Supplementary Material for
BEV-DG: Cross-Modal Learning under Bird's-Eye View for Domain Generalization of 3D Semantic Segmentation

Miaoyu Li[1], Yachao Zhang[2 ‡], Xu Ma[3], Yanyun Qu[1 ‡], Yun Fu[3]

[1]School of Informatics, Xiamen University

[2] Tsinghua Shenzhen International Graduate School, Tsinghua University

[3] Department of ECE, Northeastern University

limiaoyu@stu.xmu.edu.cn, yachaozhang@sz.tsinghua.edu.cn, yyqu@xmu.edu.cn

1. Analysis of DVM

Since the LiDAR configuration is reflected in the global density of the point cloud, the domain attribute of a sample should be characterized by its global feature. In light of this, in BEV-driven domain contrastive learning, we must transform the BEV feature map with area-level information into a feature with global point cloud information. However, the points are not uniformly distributed in the point cloud. Specifically, the density of the areas near the LiDAR is greater than the areas away from the LiDAR. Therefore, to generate the global feature with the perception of density, we cannot equally treat different areas of the point cloud.

By observing the area distribution in the BEV space, we find that the percentage of each type of area can well embody the point cloud density, as shown in Fig. 3 of the main paper. Based on this, we introduce DVM to generate the global feature with the perception of point cloud density. The whole process can be formulated as Eq. 7 in the main paper. First, we extract the feature of each type of area by max pooling, *i.e.*, $MAX\left(f^{bev}_{[1,10)}\right)/MAX\left(f^{bev}_{[10,50)}\right)/MAX\left(f^{bev}_{[50,+\infty)}\right)$. Next, we use the percentage as the weight of each type of area, *i.e.*, $\frac{N_{[1,10)}}{N_{all}}/\frac{N_{[10,50)}}{N_{all}}/\frac{N_{[50,+\infty)}}{N_{all}}$, to generate the global feature from the weighted summation of three types of area features. In this way, we generate the global feature by treating areas differently based on the area distribution, which embodies the point cloud density of different datasets. As a result, DVM can generate a density-maintained global feature (BEV vector) with distinct domain attributes.

2. Additional Qualitative Results

In order to verify the performance of our BEV-DG more comprehensively, we additionally give more qualitative results on three designed domain generalization settings, *i.e.*, A,S→N, A,N→S and N,S→A. From Fig. 1 - 3, we can observe that our BEV-DG not only precisely identify the large targets (*e.g.*, "car" and "truck"), but also works well for small targets (*e.g.*, "person" and "bike"). These results demonstrate the effectiveness of our BEV-DG.

‡Corresponding Author

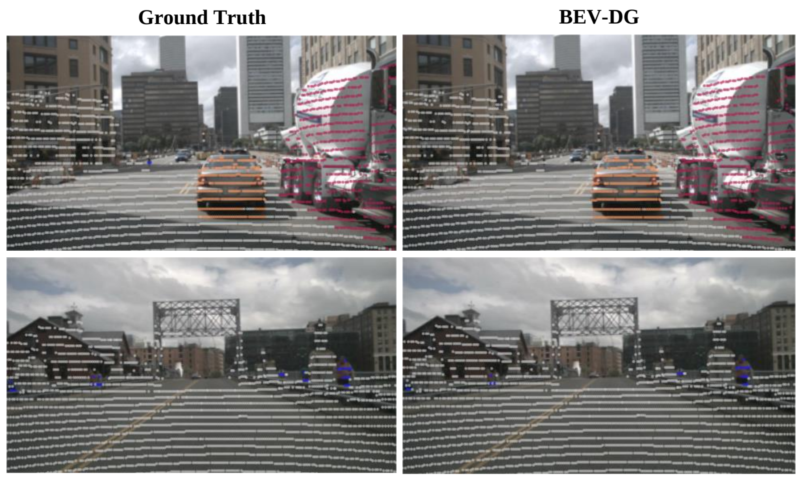

Figure 1. Results on A,S→N.

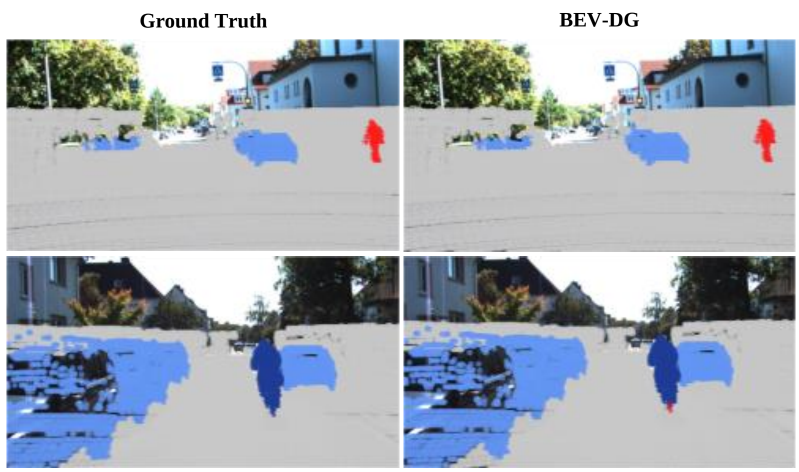

Figure 2. Results on A,N→S.

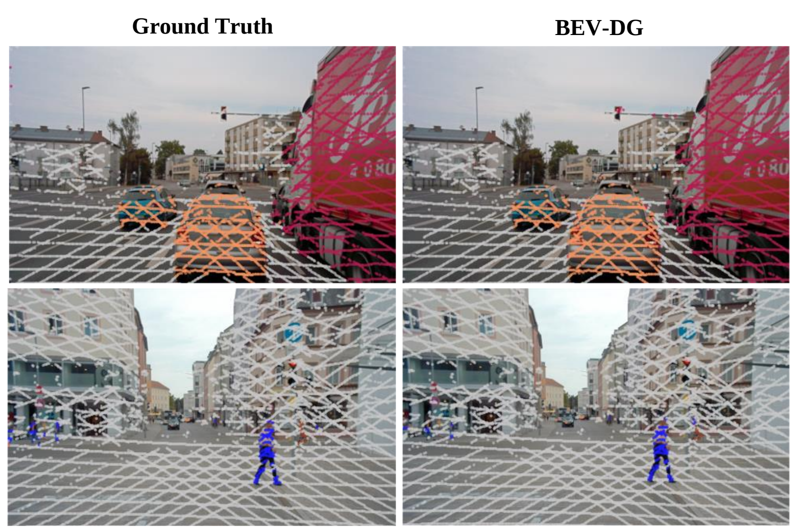

Figure 3. Results on N,S→A.