Self-Supervised Class-Agnostic Motion Prediction with Spatial and Temporal Consistency Regularizations

Supplementary Material

We provide the following contents in the supplementary.
- Implementation details on solving the optimal transport problem;
- Details of the breadth-first clustering;

1. Solving Optimal Transport Problem

The solution to an entropic optimal transport:

$$\begin{aligned} \pi^* &= \arg\min_{\pi} \sum_{i,j} C_{ij}\pi_{ij} + \varepsilon \pi_{ij} \log \pi_{ij} \\ s.t. \quad & \pi \mathbf{1}_n = \frac{1}{m}\mathbf{1}_m, \pi^T \mathbf{1}_m = \frac{1}{n}\mathbf{1}_n, \end{aligned} \tag{1}$$

is unique and has the form:

$$\pi^* = Diag(a) K Diag(b), \tag{2}$$

where $a \in \mathbb{R}^m$, $b \in \mathbb{R}^n$, and $K \in \mathbb{R}^{m \times n}$ is a matrix with $K_{ij} = \exp\left(-\frac{c_{ij}}{\varepsilon}\right)$. ε is 0.03.

We can utilize the following steps to solve this problem [1]:

1. Initialize $b = \frac{1}{n}\mathbf{1}_n$.
2. Update a by $Diag(a)Kb = \frac{1}{m}\mathbf{1}_m$.
3. Update b by $Diag(b)K^T a = \frac{1}{n}\mathbf{1}_n$.

Repeating the above process for N iterations, we can find an approximate solution of π^* according to Eq. 2. In the experiments, we set N to 3.

2. Breadth-First Clustering

We first remove the majority of ground plane points by Patchwork++ [2] for better clustering. We evaluate the effectiveness of the ground removal in Table 1. As it indicates, we can correctly segment the majority of ground points. The results of ground plane removal are shown in the second row of Fig. 1. We incorporate the neighboring cells of each cell within the Bird's Eye View (BEV) map into the cluster, and continue this process iteratively by including the neighbors of the neighbors until no more cells can be included. Details are presented in Algorithm 1.

Table 1. The precision, recall, and F1-score of the employed ground segmentation algorithm on the nuScenes dataset.

Prec.	Recall	F1
0.95	0.92	0.94

In Algorithm 1, $B = \{b_i \in \mathbb{R}^2\}_{i=1}^N$ is the 2D coordinates list, where N is the number of non-empty cells. M_l records which cluster each cell belongs to. H and W are the height and width of the BEV map respectively. We consider cells within d_c (d_c is set to 3 in the paper) City-Block distance as neighbors. The clustering results are shown in the third row of Fig. 1. We utilize two metrics, OC and UC, to evaluate the accuracy of clustering, as shown in Table 2. UC indicates the number of BEV cells in an instance that are clustered into the same cluster, providing insight into the degree of under-clustering. A 100% UC value implies that all cells within the instance are clustered into a single cluster, signifying no under-clustering. OC indicates the number of BEV cells in a cluster that belong to the same instance, providing insight into the degree of over-clustering. A 100% OC value indicates that all cells in the cluster belong to the same instance, thereby demonstrating there is no over-clustering. As depicted in the table, the breadth-first clustering algorithm can achieve a good balance between under-clustering and over-clustering (UC 90.4% (*i.e.* 9.6% under-clustering rate) and OC 88.9% (*i.e.* 11.1% over-clustering rate)). Although there will be some failure cases,

Algorithm 1 Breadth-first clustering

Input: B

$\texttt{Label} \leftarrow 1$, $M_l \leftarrow zeros(H \times W)$
repeat
  $b \leftarrow B.\texttt{pop}$
  if $M_l(d)$ is $zero$ **then**
    `continue`
  end if
  $L_b \leftarrow \texttt{list}(b)$
  repeat
    $d \leftarrow L_b.\texttt{pop}$
    for $d_n \in \texttt{Neighborhood}(d)$ **do**
      if $M_l(d_n)$ is $zero$ **then**
        $L_b.\texttt{append}(d_n)$, $M_l(d_n) \leftarrow \texttt{Label}$
      end if
    end for
  until L_b is empty
  $\texttt{Label} \leftarrow \texttt{Label} + 1$
until B is empty

Output: M_l

Table 2. Cluster Accuracy.

UC	OC
90.4%	88.9%

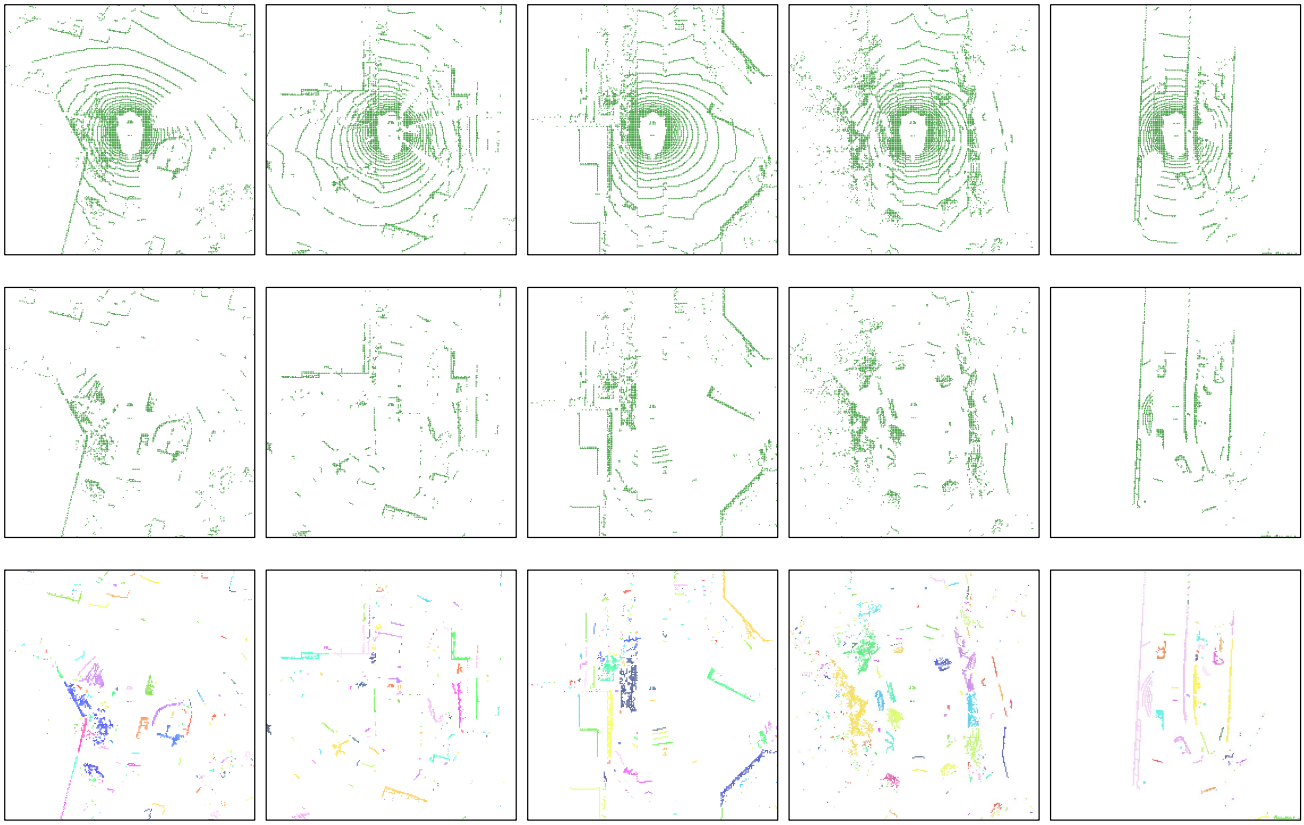

Figure 1. **Results of ground segmentation and clustering.** First row: BEV maps discretized from LiDAR point clouds; Second row: Results of ground removing; Third row: Results of clustering (clusters are with different colors). Zoom in for the best view.

it's noteworthy that cluster consistency regularization can still provide the model with more accurate information during training due to the larger number of accurate clustering outcomes.

References

[1] Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Neural Information Processing Systems (NeurIPS)*, 2013. 1

[2] Seungjae Lee, Hyungtae Lim, and Hyun Myung. Patchwork++: Fast and robust ground segmentation solving partial under-segmentation using 3d point cloud. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2022. 1