Supplementary Material: Novel Class Discovery for 3D Point Cloud Semantic Segmentation

Luigi Riz[1] Cristiano Saltori[2] Elisa Ricci[1,2] Fabio Poiesi[1]

[1]Fondazione Bruno Kessler [2]University of Trento

1. Introduction

We provide some additional material in support of the main paper. The content is organised as follows:

- Sec. 2 provides a description of the proposed datasets for the evaluation of NCD methods for point cloud semantic segmentation;
- Sec. 3 reports the implementation details of EUMS†, our adaptation for 3D point cloud data of the method proposed by Zhao et al. [5] (originally designed for NCD in 2D image semantic segmentation);
- Sec. 4 shows a collection of additional qualitative results produced with NOPS on all the different splits of SemanticPOSS-n^i and SemanticKITTI-n^i.

2. Dataset splits for 3D NCD

To evaluate the performances of NOPS, we divide SemanticKITTI [1] and SemanticPOSS [2] into four different splits. We name these splits as SemanticKITTI-n^i and SemanticPOSS-n^i, respectively, where n is the number of novel classes contained in each split and i indexes the split. In each set, the novel classes and the base classes correspond to unlabelled and labelled points. These splits are selected based on two principles, i.e. balancing the distribution of the novel classes in each split, and including semantic relationships between base and novel classes within the same split. See details about the splits in Fig. 1. The first principle allows us to avoid the case in which the most frequent novel class affects the other classes, thus in turn affecting the learning of the unsupervised points. The second principle encourages the model to exploit the supervised knowledge over some base classes to discover the unsupervised novel classes, as in the case of the novel class *rider* in SemanticPOSS-3^3, whose discovery can be facilitated by the presence of the class *bike*, that is considered as base class in this specific split.

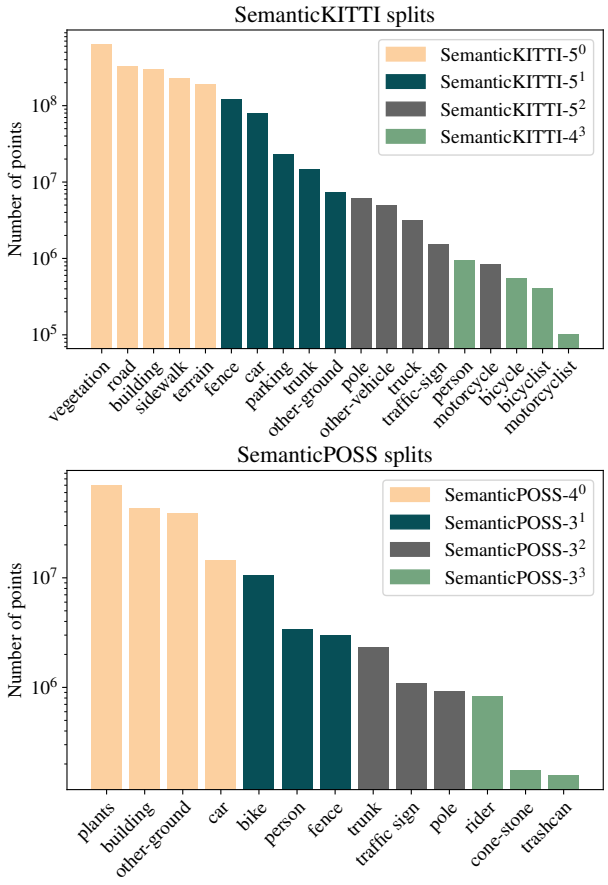

Figure 1. Histograms representing the number of points belonging to each class in SemanticKITTI [1] and SemanticPOSS [2]. Each class has been assigned the colour of the split in which it has to be considered novel (unlabelled).

3. Adapting NCD for 2D images to 3D

One of our contribution is the adaptation of EUMS [5], proposed for NCD in 2D image semantic segmentation, to 3D data. As some of the EUMS assumptions for the 2D case do not hold in the 3D point cloud domain, we introduce some changes in the proposed baseline. We name this adapted version as EUMS†.

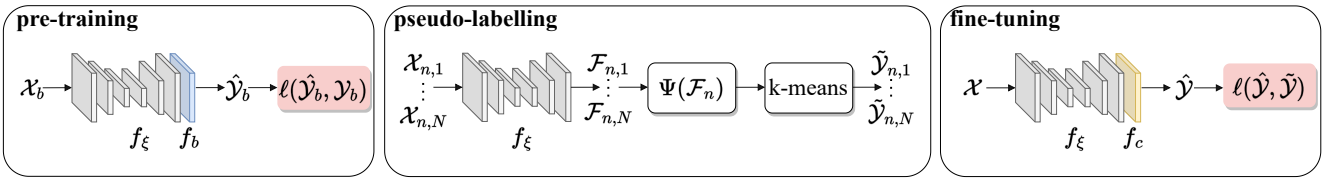

Figure 2. Overview of EUMS†, our adaptation of the method proposed by Zhao et al. [5]. We first pre-train f_ξ and f_b considering only the base points in each point cloud. Using f_ξ, we extract the features of the novel points in each scene, that are filtered with the selection function $\Psi(\cdot)$. Then, we produce the pseudo-labels for the selected novel points by using the k-means algorithm. Lastly, we plug a new segmentation head f_c into f_ξ and fine-tune the complete model on both novel and base points, considering pseudo-labels and ground-truth labels respectively.

As in the original implementation, EUMS† consists of three consecutive steps: i) pre-training, ii) pseudo-labelling and iii) fine-tuning. The block diagram of EUMS† is illustrated in Fig. 2. We first pre-train our model $f_\xi \circ f_b$ on the base classes only, where f_ξ is the feature extractor, f_b is the segmentation head for the base classes and $\circ$ is the composition operator. Then, we generate the pseudo-labels considering the features extracted by f_ξ and filtered with the selection function $\Psi(\cdot)$ working on the whole dataset, where Ψ is a random selection function. Lastly, we fine-tune the architecture $f_\xi \circ f_c$ jointly on the labelled base points and on the pseudo-labelled novel points, where f_c is the segmentation head for both base and novel classes. Here below each step is described in detail.

Pre-training. EUMS assumes that the novel classes belong to the foreground. Then, the novel classes are merged with the background class (considered as base in all the dataset splits) during the pre-training phase. The foreground pixels are obtained by an auxiliary saliency detection model [3]. The background pixels are just the output of the pre-trained model. The portion of the image belonging to both the foreground and the background masks contains the novel pixels. Because in point clouds there is no concept of foreground/background and saliency for 3D data cannot be leveraged as easily as for 2D data [4], we consider the novel points as the unlabelled points and we discard them during the pre-training phase. Therefore, the pre-training stage of EUMS† considers only the base points in each scene $\mathcal{X}_b$ and optimises $f_\xi \circ f_b$ by considering the objective function $\ell(\hat{\mathcal{Y}}_b, \mathcal{Y}_b)$, where $\hat{\mathcal{Y}}_b$ are the network predictions $\hat{\mathcal{Y}}_b = (f_\xi \circ f_b)(\mathcal{X}_b)$ and $\mathcal{Y}_b$ are the ground-truth annotations for the base points.

Pseudo-labelling. EUMS assumes that each image contains at most one novel class, allowing to compute a unique pseudo-label for each image. Authors in [5] propose to first average pool the features of the novel pixels in each image and then collect the image-level representations for the whole dataset. Finally, the hard pseudo-labels for all the novel points in each image are obtained by propagating the clustering affiliation of each image-level feature vector, determined by using the k-means algorithm.

In semantic segmentation for 3D point clouds, multiple novel classes usually occur in the same scene. Therefore, in EUMS† we propose to extract the per-point features $\mathcal{F}_{n,i}$ with f_ξ for all the novel points $\mathcal{X}_{n,i}$ contained in the i-th scene of the dataset. However, a large amount of novel points is difficult to handle due to hardware constrains. We randomly select a subset of point-level vectors using Ψ from each scene by setting a ratio (i.e. 30%) with an upper bound (i.e. 1K) on the number of points to select. Finally, we apply k-means clustering on the set of features collected over the whole dataset and obtain the pseudo-labels $\tilde{\mathcal{Y}}_{n,i}$ for the selected novel points in $\mathcal{X}_{n,i}$. To further enrich the pseudo-labels, we propagate the pseudo-label of each novel point to its nearest neighbour in the coordinate space. This allows us to increase the number of pseudo-labelled novel points.

Fine-tuning. During the last step of the EUMS†, we fine-tune the complete model following the same strategy used in [5]: given a point cloud $\mathcal{X}$, we compute the class predictions $\hat{\mathcal{Y}}$ as $\hat{\mathcal{Y}} = (f_\xi \circ f_c)(\mathcal{X})$ and we optimise the network considering the loss $\ell(\hat{\mathcal{Y}}, \tilde{\mathcal{Y}})$, where $\tilde{\mathcal{Y}} = \mathcal{Y}_b \cup \tilde{\mathcal{Y}}_n$.

4. Qualitative results

In this section, we report additional qualitative results by comparing NOPS with EUMS† [5] predictions.

Figs. 3-6, show qualitative results for SemanticPOSS from the split POSS-4^0 (Fig. 3), POSS-3^1 (Fig. 4), POSS-3^2 (Fig. 5) and POSS-3^3 (Fig. 6).

Figs. 7-10, show qualitative results for SemanticKITTI from the split KITTI-5^0 (Fig. 7), KITTI-5^1 (Fig. 8), KITTI-5^2 (Fig. 9) and KITTI-4^3 (Fig. 10). We add ground-truth labels as the supervised reference.

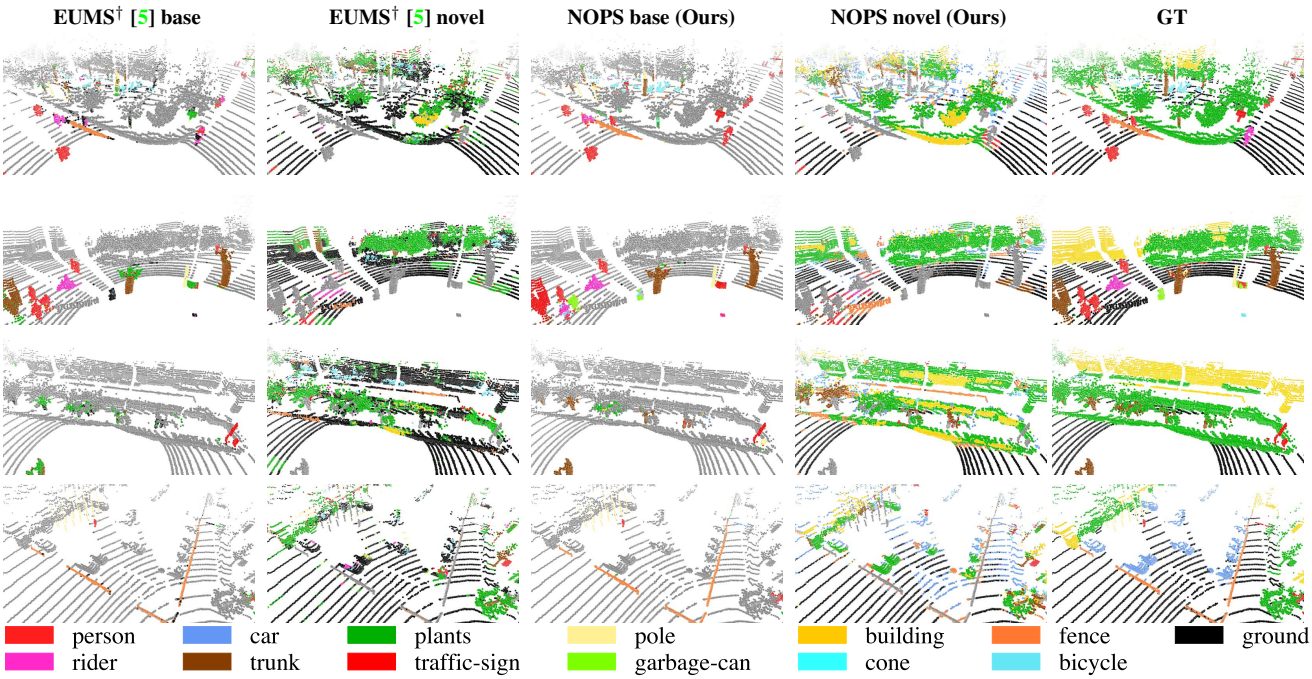

Figure 3. Qualitative comparison on SemanticPOSS from POSS-4^0. EUMS† [5] predicts wrong and cluttered predictions on the novel classes. NOPS provides improved predictions by assigning the correct classes to the majority of the points and only a minority are misclassified.

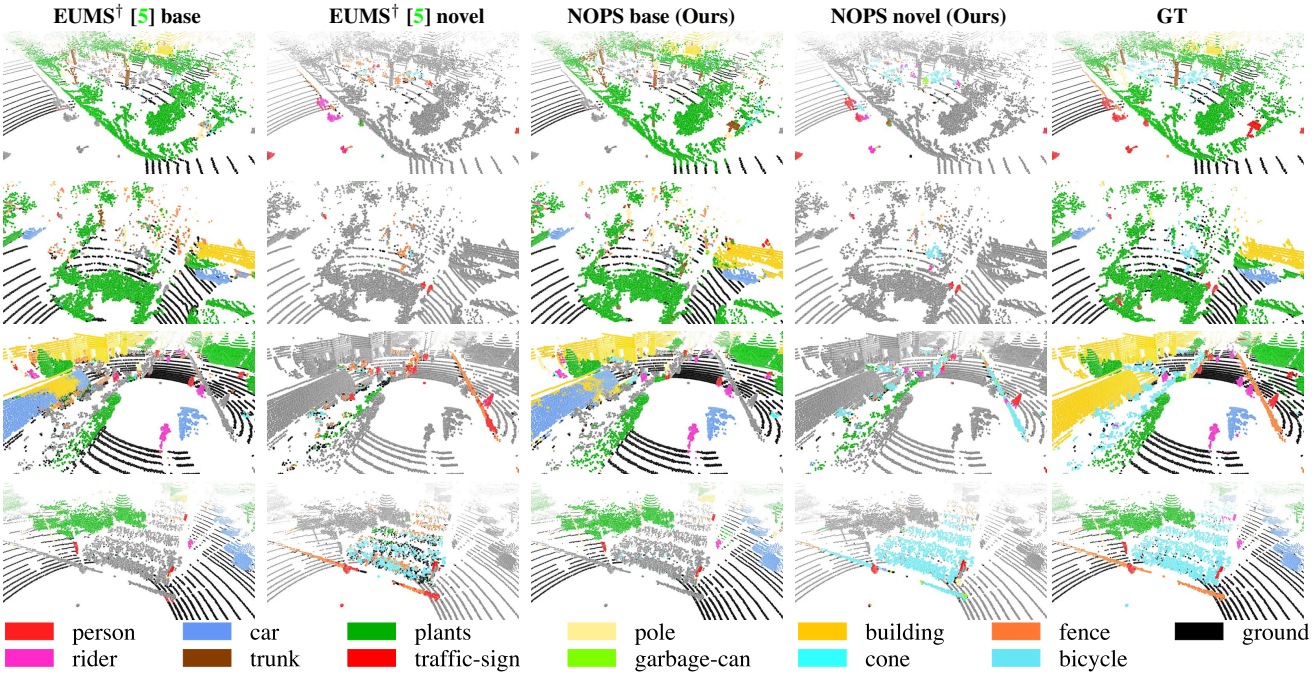

Figure 4. Qualitative comparison on SemanticPOSS from POSS-3^1. EUMS† [5] predicts wrong and cluttered predictions on the novel classes. NOPS provides improved predictions by assigning the correct classes to the majority of the points and only a minority are misclassified.

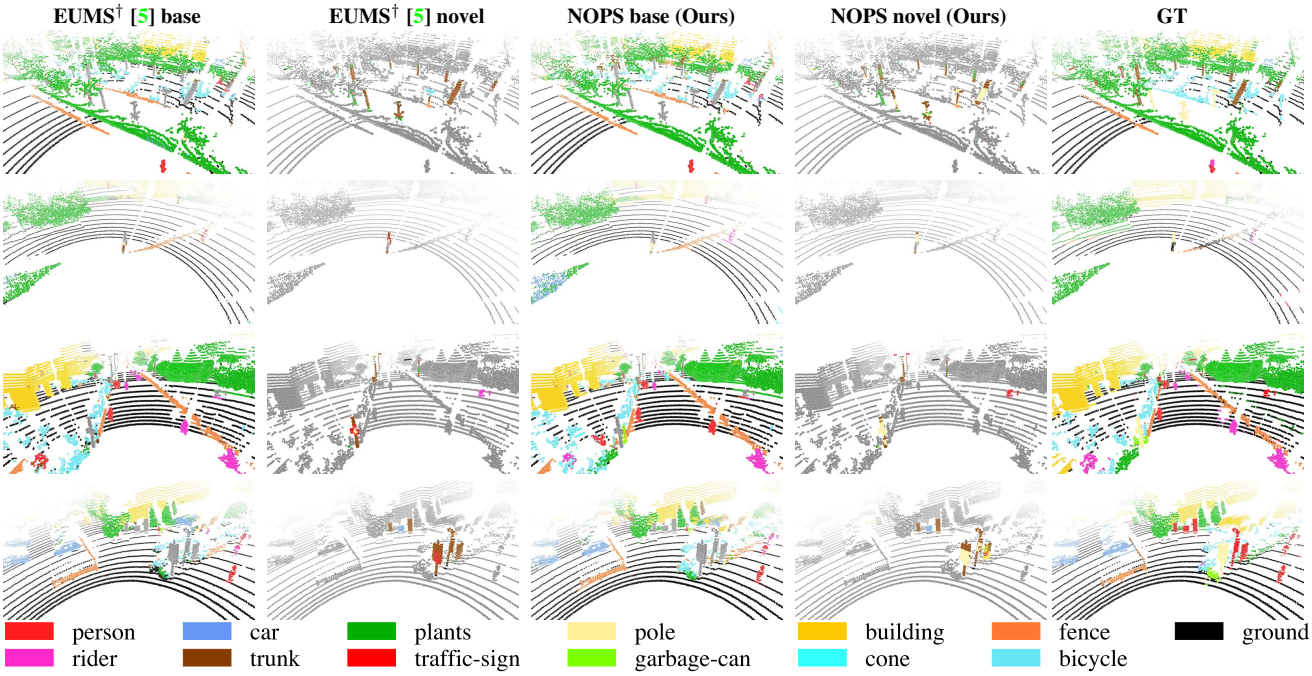

Figure 5. Qualitative comparison on SemanticPOSS from POSS-3^2. EUMS† [5] predicts wrong and cluttered predictions on the novel classes. NOPS provides improved predictions by assigning the correct classes to the majority of the points and only a minority are misclassified.

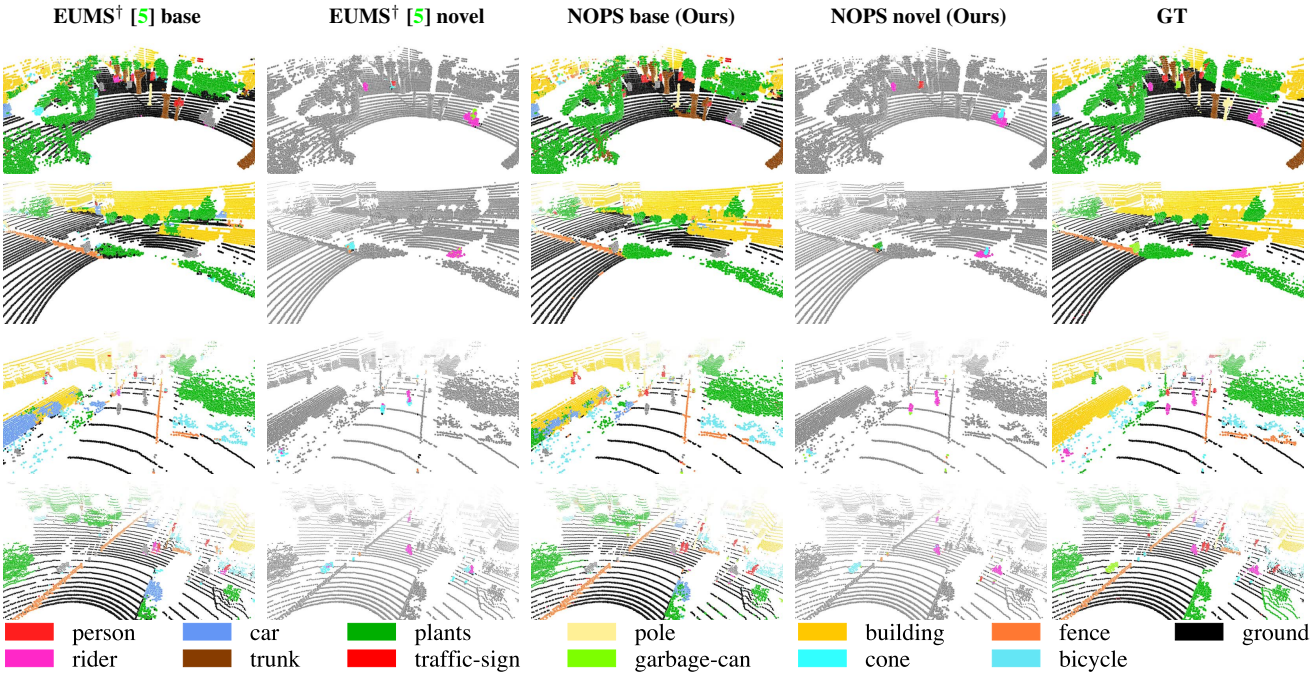

Figure 6. Qualitative comparison on SemanticPOSS from POSS-3^3. EUMS† [5] predicts wrong and cluttered predictions on the novel classes. NOPS provides improved predictions by assigning the correct classes to the majority of the points and only a minority are misclassified.

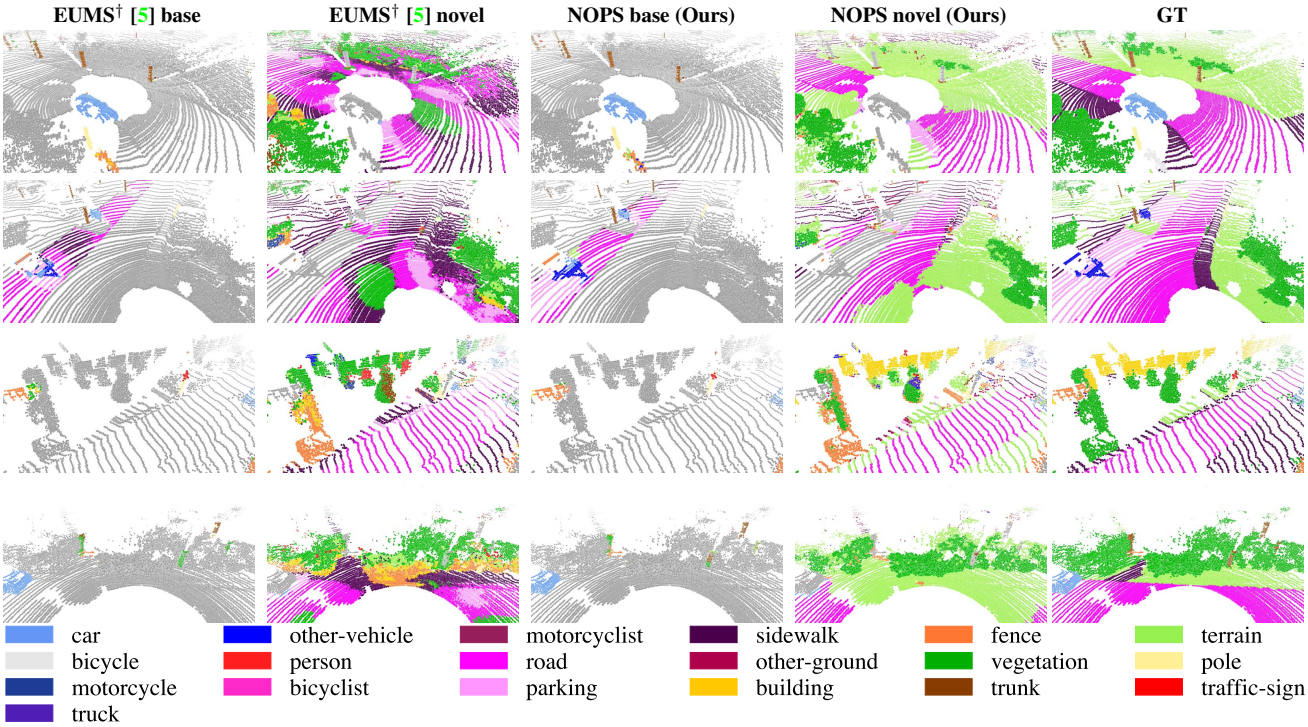

Figure 7. Qualitative comparison on SemanticKITTI from KITTI-5^0. EUMS† [5] outputs are completely or partially wrong for the novel classes. NOPS improves the performance by providing correct and more homogeneous predictions.

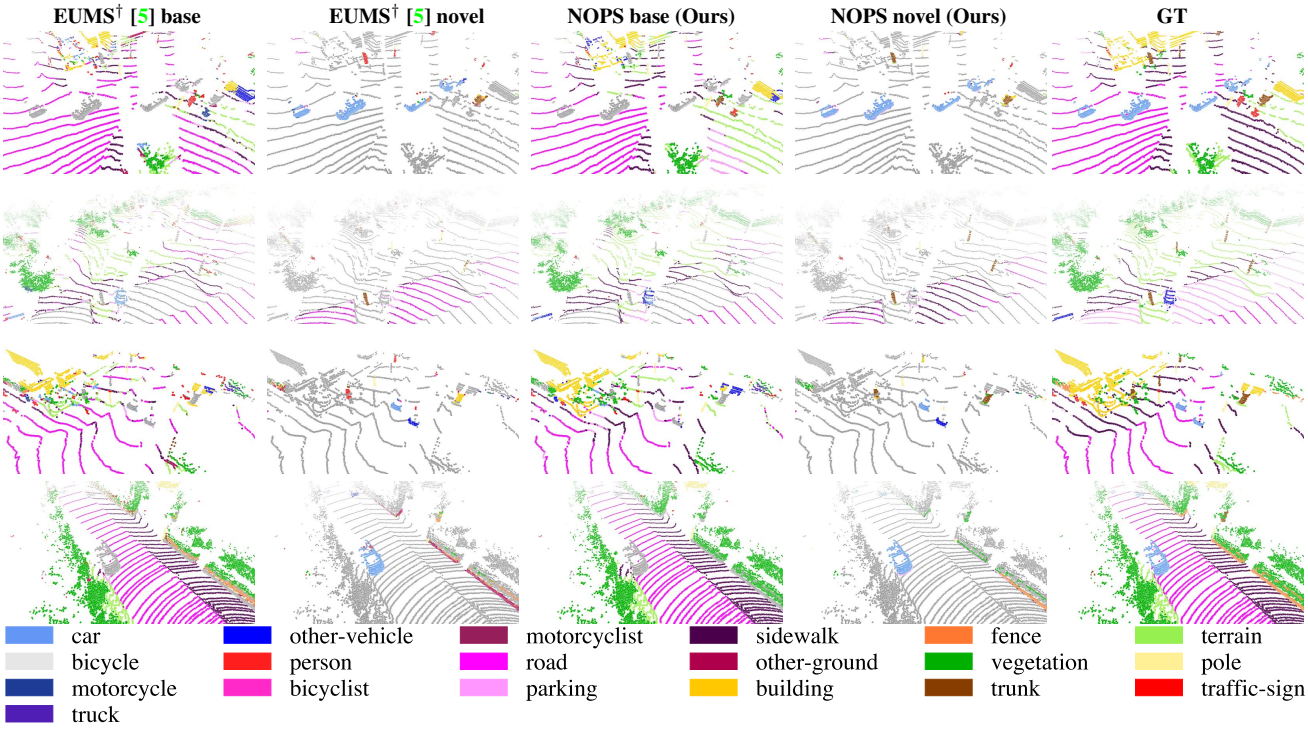

Figure 8. Qualitative comparison on SemanticKITTI from KITTI-5^1. EUMS† [5] outputs are completely or partially wrong for the novel classes. NOPS improves the performance by providing correct and more homogeneous predictions.

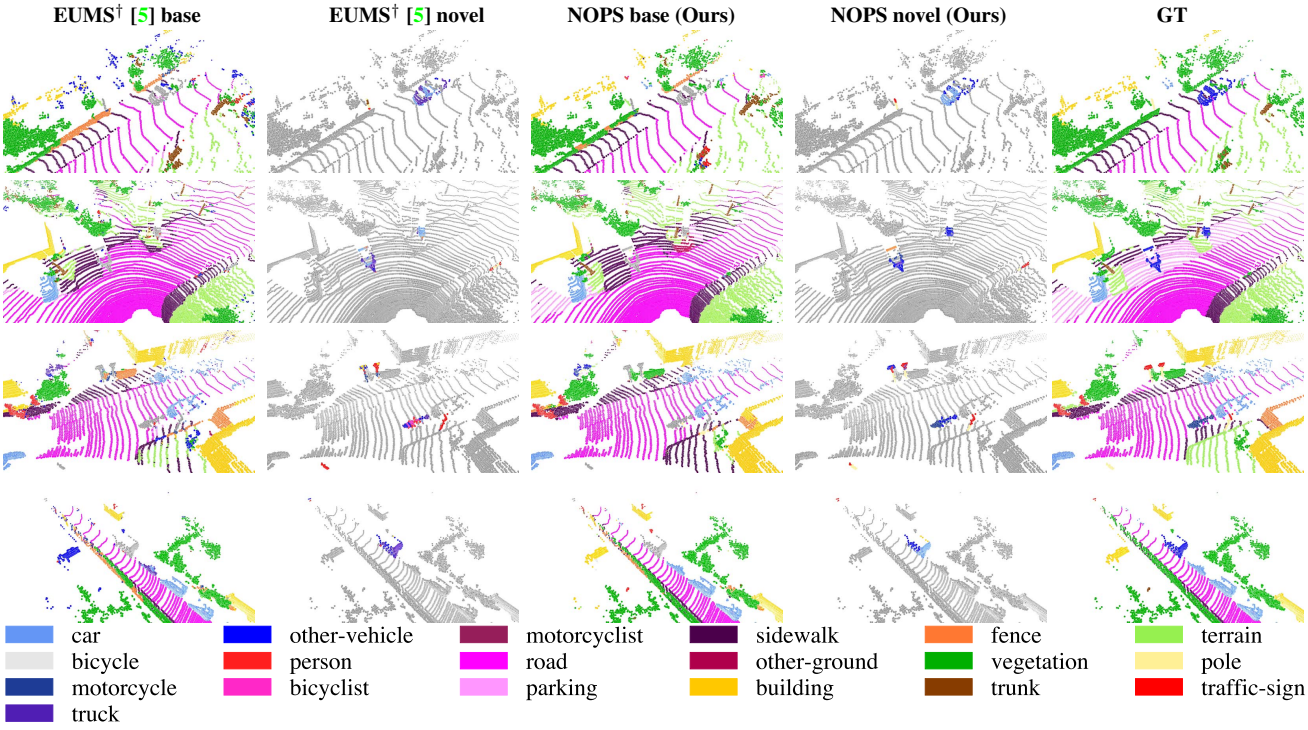

Figure 9. Qualitative comparison on SemanticKITTI from KITTI-5^2. EUMS† [5] outputs are completely or partially wrong for the novel classes. NOPS improves the performance by providing correct and more homogeneous predictions.

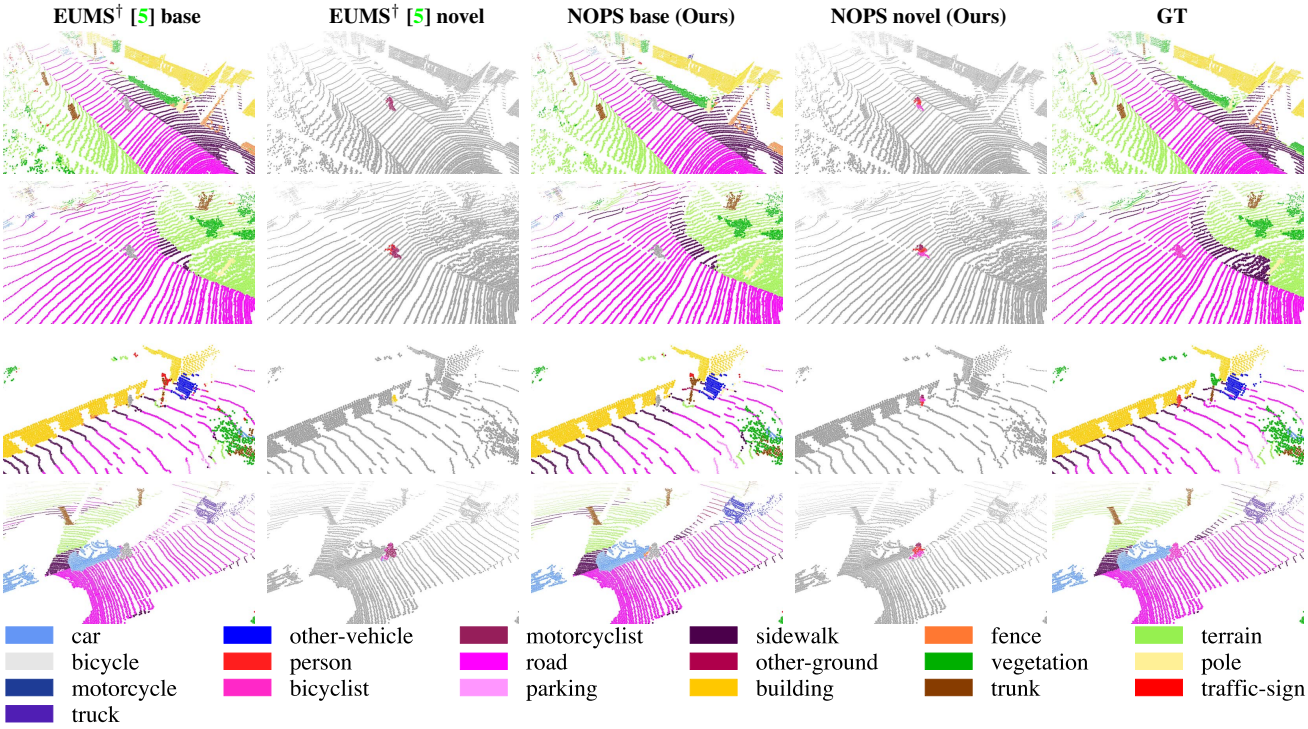

Figure 10. Qualitative comparison on SemanticKITTI from KITTI-4^3. EUMS† [5] outputs are completely or partially wrong for the novel classes. NOPS improves the performance by providing correct and more homogeneous predictions.

References

[1] J. Behley, M. Garbade, A. Milioto, J. Quenzel, S. Behnke, C. Stachniss, and J. Gall. SemanticKITTI: A dataset for semantic scene understanding of LiDAR sequences. In *CVPR*, 2019. 1

[2] Y. Pan, B. Gao, J. Mei, S. Geng, C. Li, and H. Zhao. SemanticPOSS: A point cloud dataset with large quantity of dynamic instances. In *IV*, 2020. 1

[3] X. Qin, Z. Zhang, C. Huang, C. Gao, M. Dehghan, and M. Jagersand. Basnet: Boundary-aware salient object detection. In *CVPR*, 2019. 2

[4] R. Song, W. Zhang, Y. Zhao, Y. Liu, and P.L. Rosin. Mesh saliency: An independent perceptual measure or a derivative of image saliency? In *CVPR*, 2021. 2

[5] Y. Zhao, Z. Zhong, N. Sebe, and G.H. Lee. Novel class discovery in semantic segmentation. In *CVPR*, 2022. 1, 2, 3, 4, 5, 6