RecycleLoRA: Rank-Revealing QR-Based Dual-LoRA Subspace Adaptation for Domain Generalized Semantic Segmentation

Supplementary Material

This supplement provides additional materials omitted from the main text to facilitate a deeper understanding of our proposed RecycleLoRA.

Contents

A. Implementation Details

Our method is implemented based on the MMSegmentation codebase. We use DINOv2-Large as the backbone and Mask2Former as the decode head. Following the experimental setups of Rein [64] and SoMA [69], we only utilize the default data augmentation provided in Mask2Former [12] to ensure a fair comparison. All models are trained on NVIDIA A6000 GPUs. Further details on hyperparameters are provided in Table 8. Unless stated otherwise, all experiments in the main paper were conducted using these settings.

Hyperparameter	**Synthetic-to-Real**	**Real-to-Real**
backbone	DINOv2-L	DINOv2-L
main rank	32	32
sub rank	4	2
main lr mult.	1.0	1.0
sub lr mult.	0.5	0.5
learning rate	1e-4	1e-4
backbone lr mult.	0.5	0.5
lr scheduler	PolyLR	PolyLR
AWD scheduler	Cosine	Cosine
weight decay	0.05	0.05
optimizer	AdamW	AdamW
batch size	4	4
iterations	40,000	40,000

Table 8. Hyperparameter settings for experiments.

B. Additional Experiments and Analysis

B.1. Hyperparameter Analysis

Rank Analysis. To determine the optimal configuration for our dual-adapter structure, we conducted an analysis to investigate the impact of the rank settings for both the Main and Sub Adapters on domain generalization performance.

First, for the synthetic-to-real scenario (Table 9), we found that a Main Adapter rank of 32 consistently outperformed a rank of 16. With the Main Adapter's rank fixed at 32, we observed that performance peaked at an average mIoU of 68.95 when the Sub Adapter's rank was 4. However, increasing the rank further to 8 or higher led to a noticeable degradation in performance. This finding is consistent with our hypothesis that the Sub Adapter, which modifies the VFM's major directions, requires minimal and careful adjustments. Therefore, we adopted the (32, 4) rank configuration for the synthetic-to-real experiments (e.g., Table 3, 12).

We extended this analysis to the real-to-real generalization scenario (Table 10) as well. In this setting, the best performance (72.10 mIoU) was achieved with a Main Adapter rank of 32 and a Sub Adapter rank of 2. This result suggests that for the real-to-real setting, which has a smaller domain gap, an even more conservative adjustment of the VFM's major directions is beneficial. Consequently, we used the (32, 2) rank configuration for all real-to-real experiments (e.g., Table 4).

Synthetic-to-Real Generalization

rank		lr		Params.	*Trained on GTAV*			
main	sub	main	sub		→Citys.	→BDD	→Map.	Avg.
32	2	1e-4	5e-5	13.4M	72.03	60.75	72.06	68.28
32	4	1e-4	5e-5	14.2M	**73.01**	**61.77**	**72.07**	**68.95**
32	8	1e-4	5e-5	15.7M	72.83	61.13	70.81	68.26
32	16	1e-4	5e-5	18.9M	71.20	61.16	70.87	67.74
32	32	1e-4	5e-5	25.2M	72.13	60.83	69.87	67.61
16	2	1e-4	5e-5	7.1M	71.67	61.27	70.84	67.93
16	4	1e-4	5e-5	7.9M	71.74	61.54	71.25	68.18
16	8	1e-4	5e-5	9.4M	71.40	60.56	70.41	67.46
16	16	1e-4	5e-5	12.6M	71.51	60.48	70.37	67.45

Table 9. Domain generalization results (mIoU %) for RecycleLoRA with varying rank configurations for its Main and Sub Adapters, under the synthetic-to-real setting (G→{C, B, M}). **Bold** and underlined indicate best and second-best results.

Real-to-Real Generalization

rank		lr		Params.	*Trained on Cityscapes*		
main	sub	main	sub		→BDD	→Map.	Avg.
32	2	1e-4	5e-5	13.4M	66.65	**78.14**	**72.10**
32	4	1e-4	5e-5	14.2M	**66.76**	76.64	71.70

Table 10. Domain generalization results (mIoU %) for RecycleLoRA with varying rank configurations for its Main and Sub Adapters, under the real-to-real setting (C→{B, M}). **Bold** and underlined indicate best and second-best results.

Learning rate Analysis. We conduct an analysis to determine the optimal learning rate for the Sub Adapter and validate our design choice of using a different learning rate from the Main Adapter. As shown in Table 11, we fixed the Main Adapter's learning rate to 1e-4 and varied the Sub Adapter's learning rate. The results demonstrate that the best performance is achieved when the Sub Adapter's learning rate is set to 5e-5, half that of the Main Adapter, achieving an average mIoU of 68.95. Setting the learning rate for the Sub Adapter to be either the same as the Main Adapter (1e-4) or excessively low (1e-5) resulted in a performance drop. This empirical evidence supports our strategy of applying a carefully tuned, lower learning rate to the Sub Adapter. This differentiation is crucial for enabling the complementary learning process between the two adapters, validating our overall design.

Synthetic-to-Real Generalization								
rank		lr		Params.	*Trained on GTAV*			
main	sub	main	sub		→Citys.	→BDD	→Map.	Avg.
32	4	1e-4	1e-4	14.2M	72.23	60.50	70.60	67.78
32	4	1e-4	5e-5	14.2M	**73.01**	**61.77**	**72.07**	**68.95**
32	4	1e-4	1e-5	14.2M	72.55	61.10	70.55	68.07

Table 11. Domain generalization results (mIoU %) for RecycleLoRA with varying learning rate configurations for its Main and Sub Adapters, under the synthetic-to-real setting (G→{C, B, M}). **Bold** and underlined indicate best and second-best results.

B.2. Ablation on Dual-Adapter Initialization

To assess the impact of the initialization strategy on our dual-adapter framework, we conducted an ablation study comparing our RRQR-based approach with other representative initialization methods. We compare against Kaiming uniform initialization, a standard method that does not leverage the pre-trained weight structure, and SVD-based initialization, which utilizes subspace decomposition as seen in prior work such as SoMA [69] or PiSSA [48]. Interestingly, as presented in Table 12, the other initialization methods did not synergize with the dual-adapter structure and instead exhibited performance degradation. Specifically, the dual-adapter with Kaiming initialization scored 66.31 mIoU, which is lower than the 66.89 mIoU of standard LoRA [26] (single adapter) presented in Table 3. Similarly, the SVD-based initialization achieved only 67.23 mIoU, underperforming SoMA (single adapter), which scored 68.27 mIoU as shown in Table 3. In contrast, our proposed RRQR-based initialization achieves an average mIoU of 68.95, significantly outperforming both alternatives. These results underscore the importance of the initialization method and suggest that our proposed RRQR-based strategy is a more effective choice for the proposed dual-adapter structure.

Synthetic-to-Real Generalization					
Initialization	Backbone	*Trained on GTAV*			
		→Citys.	→BDD	→Map.	Avg.
Kaiming unif.	DINOv2-L	68.96	60.57	69.41	66.31
SVD	DINOv2-L	70.40	60.35	70.93	67.23
RRQR	DINOv2-L	**73.01**	**61.77**	**72.07**	**68.95**

Table 12. Domain generalization results (mIoU %) for the dual-adapter framework with different initialization strategies, under the synthetic-to-real setting (G→{C, B, M}). **Bold** and underlined indicate best and second-best results.

C. Limitations and Future Works

While RecycleLoRA demonstrates robust state-of-the-art performance, we identify several avenues for future research that could further advance its capabilities and address its current limitations.

Further optimization of RecycleLoRA could be achieved by developing systematic methods for tuning hyperparameters, such as the ranks and learning rates of the dual adapters. The current configuration was determined empirically, and creating automated search strategies would enhance the practicality and replicability of our approach.

The scope of our work, currently focused on Domain Generalized Semantic Segmentation, could also be broadened. The core principle of recycling pre-trained knowledge through subspace analysis is likely applicable to other downstream tasks that require efficient foundation model fine-tuning, such as object detection, video analysis, and medical image segmentation. Investigating the effectiveness of our approach in these diverse contexts presents a logical direction for future work.

Another promising research direction involves revisiting our methodology's binary partitioning of the VFM's subspace into "major" and "minor" directions, which currently omits the intermediate directions. This simplification is based on the hypothesis that the most and least dominant directions are most critical for balancing knowledge preservation and new feature acquisition. However, the potential contribution of these intermediate directions remains unexplored. Future research could investigate a more nuanced allocation of the entire subspace spectrum, perhaps through a third adapter or a soft-weighting scheme that utilizes all ranked directions, which may unlock further performance gains.

D. Qualitative Results

Figures 6 and 7 present qualitative comparisons against other state-of-the-art methods. These visualizations highlight that RecycleLoRA generates more accurate and detailed segmentation maps, which are more closely aligned with the ground truth.

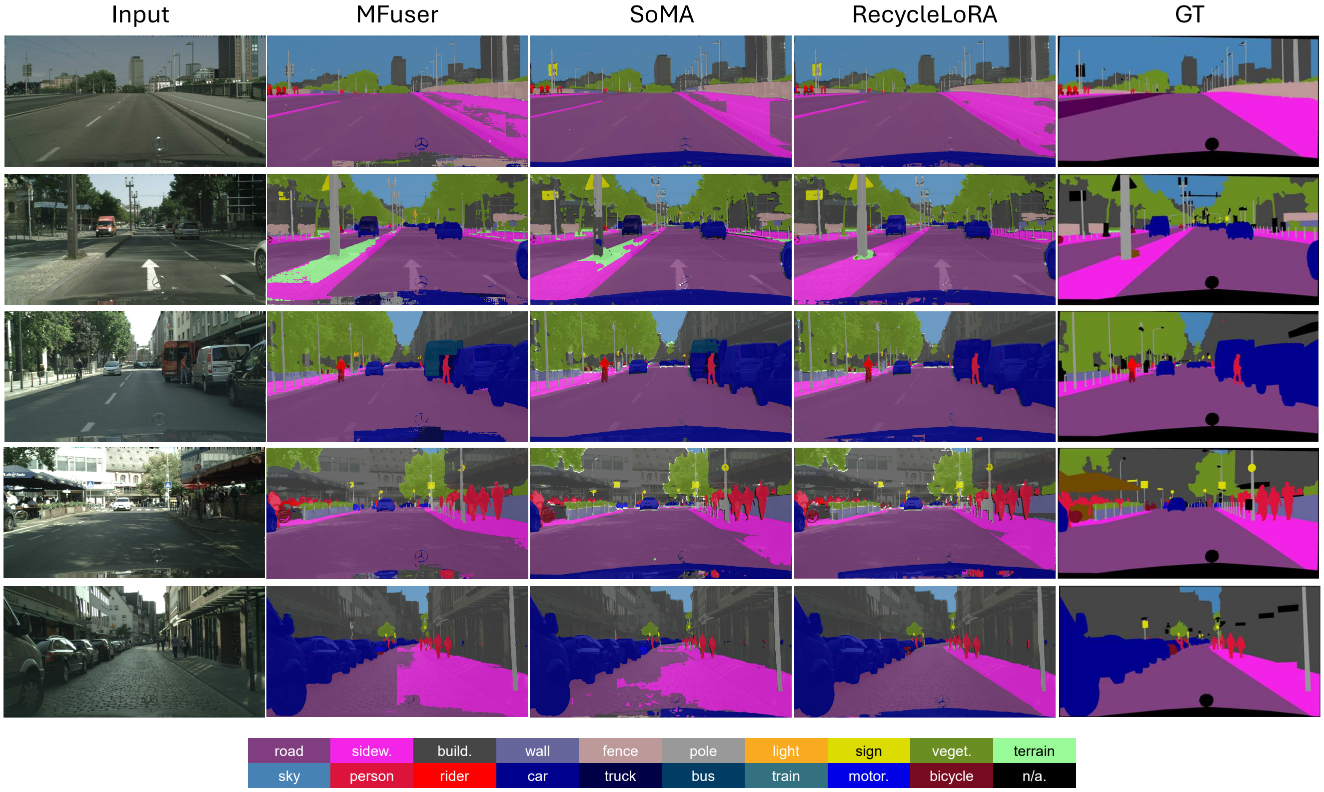

Figure 6. Qualitative comparison of semantic segmentation on the Cityscapes. All models were trained on the GTAV.

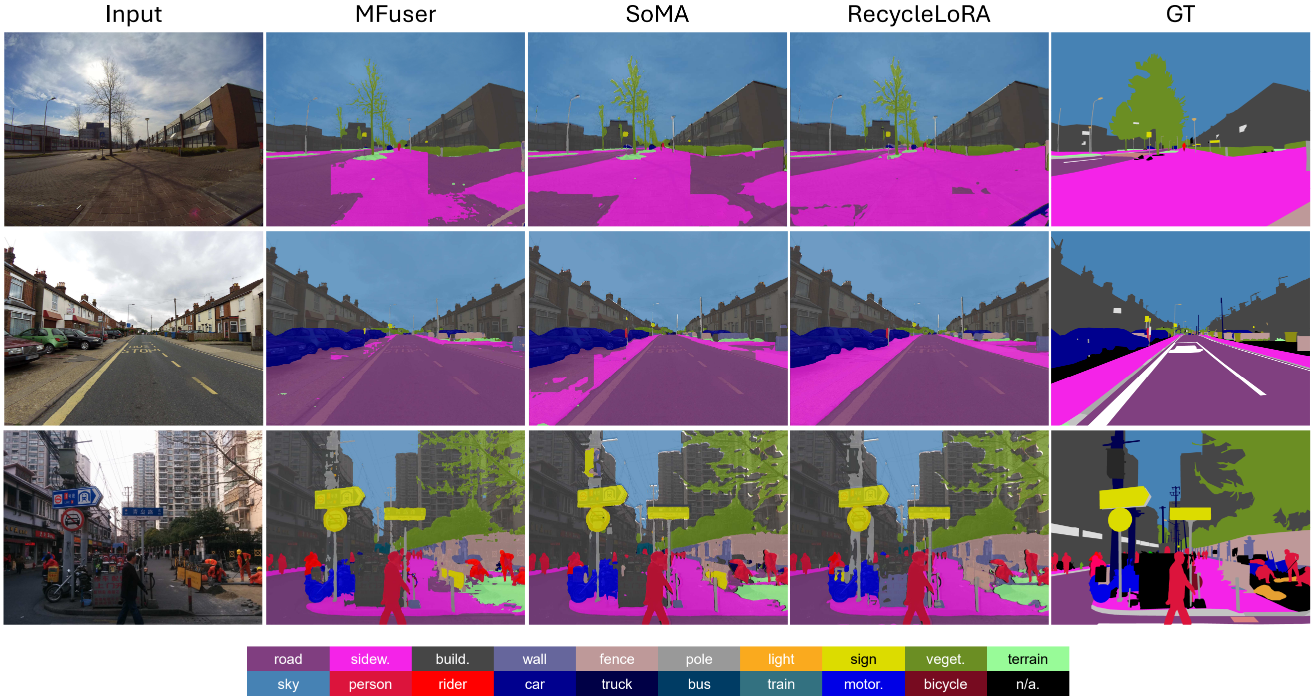

Figure 7. Qualitative comparison of semantic segmentation on the Mapillary. All models were trained on the GTAV.