Supplementary Material

A. Dataset Statistics

Dataset	Split	Train (km^2)	Val (km^2)	Overlap (km^2) / Ratio
Argoverse2	Original	3.46	1.04	0.56 / 54%
	New	3.58	0.54	**0.00 / 0%**
NuScenes	Original	2.00	0.92	0.79 / 85%
	New	1.64	0.54	**0.06 / 11%**

Table 1. Summary of the cumulative areas of the training set, validation set, and their overlap in both datasets. The original splits of both datasets exhibit a high overlap ratio, a problem substantially mitigated by our suggested re-division.

In this supplementary material, we provide an in-depth analysis of the NuScenes [1] and Argoverse2 [3] datasets to further emphasize the necessity of re-splitting both datasets for the task of map construction.

Table 1 shows the cumulative areas of the training and validation sets and their overlap across different splits of both datasets. We query all locations within a $30, m$ radius circle from the ego-vehicle's position.

In Argoverse2, the initial overlap ratio is 54%. By merging the training, validation, and testing sets together, we redivide them into a new train/validation split. We propose that an additional testing set is unnecessary, as all map data is publicly available. A fair testing set cannot be constructed if the ground-truth data is not hidden. Consequently, we divide the entire set of $1,000$ scenes into a $700/150$ train/validation split, ensuring a balanced distribution over objects, weathers, and cities, while completely eliminating any overlaps.

In NuScenes, more than 85% of locations in the validation set appear in the training set in the original split, which raises serious concerns about overfitting. However, the overlap ratio is reduced to 11% after using Roddick and Cipolla's new split [2], substantially mitigating this issue.

B. Dataset Visualization

To facilitate a more intuitive understanding, we visualize each city/district in both datasets, comparing the train/validation splits before and after the re-division. Figure 1, 2, 3, 4, 5, and 6 illustrate this comparison for Argoverse2. The *green* areas represent the training set, *blue* signifies the validation set, and *red* indicates the overlaps. We maintain a balance between training and validation data across all six cities and ensure no overlaps. Figure 7, 8, 9, and 10 showcase the comparison for NuScenes. The original split results in a high degree of overlap, which the new split significantly alleviates.

Argoverse2 exhibits a significantly higher diversity of locations compared to NuScenes, which explains the performance gap observed in our experiments.

References

[1] Holger Caesar, Varun Bankiti, Alex H Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. nuscenes: A multimodal dataset for autonomous driving. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11621–11631, 2020. 1

[2] 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. 1

[3] Benjamin Wilson, William Qi, Tanmay Agarwal, John Lambert, Jagjeet Singh, Siddhesh Khandelwal, Bowen Pan, Ratnesh Kumar, Andrew Hartnett, Jhony Kaesemodel Pontes, Deva Ramanan, Peter Carr, and James Hays. Argoverse 2: Next generation datasets for self-driving perception and forecasting. In *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks (NeurIPS Datasets and Benchmarks 2021)*, 2021. 1

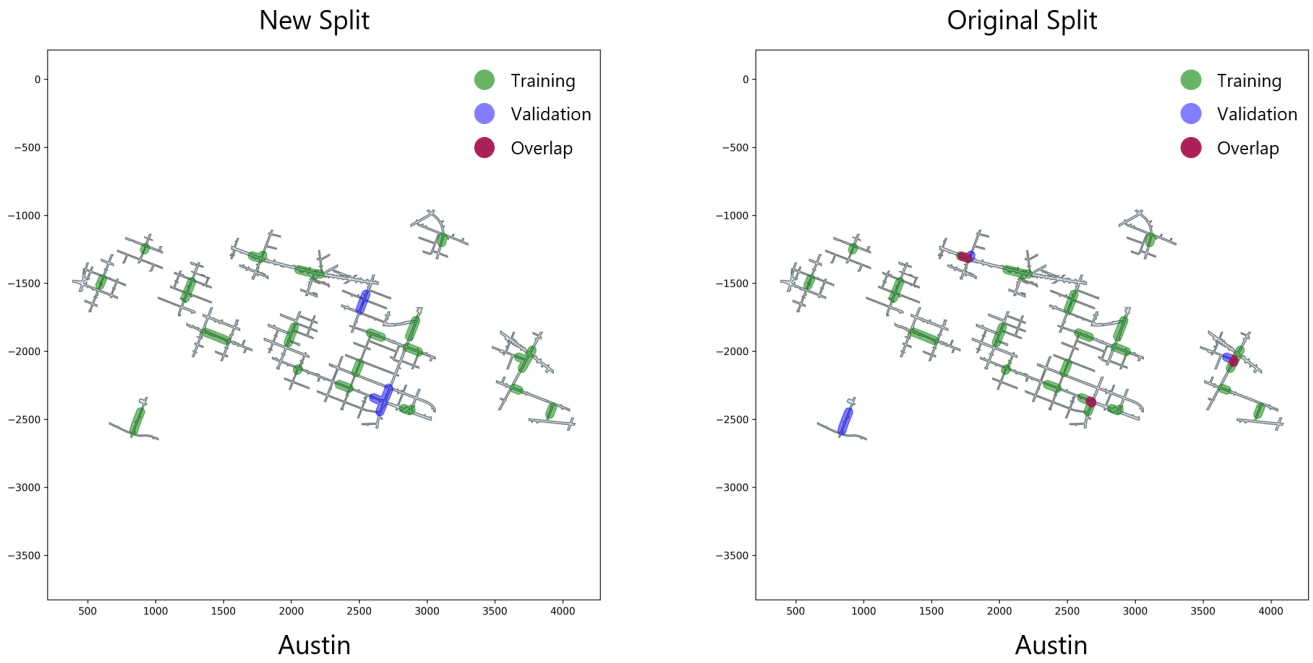

Figure 1. Comparison of the new and original splits in the Austin area of the Argoverse2 dataset. The *green* regions represent the training set, *blue* denotes the validation set, and *red* signifies the overlapping areas. Best viewed in color.

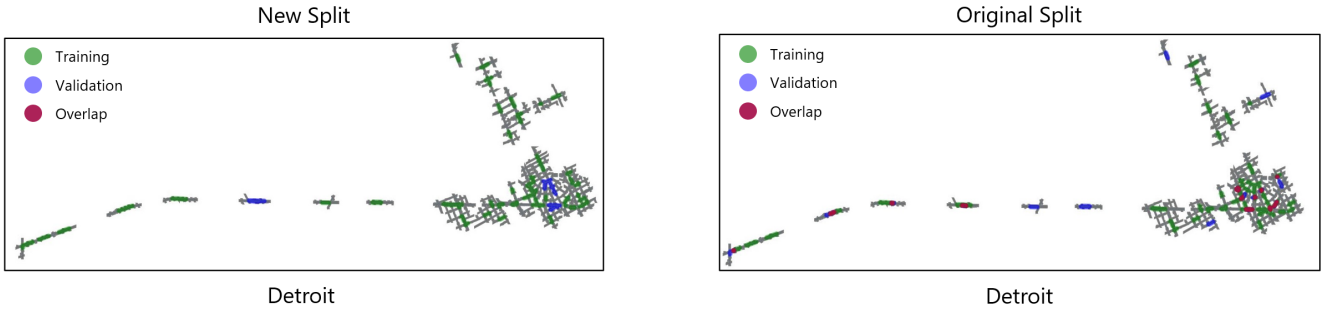

Figure 2. Comparison of the new and original splits in the Detroit area of the Argoverse2 dataset. The *green* regions represent the training set, *blue* denotes the validation set, and *red* signifies the overlapping areas. Best viewed in color.

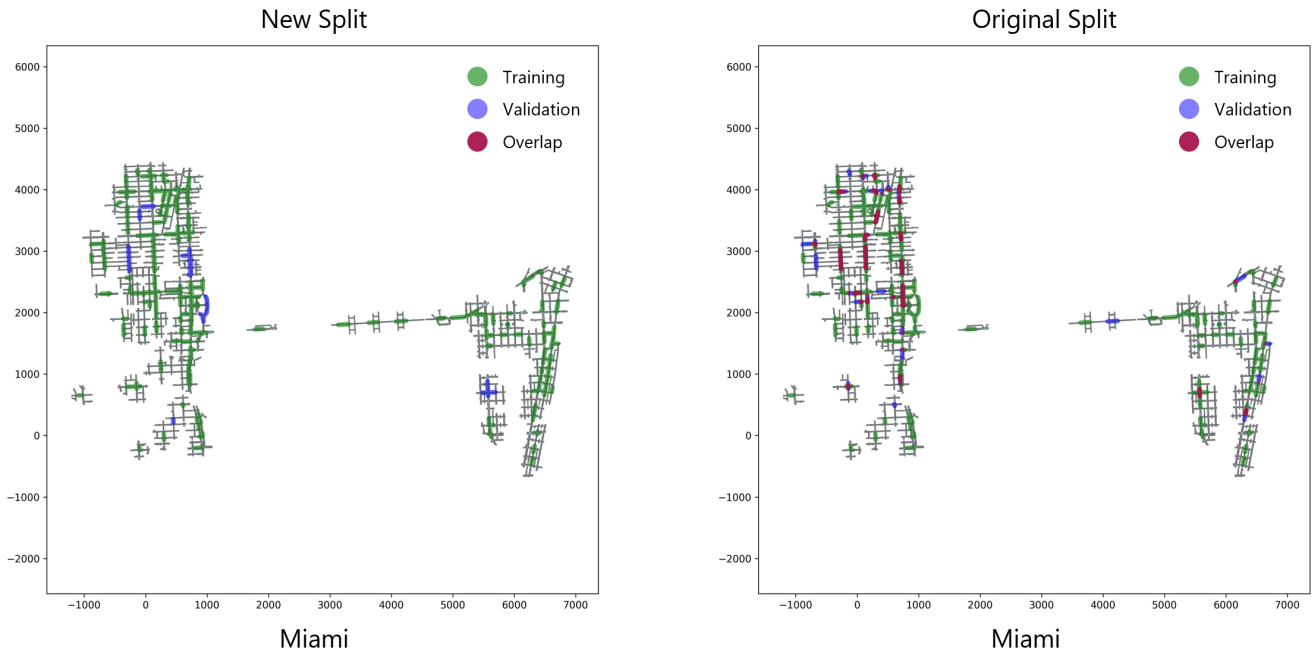

Figure 3. Comparison of the new and original splits in the Miami area of the Argoverse2 dataset. The *green* regions represent the training set, *blue* denotes the validation set, and *red* signifies the overlapping areas. Best viewed in color.

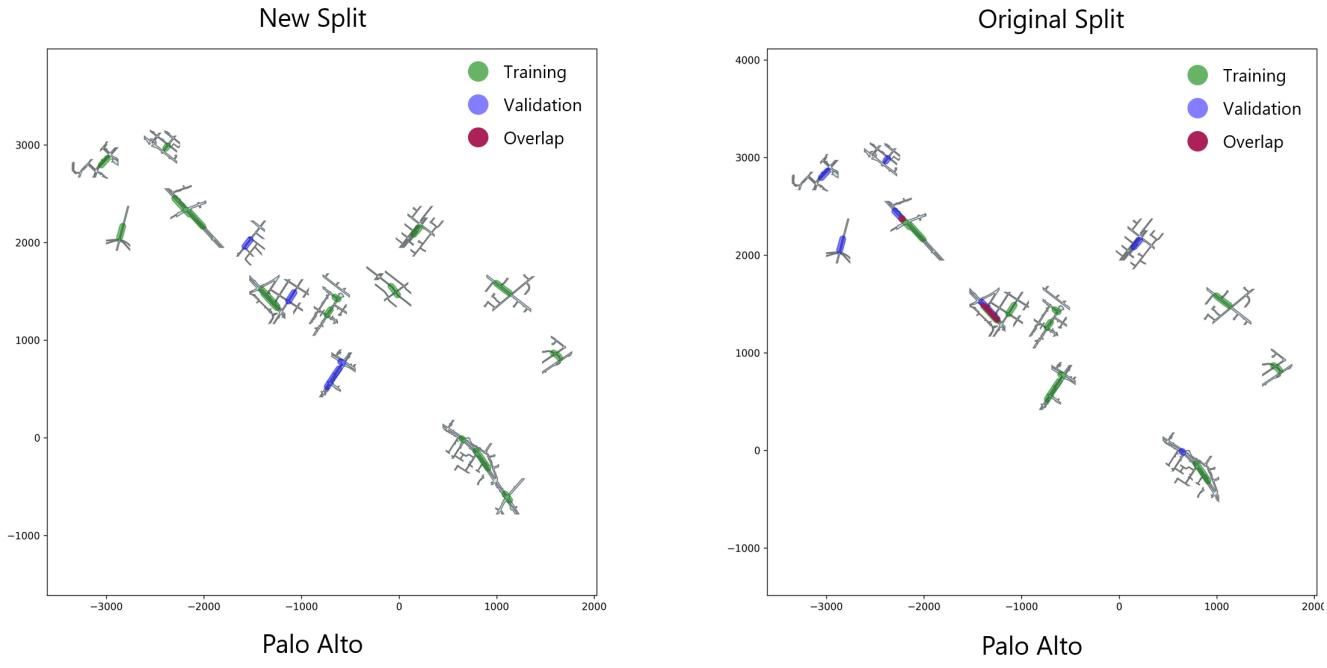

Figure 4. Comparison of the new and original splits in the Palo Alto area of the Argoverse2 dataset. The *green* regions represent the training set, *blue* denotes the validation set, and *red* signifies the overlapping areas. Best viewed in color.

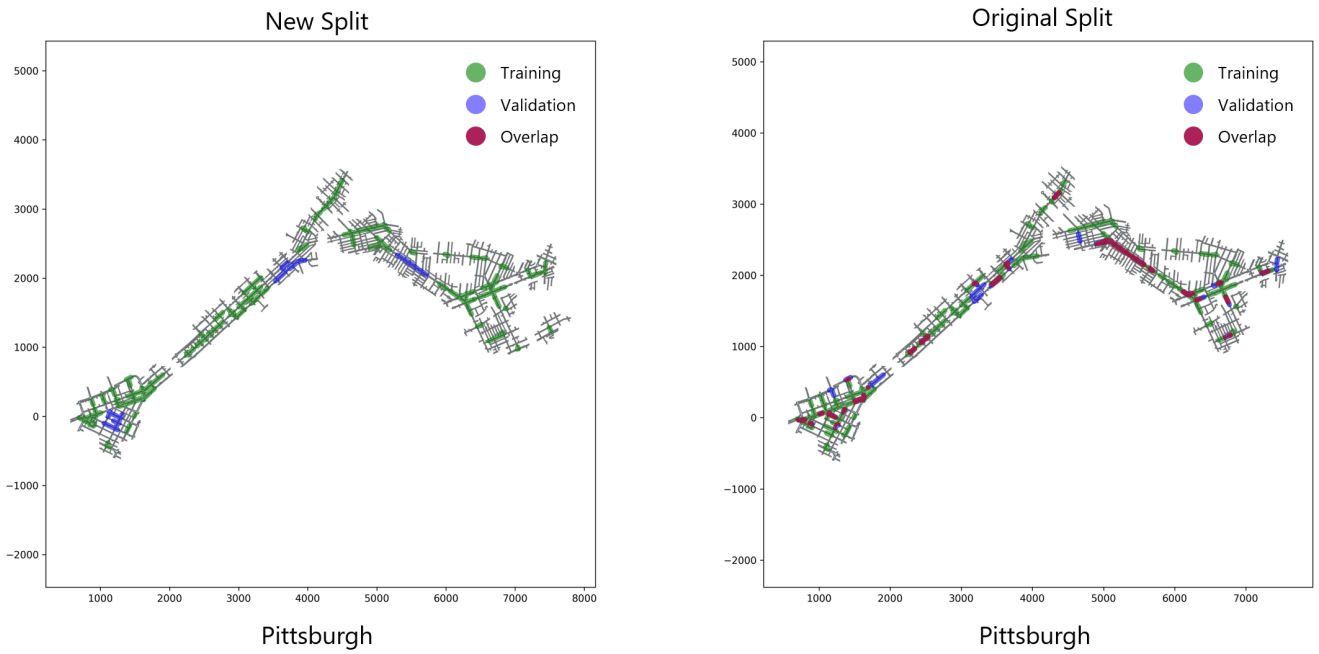

Figure 5. Comparison of the new and original splits in the Pittsburgh area of the Argoverse2 dataset. The *green* regions represent the training set, *blue* denotes the validation set, and *red* signifies the overlapping areas. Best viewed in color.

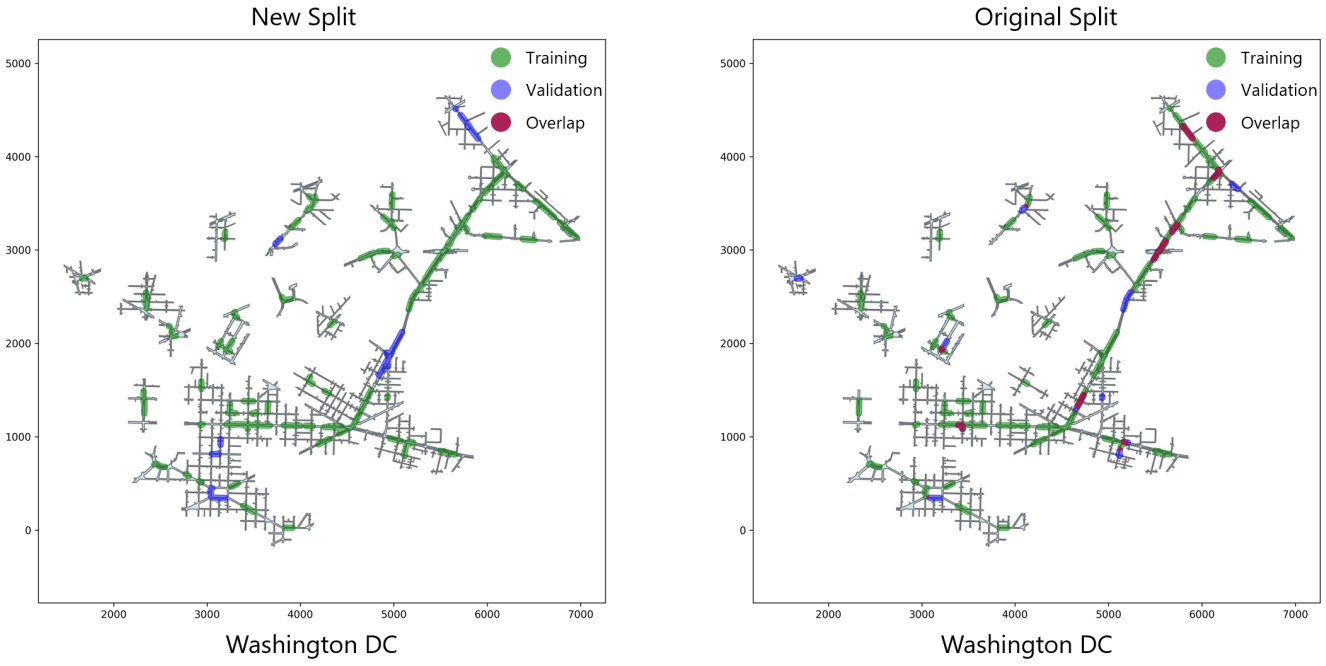

Figure 6. Comparison of the new and original splits in the Washington D.C. area of the Argoverse2 dataset. The *green* regions represent the training set, *blue* denotes the validation set, and *red* signifies the overlapping areas. Best viewed in color.

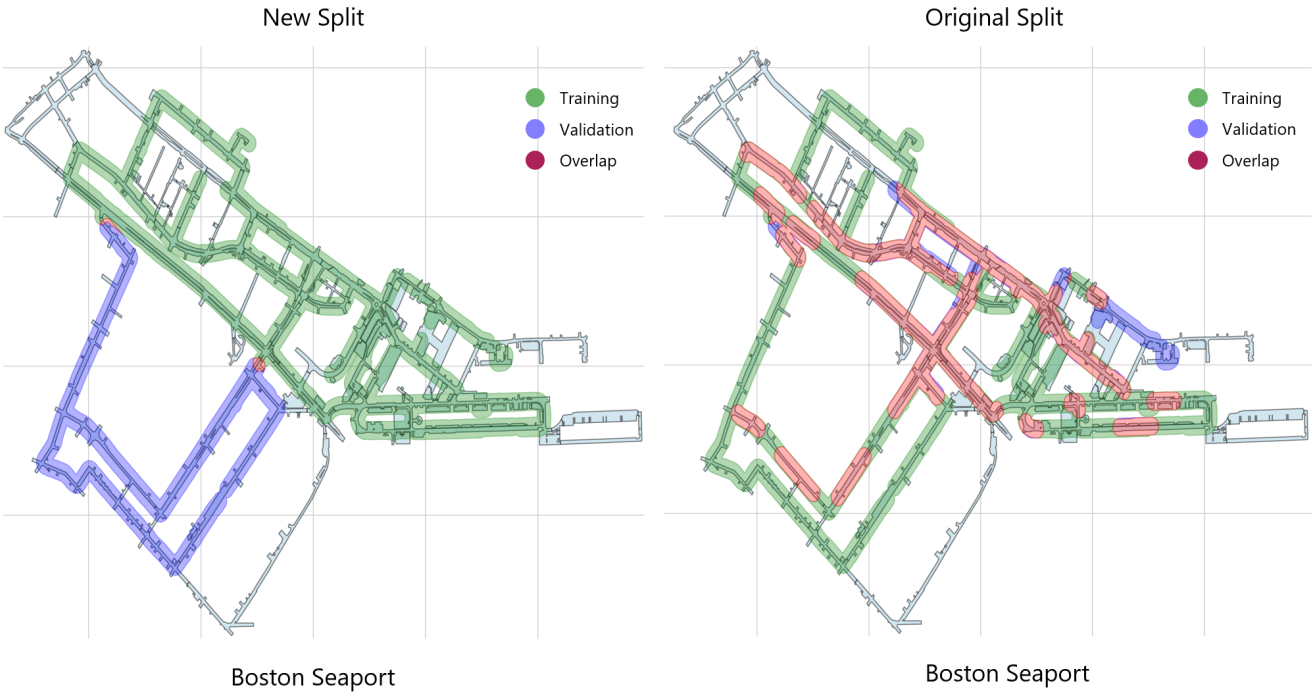

Figure 7. Comparison of the new and original splits in the Boston seaport area of the NuScenes dataset. The *green* regions represent the training set, *blue* denotes the validation set, and *red* signifies the overlapping areas. Best viewed in color.

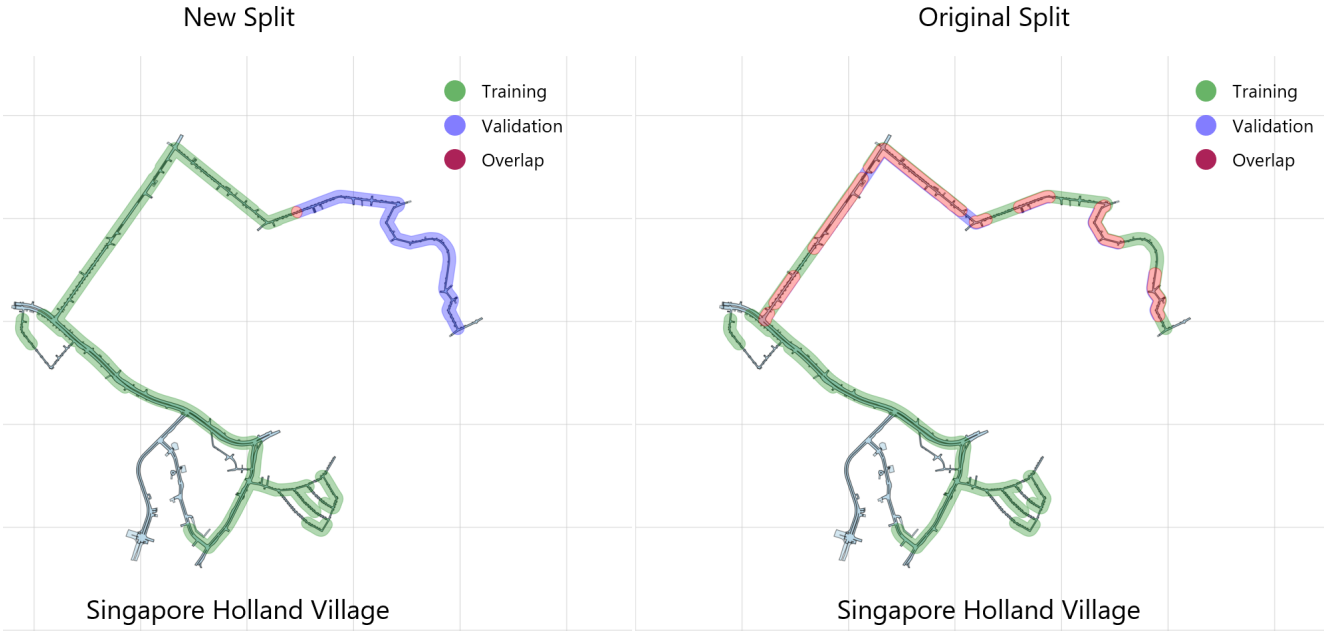

Figure 8. Comparison of the new and original splits in the Singapore Holland area of the NuScenes dataset. The *green* regions represent the training set, *blue* denotes the validation set, and *red* signifies the overlapping areas. Best viewed in color.

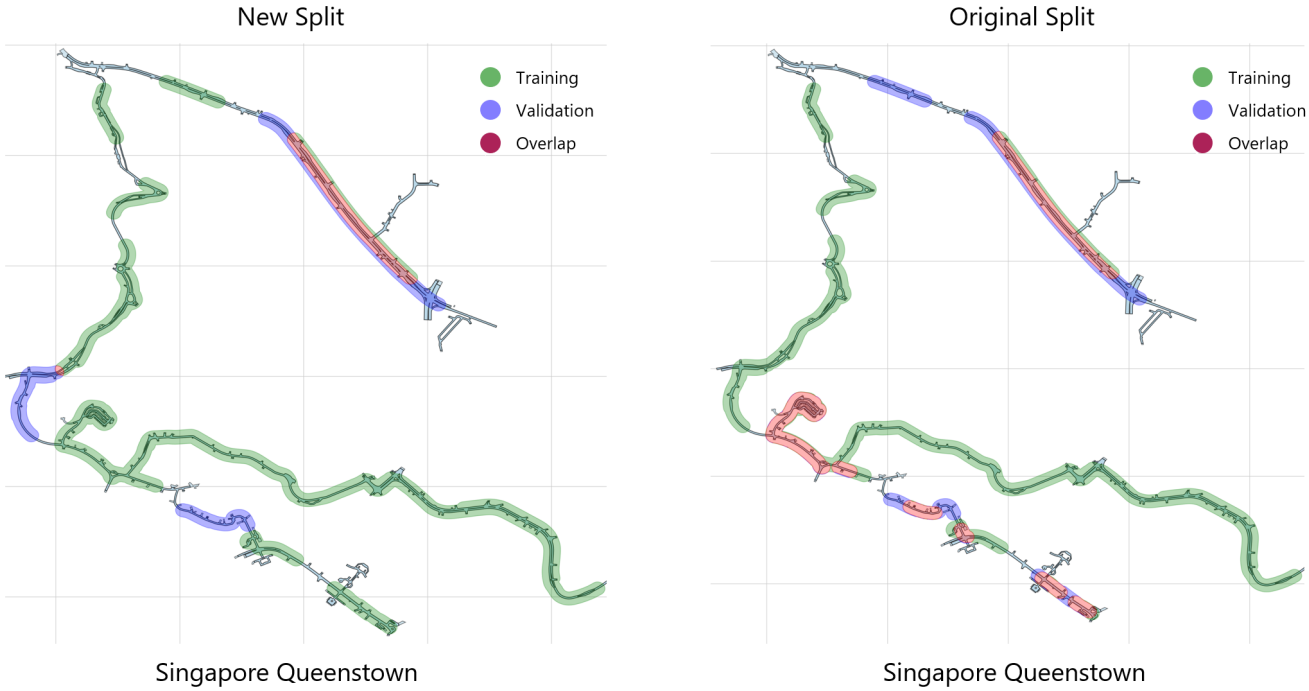

Figure 9. Comparison of the new and original splits in the Singapore Queenstown area of the NuScenes dataset. The *green* regions represent the training set, *blue* denotes the validation set, and *red* signifies the overlapping areas. Best viewed in color.

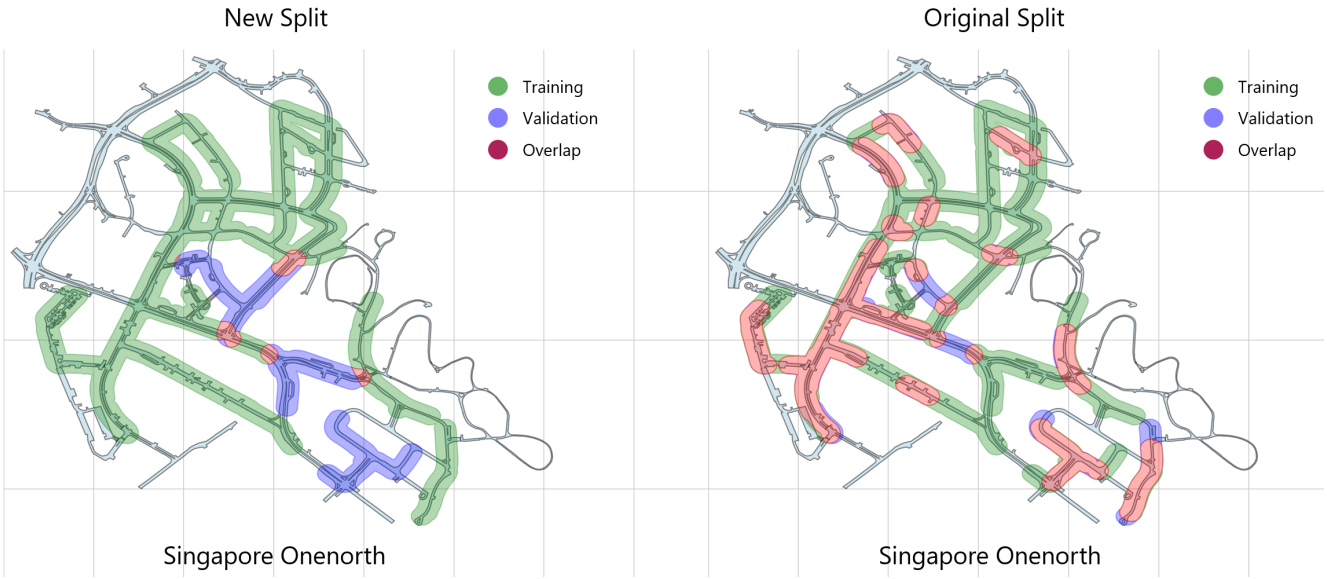

Figure 10. Comparison of the new and original splits in the Singapore One-North area of the NuScenes dataset. The *green* regions represent the training set, *blue* denotes the validation set, and *red* signifies the overlapping areas. Best viewed in color.