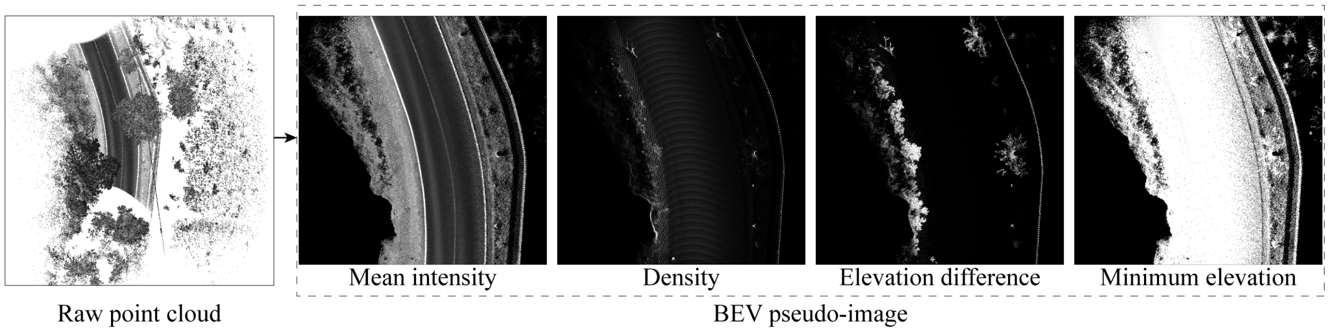
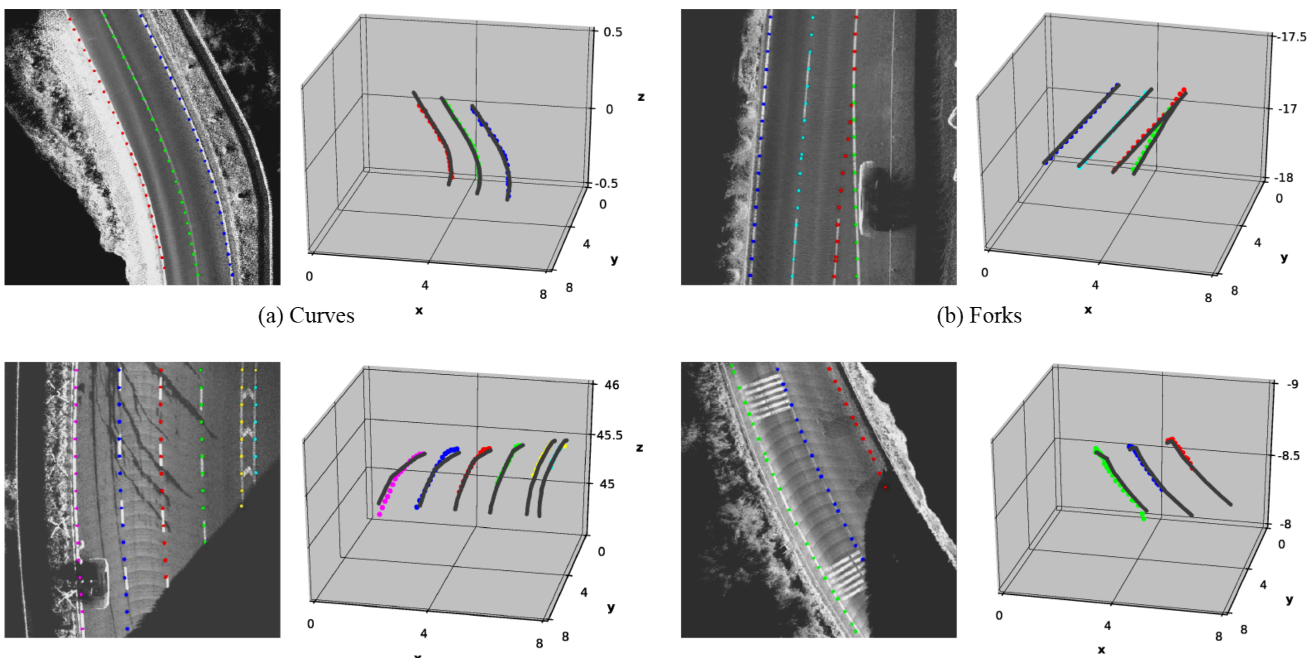


Figure 1. Raw point cloud and the 4-channel pseudo-BEV image.



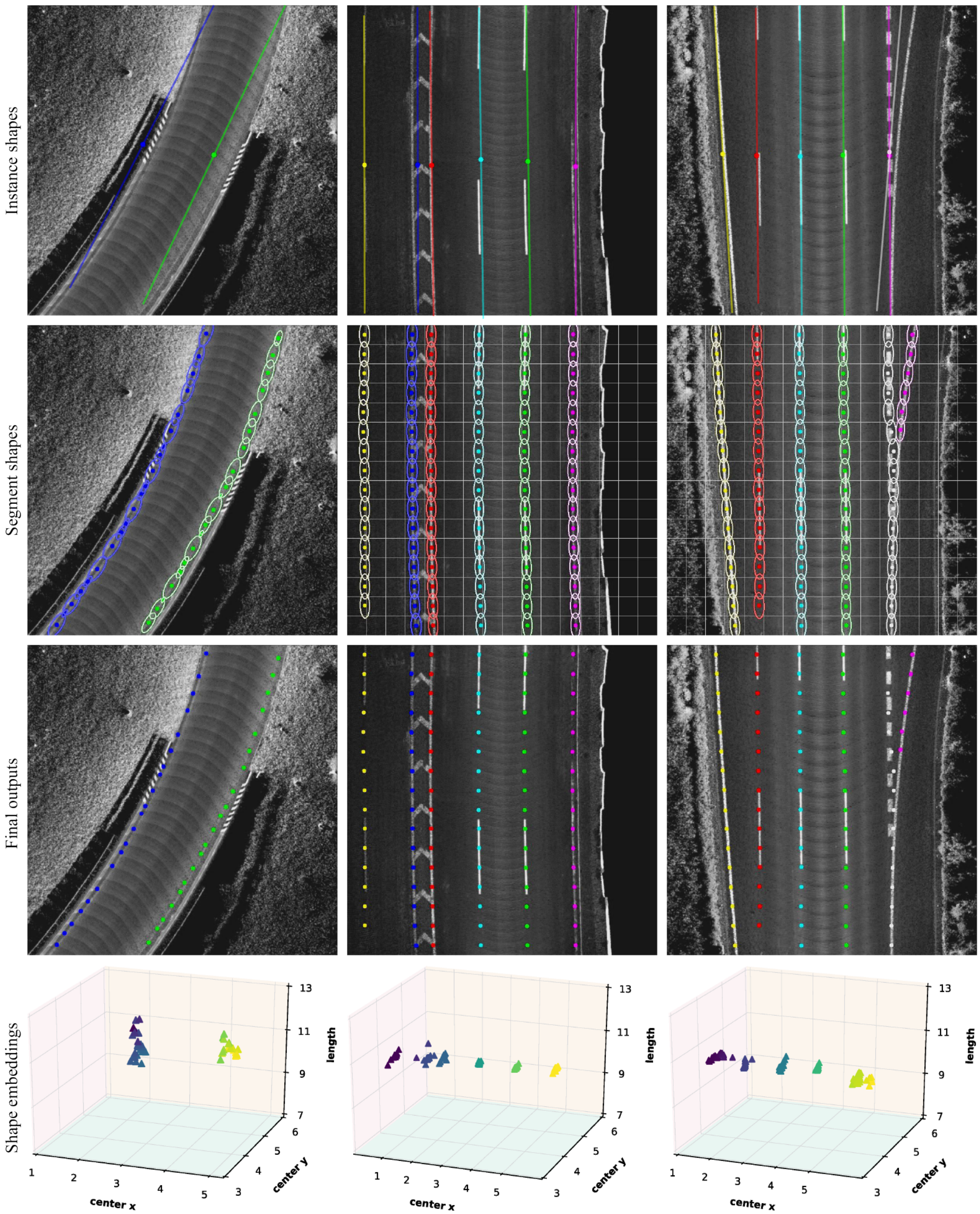
(a) Curves (b) Forks

(c) Dense lanes (d) Gradients

Figure 2. Visualizations of 3D results in more scenarios, including (a): Curves; (b): Forks; (c): Dense lanes; (d): Gradients. The 3D results are highly precise with correct instance discrimination, but remain to be improved in smoothness.

the grid, as the "Dual-Level Lane Shape Representation" section defined.

The third row shows the final predicted lane points with instance discrimination for visualizations. And the last row shows the grid-level shape embedding predictions in the feature space. The shape embeddings correspond one-to-one to the segment shapes, which are shown in Figure 3 in the form of triangles. The ground-truth 2D Gaussian distributions of shape embeddings $\bar{\mathcal{N}}_e$ are the same as $\bar{\mathcal{N}}_l$, and the predictions $\mathcal{N}_e$ are acquired with the contextual information from a local perspective. That is to say, each segment votes the instance it belongs to by measuring the differences between its corresponding shape embeddings and all the instance shapes. As shown in Figure 3, triangles distributed together are likely to belong to the same instances, while those distributed far apart tend to belong to different instances. It can be noticed that the scenarios with large curvature lanes predict imprecise shape embeddings, but the segments aggregation is well done due to the correct global guidance provided by the instance shapes.



(a) Curves (b) Dense lanes (c) Forks

Figure 3. Visualizations of the intermediate results of our DSANet. (Best viewed in colors). The three columns demonstrate complex scenarios including (a): Curves; (b): Dense lanes; (c): Forks. The four rows respectively show: the instance shapes output by the instance shape branch, the grid-level segment shapes output by the LS head, the final outputs and the grid-level shape embeddings output by the SE head.