FIFO: Learning Fog-invariant Features for Foggy Scene Segmentation
— *Supplementary Material* —

Sohyun Lee
GSAI, POSTECH
lshig96@postech.ac.kr

Taeyoung Son*
NALBI
taeyoung@nalbi.ai

Suha Kwak
GSAI, POSTECH
suha.kwak@postech.ac.kr

http://cvlab.postech.ac.kr/research/FIFO/

1. Algorithm of FIFO

We present the training procedure of FIFO in Algorithm 1.

Algorithm 1 : Training FIFO

Input: Pretrained fog-pass filtering module for the l^{th} layer: $F^l(\cdot)$, Segmentation network: $S(\cdot)$, Number of layers: L, Batch size per domain: m, Segmentation prediction: P, Segmentation label: Y, Input image set $\{I^{\mathsf{CW}}, I^{\mathsf{SF}}, I^{\mathsf{RF}}\}$: x, Subset of two elements from domain set $\{\mathsf{CW},\mathsf{SF},\mathsf{RF}\}$: $\{a, b\}$ and Segmentation label set $\{Y^{\mathsf{CW}}, Y^{\mathsf{RF}}\}$: y.
Output: Optimized segmentation network $S(\cdot)$.

for $\{1, \ldots,$ # of training iterations$\}$ **do**
  Sample mini-batch $\{x_i\}_{i=1}^m$
  for $\{l \longleftarrow 1$ to $L\}$ **do**
    $\mathcal{L}_{\text{F}^l} \longleftarrow \mathcal{L}_{\text{F}^l}(\{\mathbf{f}_i^l\}_{i=1}^m)$
    Update the fog-pass filtering module F^l
  end for
  Sample mini-batch $\{x_j\}_{j=1}^m$ and $\{y_j\}_{j=1}^m$
  Sample the pair $\{I^a, I^b\} \in x_j$
  for $\{l \longleftarrow 1$ to $L\}$ **do**
    $\{\mathbf{f}_j^{a,l}\}_{j=1}^m, \longleftarrow \{F^l(\mathbf{u}_j^{a,l})\}_{j=1}^m$
    $\{\mathbf{f}_j^{b,l}\}_{j=1}^m, \longleftarrow \{F^l(\mathbf{u}_j^{b,l})\}_{j=1}^m$
    $\mathcal{L}_{\text{fsm}}^l \longleftarrow \{\mathcal{L}_{\text{fsm}}^l(\mathbf{f}_j^{a,l}, \mathbf{f}_j^{b,l})\}_{j=1}^m$
  end for
  if $\{a, b\} == \{\mathsf{CW}, \mathsf{SF}\}$ **then**
    $\mathcal{L}_{\text{con}} \longleftarrow \sum_i \text{KLdiv}(P_i^a, P_i^b)$
  end if
  if $\{a, b\} \cap \{\mathsf{CW}, \mathsf{SF}\} \neq \emptyset$ **then**
    $\mathcal{L}_{\text{seg}} \longleftarrow -\frac{1}{n} \sum Y \log P$
  end if
  $\mathcal{L}_{\text{S}} \longleftarrow \sum_l \mathcal{L}_{\text{fsm}}^l + \mathcal{L}_{\text{con}} + \mathcal{L}_{\text{seg}}$
  Update the segmentation network S
end for

*This work was done while Taeyoung Son was in POSTECH.

Consequently, the total objective of FIFO is following:

$$\sum_l \min_{F^l} \mathcal{L}_{F^l}^l + \min_S (\sum_l \mathcal{L}_{\text{fsm}}^l + \mathcal{L}_{\text{con}} + \mathcal{L}_{\text{seg}}), \quad (1)$$

where l is the layer index.

2. Generalization to Other Weather Conditions

We investigate the generalization ability of FIFO on the other weather conditions, *rainy* [7] and *frosty* [6] versions of the Cityscapes [2] dataset, according to the severity of the corruptions. Figure A1 presents the performance of baseline [9], an ordinary segmentation model trained on clear weather images, and FIFO on varying the severity of frosty and rainy corruptions. FIFO tends to be robust to each corruption than the baseline, even when the corruption gets severe. Table A1 and Table A2 show detailed quantitative results of baseline and FIFO on frosty and rainy corruptions, respectively. Additional qualitative results are presented in Figure A7.

We also evaluate FIFO on ACDC, the real-world adverse conditions dataset for semantic driving scene understanding. For fair comparisons on ACDC [11], FIFO is trained on the Cityscapes, Foggy Cityscapes-DBF, and Foggy Zurich datasets, following the unsupervised learning setting of the benchmark. As summarized in Table. A3, FIFO outperforms the existing foggy scene segmentation methods reported in [11] for all four conditions.

		Corruption Severity				
		1	2	3	4	5
Frost	Baseline	45.53	23.59	14.97	13.60	10.66
	FIFO	46.85	30.64	22.66	20.88	17.47

Table A1. Quantitative results on Frosty Cityscapes according to the severity of corruptions.

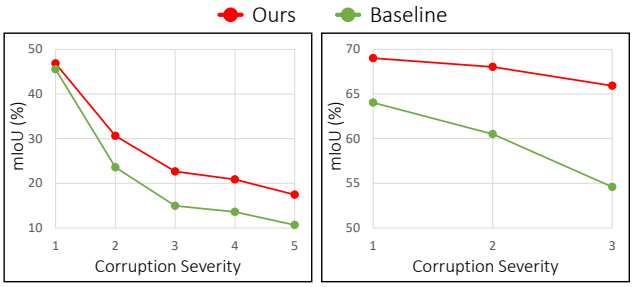

Figure A1. Performance (mIoU) versus the corruption severity. Ours (FIFO) and baseline are evaluated on Frosty Cityscapes and Rainy Cityscapes.

		Corruption Severity			
		1	2	3	all
Rain	Baseline	64.03	60.51	54.62	57.60
	FIFO	69.01	68.03	65.92	67.62

Table A2. Quantitative results on Rainy Cityscapes according to the severity of corruptions.

Method	Fog	Rain	Snow	Night	Avg.
RefineNet	46.4	52.6	43.3	29.0	43.7
SFSU [2]	45.6	51.6	41.4	29.5	42.9
CMAda	51.2	53.4	47.6	32.0	47.1
FIFO	**54.1**	**58.8**	**51.8**	**32.5**	**49.4**

Table A3. Quantitative results on the ACDC dataset.

3. Independence Analysis of Fog Factors

In this section, we quantitatively evaluate the independence of the fog factors from the image content compared to that of the Gram matrices from the content. To this end, we design a content-pass filtering module that is optimized to extract content-relevant information, which we call content factors.

Training Content-pass Filtering Module. Let I^a and I^b be a pair of images from the mini-batch, and C^l denote the content-pass filtering module attached to the l^{th} layer of the segmentation network. Let $\mathbf{u}^{a,l}$ and $\mathbf{u}^{b,l}$ be the vectorized upper triangular parts of the Gram matrices computed from the l^{th} feature maps of I^a and I^b. Then the content factors of the two images are computed by $\mathbf{c}^{a,l} = C^l(\mathbf{u}^{a,l})$ and $\mathbf{c}^{b,l} = C^l(\mathbf{u}^{b,l})$. In contrast to the fog-pass filtering module, this module is optimized to learn an embedding space of content factors where the pairs having the same content, *i.e.*, CW–SF are grouped closely and else pairs are far from each other. Given the set of every image pair $\mathcal{P}$ in the mini-batch, the loss function for C^l is designed accordingly as follows:

$$\mathcal{L}_{C^l} = \sum_{(a,b)\in\mathcal{P}} \Big\{ \big(1-\mathbb{I}(a,b)\big)\Big[m - d\big(\mathbf{f}^{a,l},\mathbf{f}^{b,l}\big)\Big]_+^2 + \mathbb{I}(a,b)\Big[d\big(\mathbf{f}^{a,l},\mathbf{f}^{b,l}\big) - m\Big]_+^2 \Big\}, \tag{2}$$

where $d(\cdot)$ is the cosine distance, m is a margin, and $\mathbb{I}(a,b)$ denotes the indicator function that returns 1 if the pair of I^a and I^b is a CW–SF pair and 0 otherwise, respectively.

Independence Analysis of Fog Factors. We design the independence score to quantitatively evaluate and compare the independence of fog factors and that of Gram matrices from content factors. We first measure the score of the independence of fog factors from content factors. To this end, we select one image I_i, then choose k images $\{I_n\}$ whose fog factors are most similar to the fog factor f_i of the selected image I_i. Then, we also choose k images $\{I_m\}$ whose content factors are most similar to the content factor c_i of the selected image I_i. After that, we compute the proportion of the number of overlapped images $|\{I_n\} \cap \{I_m\}|$ between $\{I_n\}$ and $\{I_m\}$. Then, we repeat the process for all N images and calculate the average proportion as the independence score.

Let I, f, and c be an image, a fog factor, and a content factor, then, the independence score is calculated as follows:

$$\text{IndependenceScore}(\mathcal{F},\mathcal{C}) = 1 - \frac{1}{N}\sum_{i=1}^{N}\frac{1}{k}\Big\{\big|\{I_n | f_n \in \mathcal{F}, d(f_i,f_n) \le d(f_i,f_k)\} \cap \{I_m | c_m \in \mathcal{C}, d(c_i,c_m) \le d(c_i,c_k)\}\big|\Big\}, \tag{3}$$

where $d(\cdot)$ and k are a cosine distance and a number of selecting similar factors set to 200, f_k and c_k are the k th most similar fog factor from f_i and the k th most similar content factor from c_i, where $\mathcal{F}$ and $\mathcal{C}$ denote the set of fog factors and content factors, respectively. We then replace the fog factors with Gram matrices, then repeat the same process for calculating the independence score of Gram matrices from the content factors.

Figure A2 presents the independent score of fog factors and Gram matrices from content factors. Note that the experiment settings and dataset configurations are all the same as in the main paper. Figure A2 proves that fog factors are more independent to content factors compared to Gram matrices, as desired. It indicates that the fog-pass filtering module extracts only fog-relevant information apart from the image content.

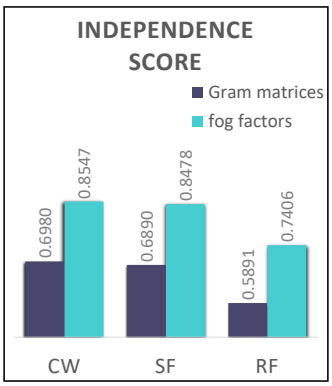

Figure A2. Independence score of fog factors and Gram matrices from content factors.

4. Empirical Verification Using Evaluation Dataset

We present additional results of the empirical verification using evaluation splits of the datasets, *i.e.*, Cityscapes (500 images) as CW, Foggy Cityscapes-DBF (500 images) as SF, and Foggy Zurich-test v2 (40 images) and Foggy Driving (101 images) as RF. Fig. A3 shows that the tendency of the results is consistent with that of Fig. 4 of the main paper.

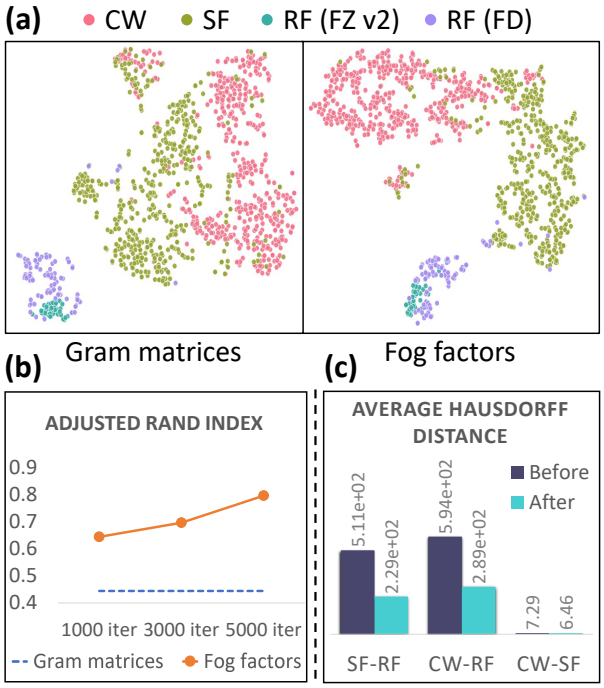

Figure A3. Results of the empirical verification using the evaluation datasets. (a) t-SNE visualization of distributions of Gram matrices and their fog factors. (b) Comparison between the quality of k-means clustering of the fog factors and Gram matrices in adjusted Rand index. (c) The fog-style gap between different domains before and after training with FIFO.

5. Impact of Fog Factors

We present additional comparison results for the quality of k-means clustering [5] of the Gram matrices and that of the corresponding fog factors in other measures, normalized mutual information [4], and adjusted mutual information [12]. All of the measures prove the impact of fog factors in that they are more clustered than Gram matrices according to each fog condition as shown in Figure A4 and Table A4.

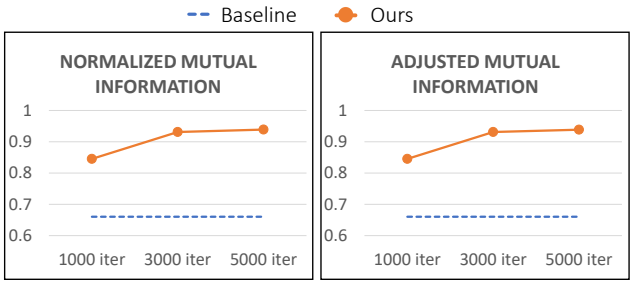

Figure A4. Comparison between fog factors and Gram matrices for the quality of k-means clustering by normalized mutual information and adjusted mutual information.

Comparison	1000 iter	3000 iter	5000 iter
Normalized Mutual Information			
Gram matrix	0.6602	0.6602	0.6602
Fog factor	0.8453	0.9313	0.9387
Adjusted Mutual Information			
Gram matrix	0.6601	0.6601	0.6601
Fog factor	0.8452	0.9313	0.9387
Adjusted Rand Index			
Gram matrix	0.6304	0.6304	0.6304
Fog factor	0.8683	0.9533	0.9596

Table A4. Quantitative results of the quality of k-means clusters of the Gram matrices and fog factors in normalized mutual information, adjusted mutual information, and adjusted Rand index [8].

6. Effect of Fog Style Matching Loss

This section conducts extensive experiments to investigate the effect of the fog style matching loss $\mathcal{L}_{fsm}$. In FIFO, the fog style matching loss $\mathcal{L}_{fsm}$ is carried out by bidirectionally matching each fog condition (*i.e.*, CW, SF, and RF), so we denote it as a 'Bidirectional' setting in this section. We conduct additional experiments about variants of the fog style matching loss $\mathcal{L}_{fsm}$ in the 'Unidirectional' setting (from Fog to Clear, from Clear to Fog). For 'Fog

Method	Image Pair			FZ	FDD	FD	C-Lindau
	CW & SF	CW & RF	SF & RF	mIoU (%)	mIoU (%)	mIoU (%)	mIoU (%)
1 pair							
Unidirectional (Fog to Clear)		CW $\leftarrow$ RF		38.5	36.3	45.6	67.1
Unidirectional (Clear to Fog)		CW $\rightarrow$ RF		36.6	36.9	46.2	64.7
Bidirectional		CW $\leftrightarrow$ RF		37.7	40.3	47.2	66.0
2 pairs							
Unidirectional (Fog to Clear)	CW $\leftarrow$ SF		SF $\leftarrow$ RF	43.3	39.2	48.8	68
Unidirectional (Clear to Fog)	CW $\rightarrow$ SF		SF $\rightarrow$ RF	43.4	39.3	48.4	63.3
Bidirectional	CW $\leftrightarrow$ SF		SF $\leftrightarrow$ RF	46.0	47.6	50.0	62.3
3 pairs							
Unidirectional (Fog to Clear)	CW $\leftarrow$ SF	CW $\leftarrow$ RF	SF $\leftarrow$ RF	44.4	42.6	47.1	68.8
Unidirectional (Clear to Fog)	CW $\rightarrow$ SF	CW $\rightarrow$ RF	SF $\rightarrow$ RF	44.1	36.5	46	64.5
Bidirectional (FIFO)	CW $\leftrightarrow$ SF	CW $\leftrightarrow$ RF	SF $\leftrightarrow$ RF	**48.4**	**48.9**	**50.7**	64.8

Table A5. Analysis on the impact of the fog style matching loss. CW, SF, and RF denote clear weather, synthetic fog, and real fog, respectively.

to Clear' settings, fog styles of real foggy images are unidirectionally matched to those of clear weather images, which is regarded as feature-level dehazing on real foggy images. This is implemented simply by detaching the gradient flows from the fog style matching loss $\mathcal{L}_{fsm}$ to clear weather images. For 'Clear to Fog' settings, fog styles of clear weather images are unidirectionally matched to real foggy images similar to feature-level fog synthesis on clear weather images. This is also implemented by detaching the gradient from the fog style matching loss $\mathcal{L}_{fsm}$ to real foggy images.

Table A5 summarizes the results. We found that the bidirectional fog style matching outperforms its unidirectional counterpart when the same domain pairs are involved; this result justifies the fog style matching loss in FIFO. In addition, unidirectional (Fog to Clear) models have superior performance on the clear weather dataset [3] compared to others due to the effect of focusing on clear weather conditions.

7. Generalization on Deep Features

This section empirically investigates the generalization of fog-invariant learning of our method on deep features. It has been reported in domain adaptation and generalization literature [1,13] that domain alignment at bottom layers closes the domain gap of deeper layer features. As shown in Fig. A5, we empirically verify our case: The average Hausdorff distances between ResBlock4 features from different domains decrease noticeably by FIFO.

8. Comparison with UDA

In this section, we discuss the reason for failure when UDA methods are applied to the foggy scene segmentation

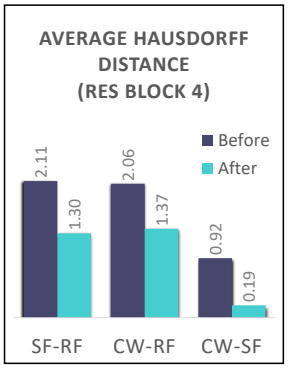

Figure A5. Distances between sets of deep features from different domains before and after training with FIFO.

in Table 1 of the main paper.

Analysis on the failure of FDA. We suspect this is because the style representation of Fourier domain adaptation (FDA) is not suitable for handling foggy scenes: FDA considers the low-frequency spectrum of an image as its style, but in the case of a foggy image, both its style and content lie in its low-frequency spectrum. We verify this via qualitative results of the spectral style transfer, an intermediate step in FDA. Fig. A6 presents examples of the style transfer from CW to RF and from RF to CW. The former causes severe artifacts on RF as the content CW as well as its style is transferred. On the other hand, the latter applies fog effects to CW, but the result is not realistic.

Superiority of FIFO over Domain Adversarial Learning. First, FIFO minimizes the entire objective function as presented in Section 1 while domain adversarial learning optimizes the min-max loss. Hence, FIFO does not suffer from the instability issue of adversarial learning in train-

ing. Second, FIFO can model and exploit within-domain fog style variations better than the domain adversarial learning (*e.g.*, DANN). This is crucial since images of the same fog condition have different fog styles in general. FIFO achieves this property by the losses in Eq. (1) and Eq. (3) of the main paper; the former motivated by metric learning enables the fog-pass filter to learn within-domain fog style variations, and the latter enables the segmentation model to keep such variations while closing style gaps only between different fog domains. Accordingly, thanks to the superiority of FIFO over domain adversarial learning, FIFO clearly outperforms DANN in Table 1 of the main paper.

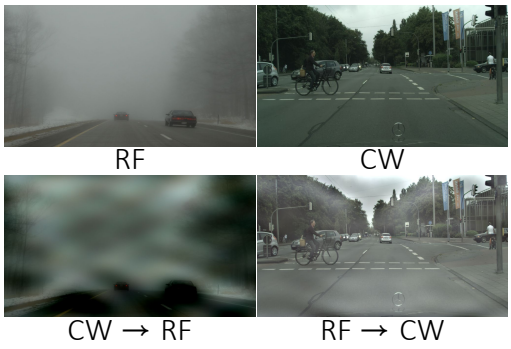

Figure A6. Outputs of spectral style transfer in FDA.

9. Comparison with Variants of CMAda

This section presents the comparison of FIFO to the variants of CMAda reported in [3], which suggests the current best performing model, CMAda3+. Table A6 demonstrates the superiority of FIFO over the variants of CMAda. In Table A6, CMAda models denoted '+' are conducted the additional procedure of fog densification for making the fog density of real foggy training images similar to target fog density of the test real foggy images. FIFO outperforms all variants of CMAda regardless of their number of stages and densification procedure.

method	FZ test v2	FDD	FD
CMAda1 [3]	38.9	36.6	46.0
CMAda2 [10]	42.9	37.3	48.5
CMAda3 [3]	43.7	40.6	48.9
CMAda2+ [3]	43.4	40.1	49.9
CMAda3+ [3]	46.8	43.0	49.8
FIFO	**48.4**	**48.9**	**50.7**

Table A6. Comparison of FIFO to variants of CMAda. '+' denotes models applied the additional procedure of fog densification for real foggy training datasets. The numbers attached to CMAda means the number of stages for curriculum learning.

10. Additional Qualitative Results

This section presents additional qualitative results omitted in the main sections due to the space limit. More segmentation results of FIFO are illustrated in Figure A8. We compare the results between FIFO, a variant of FIFO, by directly reducing the gap between Gram matrices and baseline. Overall, FIFO offers higher quality segmentation results than the baseline regardless of fog density and datasets. Specifically, FIFO seems best performing on parts where dense fog is laid while other models fail, which indicates FIFO working as desired. Figure A9 exihibits additional qualitative results on image reconstruction. Likewise, the image quality where dense fog is laid is improved, which implies FIFO extract fog-invariant features. In addition, clear weather images, as well as foggy images, become more clear when the features are trained by FIFO.

References

[1] Sungha Choi, Sanghun Jung, Huiwon Yun, Joanne T Kim, Seungryong Kim, and Jaegul Choo. Robustnet: Improving domain generalization in urban-scene segmentation via instance selective whitening. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021.

[2] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[3] Dengxin Dai, Christos Sakaridis, Simon Hecker, and Luc Van Gool. Curriculum model adaptation with synthetic and real data for semantic foggy scene understanding. *International Journal of Computer Vision (IJCV)*, 2020.

[4] Pablo A Estévez, Michel Tesmer, Claudio A Perez, and Jacek M Zurada. Normalized mutual information feature selection. *IEEE Transactions on neural networks*, 2009.

[5] John A Hartigan and Manchek A Wong. Algorithm as 136: A k-means clustering algorithm. *Journal of the royal statistical society. series c (applied statistics)*, 1979.

[6] Dan Hendrycks and Thomas Dietterich. Benchmarking neural network robustness to common corruptions and perturbations. In *Proc. International Conference on Learning Representations (ICLR)*, 2019.

[7] Xiaowei Hu, Chi-Wing Fu, Lei Zhu, and Pheng-Ann Heng. Depth-attentional features for single-image rain removal. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[8] Lawrence Hubert and Phipps Arabie. Comparing partitions. *Journal of classification*, 1985.

[9] Vladimir Nekrasov, Chunhua Shen, and Ian Reid. Lightweight refinenet for real-time semantic segmentation. In *Proc. British Machine Vision Conference (BMVC)*, 2018.

[10] Christos Sakaridis, Dengxin Dai, Simon Hecker, and Luc Van Gool. Model adaptation with synthetic and real data

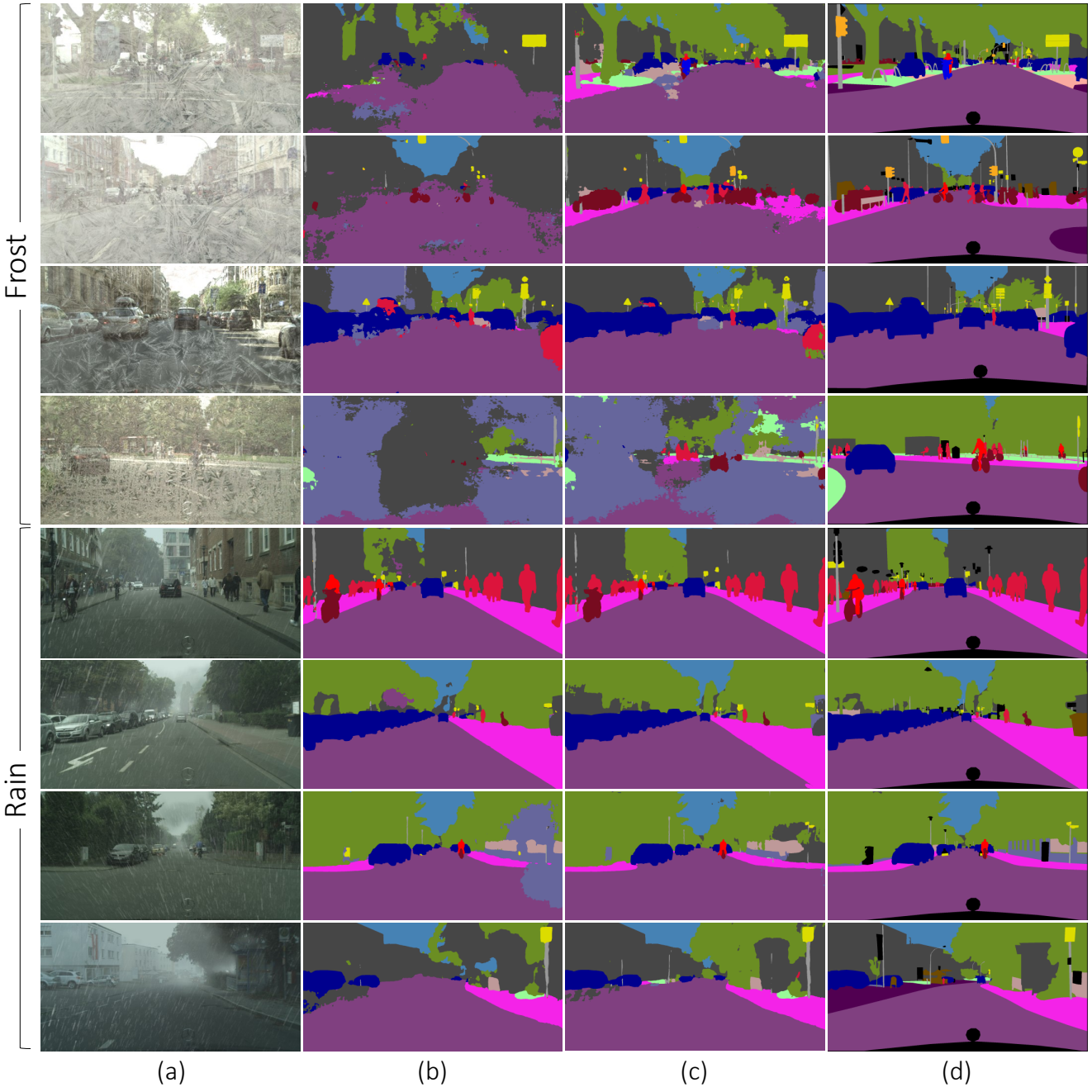

Figure A7. Additional qualitative results on *frost* and *rain* weather corruptions. (a) Weather corrupted input images. (b) Baseline. (c) FIFO. (d) Groundtruth.

for semantic dense foggy scene understanding. In *Proc. European Conference on Computer Vision (ECCV)*, 2018.

[11] Christos Sakaridis, Dengxin Dai, and Luc Van Gool. ACDC: The adverse conditions dataset with correspondences for semantic driving scene understanding. In *Proc. IEEE International Conference on Computer Vision (ICCV)*, 2021.

[12] Nguyen Xuan Vinh, Julien Epps, and James Bailey. Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance. *Journal of Machine Learning Research (JMLR)*, 2010.

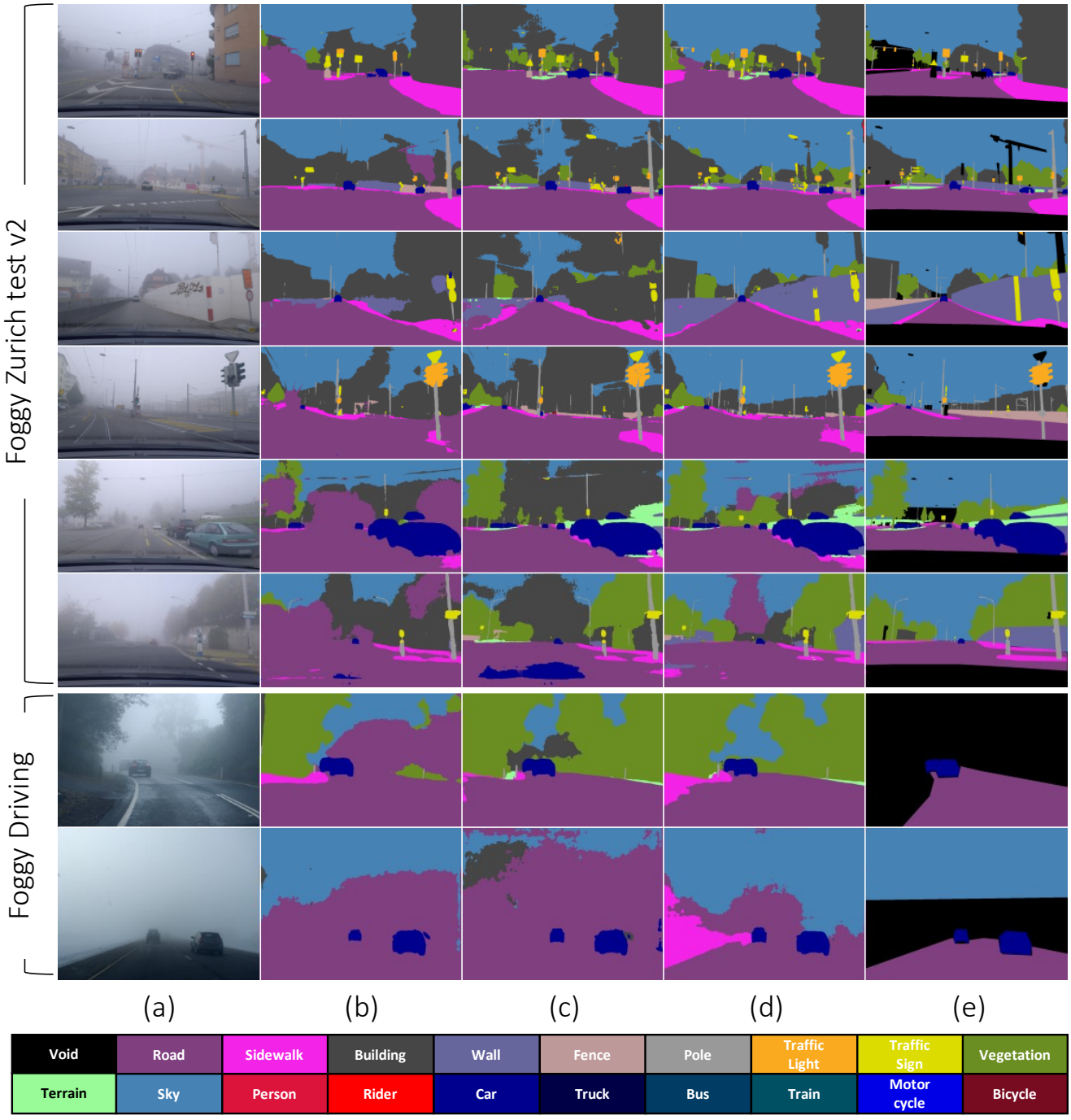

Figure A8. Additional qualitative results on the real foggy datasets. (a) Real foggy images. (b) Baseline. (c) Reduced version of FIFO closing the gap between gram matrices. (d) FIFO. (e) Groundtruth.

[13] Kaiyang Zhou, Yongxin Yang, Yu Qiao, and Tao Xiang. Domain generalization with mixstyle. In *Proc. International Conference on Learning Representations (ICLR)*, 2021.

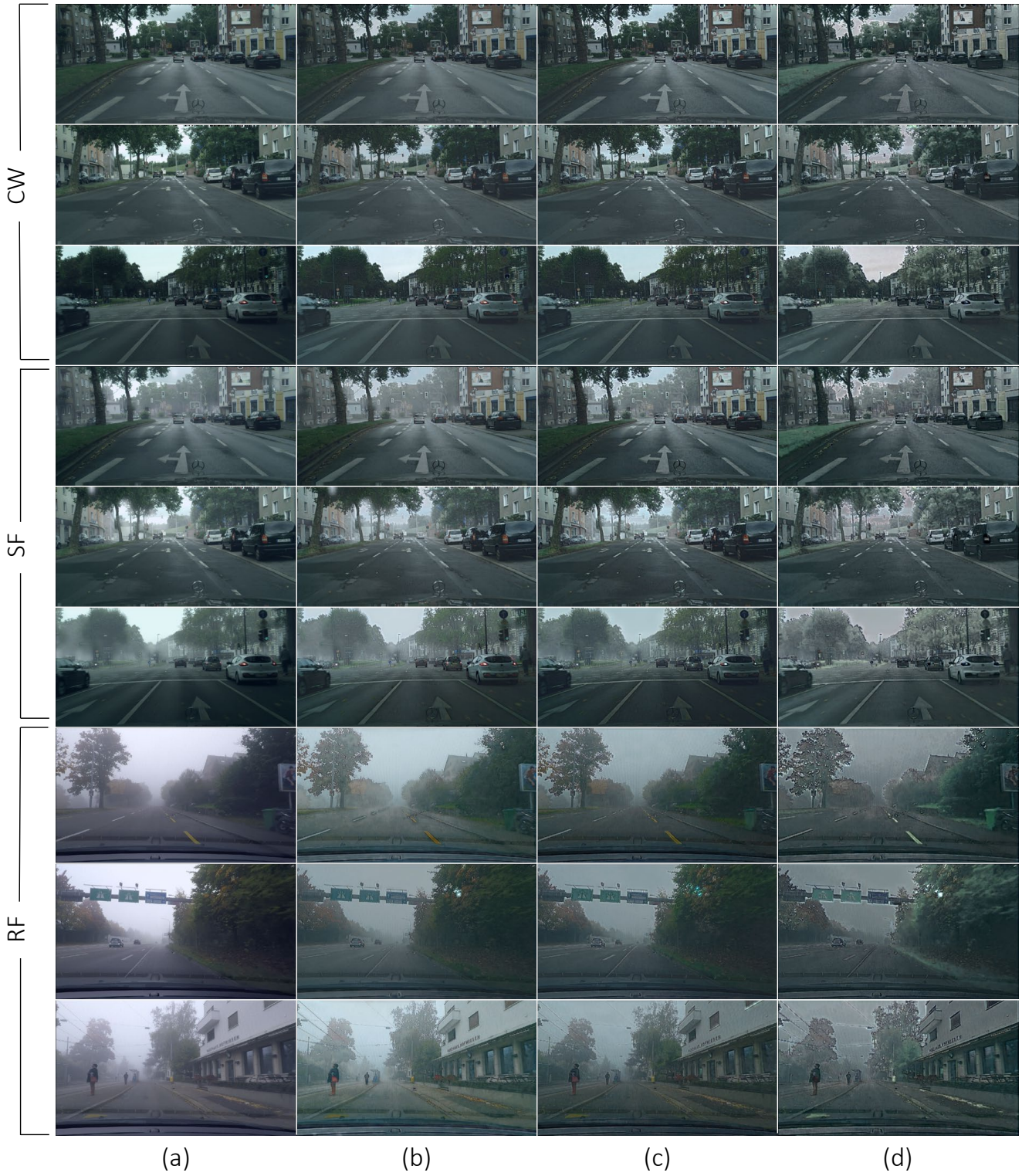

Figure A9. Additional qualitative results on image reconstruction. (a) Real foggy images. (b) Baseline. (c) Reduced version of FIFO closing the gap between gram matrices. (d) FIFO.