

# Supplementary Materials: Generalized Source-free Domain-adaptive Segmentation via Reliable Knowledge Propagation

Anonymous Authors

## 0.1 Specific Algorithm of our LAM

The core design of LAM is described in the original paper, and we will elaborate on more specific implementation details. Affine transformations can achieve translation and rotation, but the semantic spatial layout in street scenes is relatively fixed. Therefore, when implementing LAM, we impose angle constraints on the mixed segments to prevent drastic angle changes. Given the affine transformation matrix  $\omega = \begin{pmatrix} \omega_{11} & \omega_{12} & \omega_{13} \\ \omega_{21} & \omega_{22} & \omega_{23} \end{pmatrix}$ , the rotation angle can be obtained using the following formula:  $\theta = \text{atan2}(\omega_{21}, \omega_{11}) \cdot \frac{180}{\pi}$ . Furthermore, LAM can be exploited multiple times to fill in unreliable regions of prediction with reliable segments. Thus, we use the Mixing process iteratively and set the termination conditions. Stop mixing when the mean uncertainty falls below 10%. The mean uncertainty is defined as  $\text{mean}(u) = \frac{1}{H \times W} \sum_{h=1}^H \sum_{w=1}^W u_{hw}$ ,  $W$  and  $H$  is the height and width. The details are in Algorithm 1:

### Algorithm 1 LAM Algorithm

**Input:** Reliable discrete segment set  $R = \{(x_{tr}^i, r^i)\}$ , and the prediction result set  $D_t = \{(x_t^i, \hat{y}_t^i, u_t^i)\}$  for target data  
**Output:** Set of reliable mixed samples  $\{(x_{tm}^i, \hat{y}_{tm}^i, u_{tm}^i)\}$

- 1: **for** each target data  $\{x_t^i, \hat{y}_t^i, u_t^i\}$  in  $D_t$  **do**
- 2:   **if** mean uncertainty  $\text{mean}(u)$  is less than 10% **then**
- 3:     continue
- 4:   Perform connected component analysis on the predicted uncertainty  $u_t^i$  to obtain the mask  $O$  for the largest area
- 5:   Find suitable segment  $r_d$  in the set  $R$  using Eq. (3)
- 6:   Initialize the localization network  $L(\omega)$
- 7:   Optimize  $L(\omega)$  using Eq. (5) with  $r_d$  and  $u_t^i$  to obtain the optimal affine transformation parameters  $\hat{\omega}$
- 8:   Calculate the learned rotation angle  $\theta$
- 9:   **if**  $-20^\circ > \theta$  or  $\theta > 20^\circ$  **then**
- 10:     continue
- 11:   Perform mixing according to Eq. (6), Eq.(7), Eq.(8) using  $\hat{\omega}$
- 12: Obtain the mixed prediction set  $\{(x_{tm}^i, \hat{y}_{tm}^i, u_{tm}^i)\}$

## 0.2 More Visualization of LAM

We provide more visualizations of the mixed target images, mixed pseudo-labels, and mixed uncertainties generated by our LAM in Fig. A. As shown in the second column of Fig. A, the regions in the yellow rectangular boxes of pseudo-labels are those with high uncertainty, i.e., corresponding to the brighter regions in our uncertainty maps of the third column. Subsequently, the most appropriate reliable segments associated with these high-uncertainty regions are selected and cover on them, obtaining mixed target images, mixed pseudo-labels, and mixed uncertainties, as shown in the

$\tau$	0.10	0.25	0.50	0.75
G $\rightarrow$ C: Source	60.1	61.0	61.4	60.9
G $\rightarrow$ C: Target	58.5	58.8	59.3	58.6

**Table A: The impact of hyper-parameter  $\tau$  on the adaptation performance (mIoU %) in the GTA  $\rightarrow$  Cityscapes task.**

last three columns of Fig. A. By comparing the third and sixth columns of Fig. A, we can see that the high-uncertainty regions in the pseudo-labels has almost been replaced. By comparing the second and fifth columns of Fig. A, the mixed pseudo-labels have less noise while maintaining the complete semantic layout. The generated high-quality paired data provides support for fine-tuning the diffusion model to the target domain, driving the synthesis of target-style and out-of-target style data.

## 0.3 More Synthesized Data via Our Fine-tuned text-to-image Diffusion Model

We provide more visualizations of the data synthesized by our fine-tuned diffusion model in Fig. B and Fig. C. These data are synthesized based on mixed layouts and reliable layouts, respectively. In addition to the semantic layouts, given various domain factors ‘fog’, ‘snow’, ‘hail’, ‘rain’, ‘heatwave’, ‘night’, ‘windy’, and ‘cold\_snap’ as prompts, the synthesized data significantly exhibit different out-of-domain styles and well maintains the essential semantic content (i.e., undistorted). These synthesized high-quality out-of-domain data provide supporting material for the model to learn strong generalization capabilities.

## 0.4 Impact of Hyper-parameter $\tau$

In Table A, we provide sensitivity analysis of the hyper-parameter  $\tau$  by setting it from 0.10 to 0.75. Table A shows that larger or smaller  $\tau$  causes a slight degradation in network performance. We analyze that this is because when  $\tau$  is large, too much noise is introduced; when  $\tau$  is small, the segment candidates are less rich. Finally, we choose  $\tau$  as 0.5 for all adaptation tasks.

## 0.5 More Comparisons of Segmentation Results

Between our method and the existing SFDA method DTST [4], we provide more comparisons of their segmentation results on the unseen domain BDD-100K [3] in Fig. D. As shown in Fig. D, under various weather conditions, our method consistently produces more accurate predictions than the DTST method and our results have a more complete semantic structure. This contributes to the fact that our network achieves strong generalization ability by retraining on data with multi-domain characteristics.

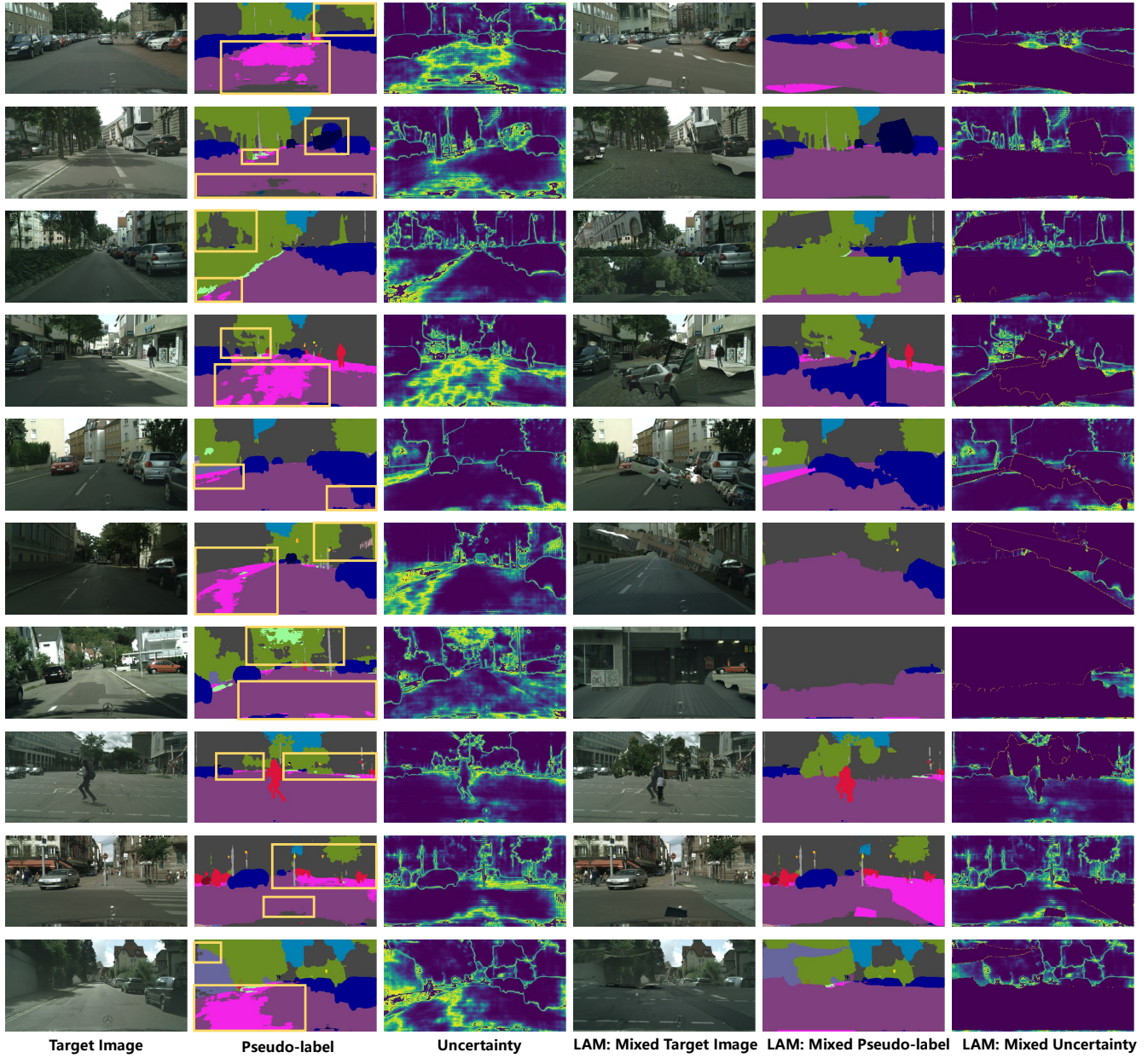


Figure A: More visualization of our proposed LAM, including mixed target images, mixed pseudo-labels, and mixed uncertainty.

## 0.6 Details on Unseen Domain Datasets.

**BDD-100K Dataset [3].** The BDD-100K dataset is an urban dataset of real scenes collected from dashcam video frames in various locations of the United States. It consists of 7,000 training, 1,000 validation, and 2,000 testing images of resolution  $1,280 \times 720$ .

**Mapillary Dataset [1].** The Mapillary dataset contains tens of thousands of real images collected from street scenes all around the world. Pixel-level labels with 66 classes are provided, but only 19 classes shared are used in our experiments. The training and validation sets consist of 18,000 and 2,000 images. The resolution of the image is  $1,920 \times 1,080$ .

**ACDC Dataset [2].** The ACDC dataset shares the same semantic classes with Cityscapes and is collected in four different adverse visual conditions: Fog, Night, Rain, and Snow. It comprises 1,600 training, 406 validation, and 2,000 test images with a resolution of  $1,920 \times 1,080$ .

## REFERENCES

- [1] Gerhard Neuhold, Tobias Ollmann, Samuel Rota Buló, and Peter Kotschieder. 2017. The mapillary vistas dataset for semantic understanding of street scenes. In *Proceedings of the IEEE international conference on computer vision*. 4990–4999.
- [2] Christos Sakaridis, Dengxin Dai, and Luc Van Gool. 2021. ACDC: The adverse conditions dataset with correspondences for semantic driving scene understanding.

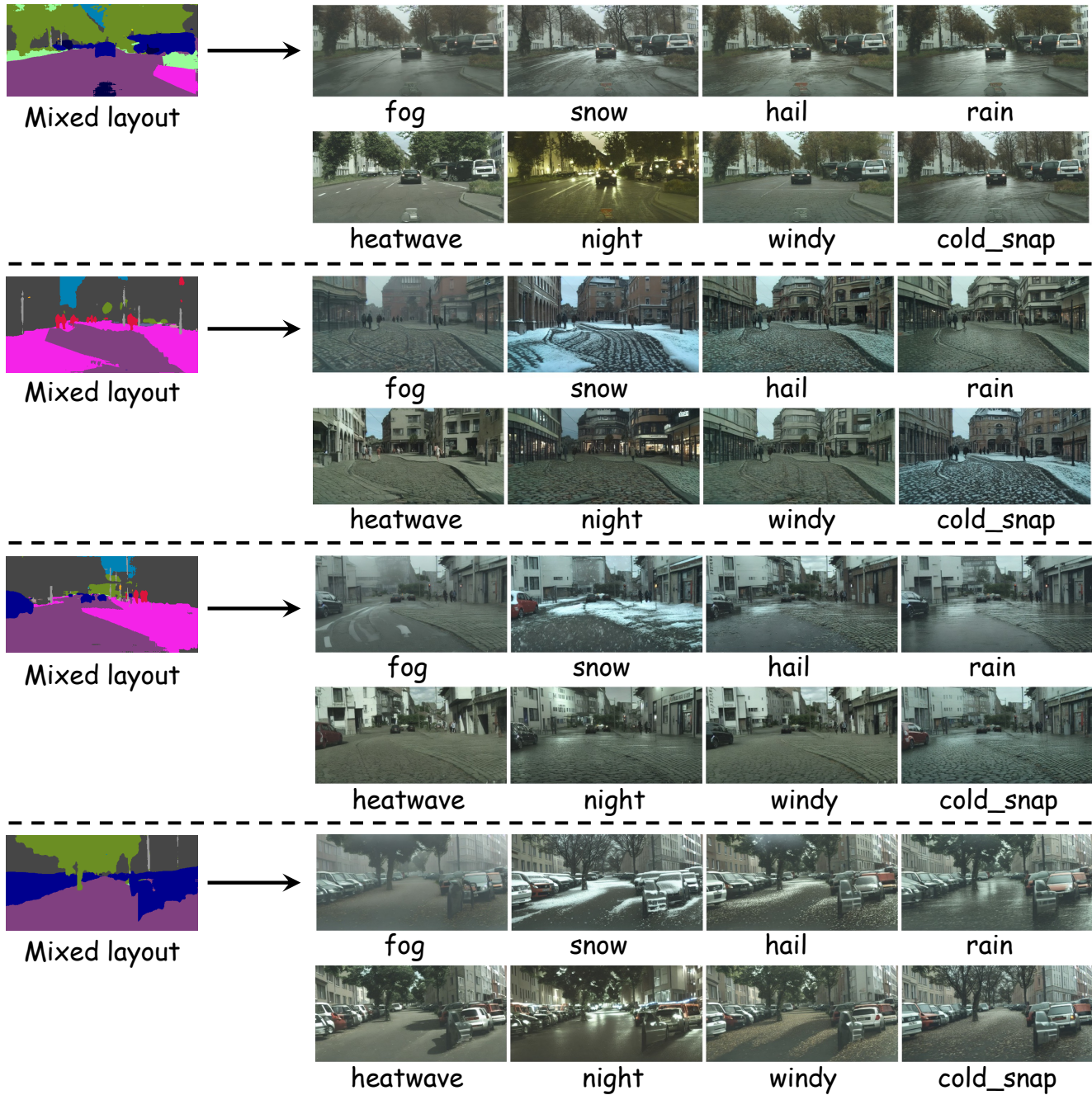


Figure B: More visualization of the synthesized image based on mixed layouts.

In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 10765–10775.

- [3] Fisher Yu, Haofeng Chen, Xin Wang, Wenqi Xian, Yingying Chen, Fangchen Liu, Vashisht Madhavan, and Trevor Darrell. 2020. Bdd100k: A diverse driving dataset for heterogeneous multitask learning. In *Proceedings of the IEEE/CVF conference*

on computer vision and pattern recognition. 2636–2645.

- [4] Dong Zhao, Shuang Wang, Qi Zang, Dou Quan, Xiutiao Ye, and Licheng Jiao. 2023. Towards Better Stability and Adaptability: Improve Online Self-Training for Model Adaptation in Semantic Segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 11733–11743.

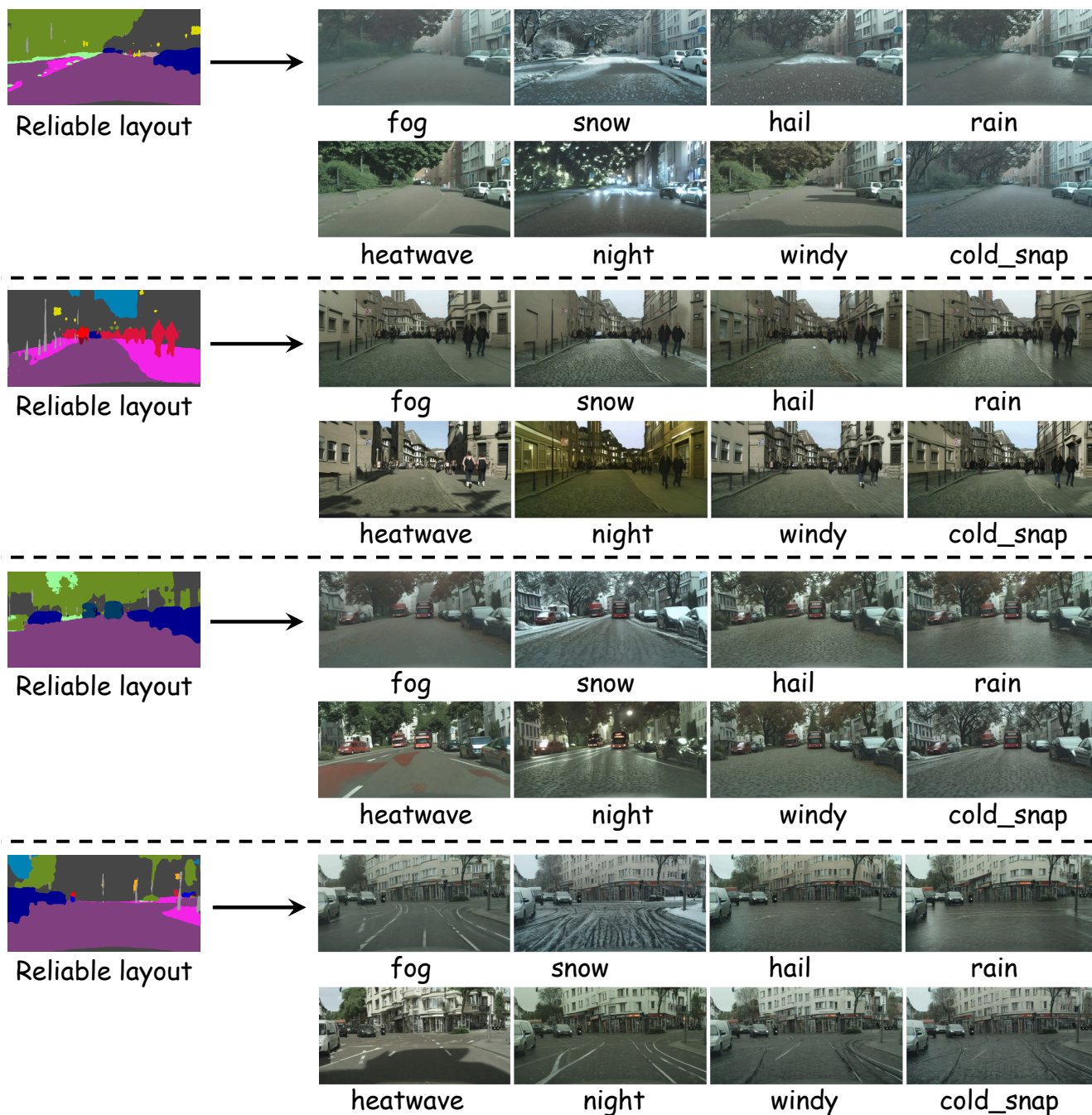


Figure C: More visualization of the synthesized image based on reliable layouts.

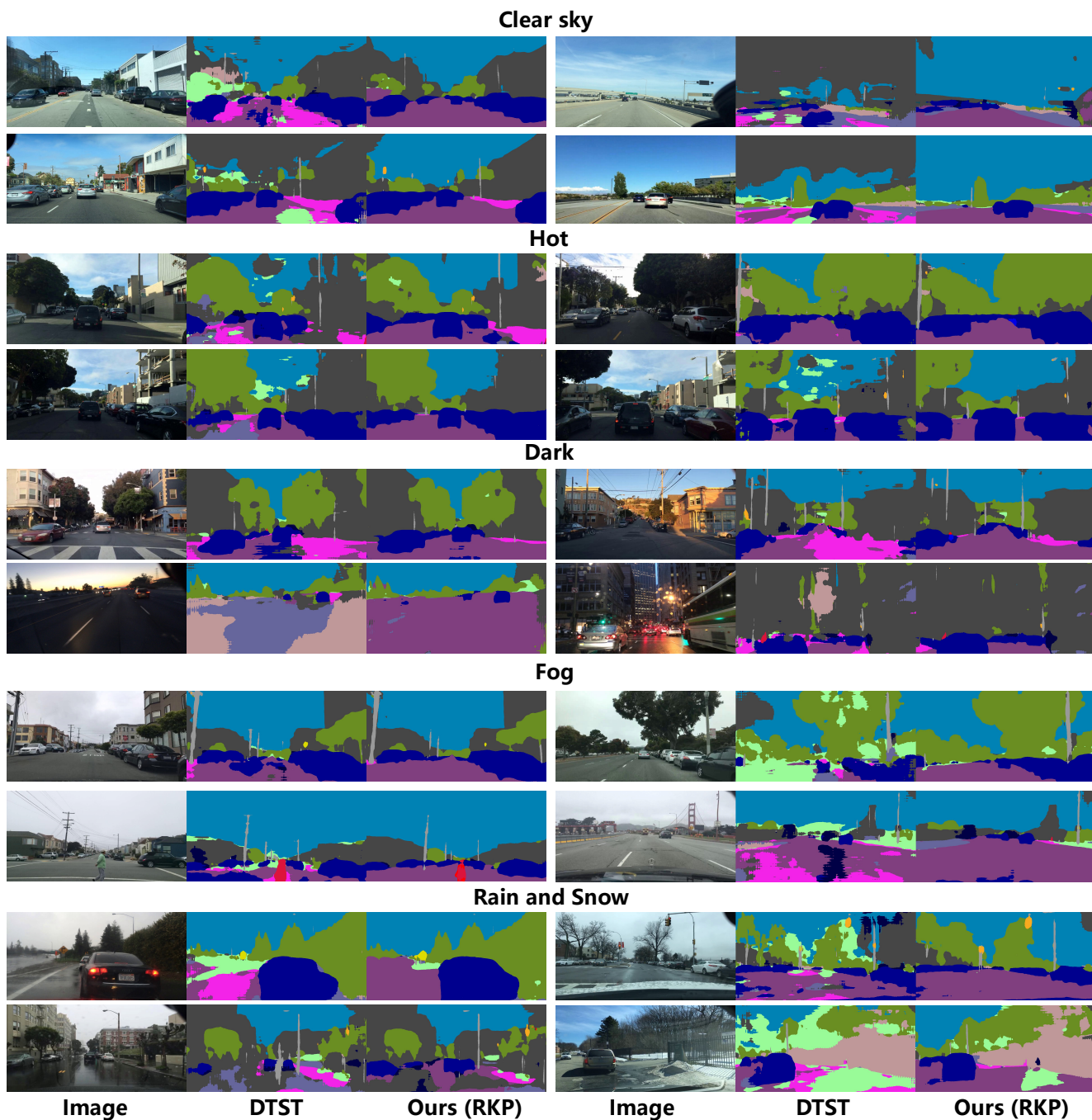


Figure D: More visualization of adaptation results on unseen domain. Comparison of our method and the SFDA method DTST [4] on unseen domain BDD-100K [3] with various weather conditions.