Supplementary material (Bird's Eye View Segmentation Using Lifted 2D Semantic Features)

BMVC 2021 Submission # 772

1 Additional EPOSH dataset details

1.1 Dataset creation

As mentioned in the main draft, we used COLMAP to reconstruct a 3D dense point cloud given a video clip. As a first step, we sampled images with resolution 1920×1080 at 12 frames per second from the input video. Apart for the following settings, default settings were used everywhere during reconstruction,

- radial 'RADIAL_FISHEYE' camera model
- 'exhaustive_matcher' feature matcher is used for the feature matching step.
- Memory cache size is set to 40 GB wherever applicable.
- 'PatchMatchStereo.window_radius' is set to 15 to make reconstruction more dense. During image undistortion, the parameter 'max_image_size' is set to 4000.

The output produced by COLMAP is not to metric scale. Therefore, each point cloud reconstructed by COLMAP is rescaled to metric scale.

1.2 Dataset classes and attributes

In Table 1 (of the main draft), all topology related classes in the first column except for 'Road curb line' are annotated with polygons. 'Road curb line' is annotated with polylines[1].

Lane lines and cross-walks have orientation as an attribute. We defined orientation similar to how [2] have defined for the BDD100k dataset. Orientation is either parallel or perpendicular to the direction of movement of ego car.

Each lane instance is also annotated for affordance, defined as the set of action choices 'afforded' to a vehicle, given its location. The affordance classes are defined in the last row of the planning related section in Table 1 (of the main draft). Figure 9 (b) shows an example of affordance annotation from the EPOSH dataset. Figure (b) shows the different types of Symbolic Road Markings annotated in the dataset. Figure 10 illustrates the 'inside intersection area' class. Figure 11 shows some examples from the EPOSH perspective dataset. Figure 12 shows some examples from the EPOSH BEV dataset.

© 2021. The copyright of this document resides with its authors.
It may be distributed unchanged freely in print or electronic forms.

[1] A polyline is a connected sequence of line segments created as a single object.

Figure 9: Subplot (a) shows the 5 different Symbolic Road Marking attributes annotated in the the EPOSH dataset dataset - Straight, Left, Right, Straight-Left & Straight-Right. All other SRMs are annotated as 'other'. Subplot (b) shows an example of affordance annotations from the EPOSH dataset.

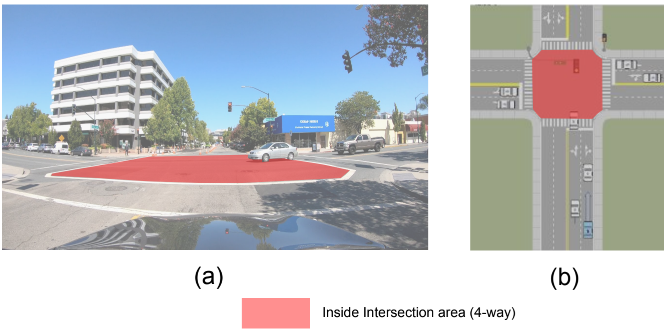

Figure 10: An example of the class 'inside intersection area'. The class has attributes as the type of intersection (3-way or 4-way). In the example in the figure, this type is 4-way intersection.

2 Additional qualitative results

The supplementary material folder also contains a video titled '**output.mp4**' which shows qualitative results from the NuScenes dataset using our method. All the 6 surround cameras from the NuScenes dataset are used as input. Note that the our method uses monocular camera image as input for making predictions. Monocular BEV segmentation prediction for each camera is stitched to create surround prediction. The corresponding ground truth is also shown in the video. In the Ground Truth (GT) for NuScenes BEV dataset produced by [1], occluded areas are masked as seen in the video. These areas are masked in the prediction as well.

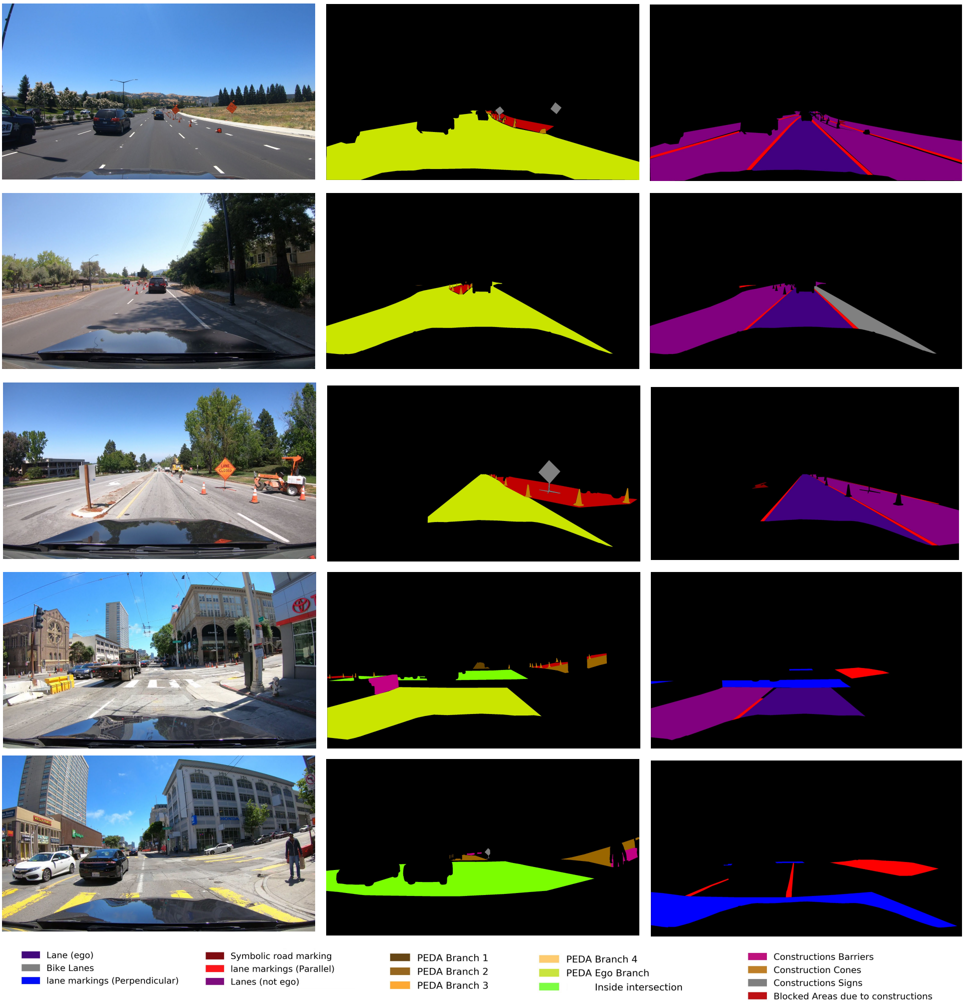

Figure 11: Some examples from the EPOSH persepctive dataset. The first column shows the image captured from camera, the second column shows planning related classes and the third column shows topology related classes. Some classes like crosswalks, lane lines are merged and shown with a single color for easier visualization. All attributes in the dataset are also not shown in the visualization.

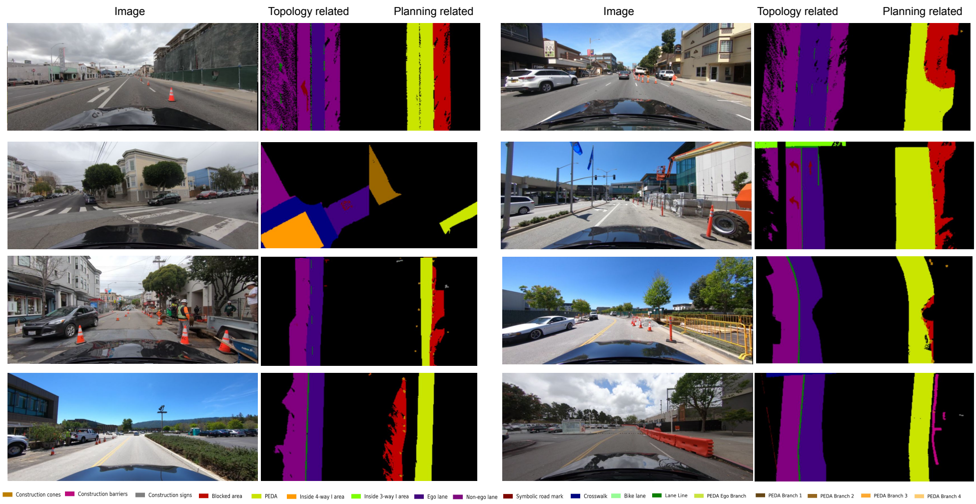

Figure 12: Some examples from the EPOSH BEV dataset.

References

[1] Thomas Roddick and Roberto Cipolla. Predicting semantic map representations from images using pyramid occupancy networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11138–11147, 2020.

[2] Fisher Yu, Haofeng Chen, Xin Wang, Wenqi Xian, Yingying Chen, Fangchen Liu, Vashisht Madhavan, and Trevor Darrell. Bdd100k: A diverse driving dataset for heterogeneous multitask learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2636–2645, 2020.