Supplementary Material: Cluster, Split, Fuse, and Update: Meta-Learning for Open Compound Domain Adaptive Semantic Segmentation

Rui Gong [1], Yuhua Chen [1], Danda Pani Paudel [1], Yawei Li [1], Ajad Chhatkuli [1],
Wen Li [3], Dengxin Dai [1], Luc Van Gool [1,2]
[1] Computer Vision Lab, ETH Zurich, [2] VISICS, KU Leuven, [3] UESTC
{gongr, yuhua.chen, paudel, yawei.li, ajad.chhatkuli, dai, vangool}@vision.ee.ethz.ch,
liwenbnu@gmail.com

In this supplementary, we provide additional information for,

S1 implementation details of our MOCDA model,

S2 more detailed information about the datasets in our experiments,

S3 additional experimental quantitative results and qualitative results on the OCDA benchmark,

S4 additional visualization results for the style code and hypernetwork prediction.

S1. Detailed Implementation of our MOCDA model

In the main paper, we introduce our MOCDA model in the Sec. 3 and the implementation details in the Sec. 4.1. Here we provide more detailed implementation of different modules in our MOCDA model, separately.

Cluster. In the cluster module, we train the MUNIT [5] model to translate between the source domain images and the compound target domain images in the unsupervised way. We follow the experimental set up in the urban scene image translation set up in MUNIT [5]. The shortest side of the images are firstly resized to 512, and then the images are randomly cropped with the size of 400×400. The loss weights for image reconstruction loss, style reconstruction loss, content reconstruction loss, and domain-invariant perceptual loss are set as 10, 1, 1, and 1, respectively. The Adam optimizer [7] is adopted with $\beta_1 = 0.5, \beta_2 = 0.999$, and the learning rate is set as 0.0001. Also, the dimension of the style code is set as 8. The number of the clusters K is set as 4.

Split, Fuse, and Update. In the split and fuse module, we have the semantic segmentation network and the discriminator. We adopt the DeepLab-VGG16 [2, 13] with synchronized batch normalization layer [6] for the semantic segmentation network. And we adopt the discriminator structure in [14]. The compound target domain images and the open domain images, from BDD100K [16], Cityscapes[3], WildDash [17] and KITTI [1], are resized to 1024×512, and the source domain images from GTA5 [12] and SYNTHIA-SF [4] are resized to 1280×720. The λ_1 in Eq. (7), and λ_2 in Eq. (14) of the main paper are set as 0.001. In the update module, during the training stage, the δ in Eq. (17) is set as 0.0001. In the split, fuse and update module, we adopt the SGD optimizer to train the hypernetwork and the semantic segmentation network, where the momentum is 0.9 and the weight decay is 5×10^{-4}. The learning rate is set as 2.5×10^{-4}, and uses the polynomial decay strategy with power of 0.9 as done in [14]. We keep the same learning rate for online updating the hypernetwork and the semantic segmentation network. Also, we adopt the Adam optimizer [7] for training the discriminator with $\beta_1 = 0.9, \beta_2 = 0.99$. The learning rate is set as 1.0×10^{-4} and uses the polynomial decay strategy with power of 0.9. And our MOCDA model is implemented with PyTorch [11].

S2. Datasets Overview

In Sec. 4 of the main paper, we introduce the experiments setup of the OCDA benchmark, and there are six datasets in total, GTA5 [12], SYNTHIA-SF [4], BDD100K [16], Cityscapes [3], WildDash [17] and KITTI [1], involved in the experiments. Here we provide detailed information of involved datasets.

GTA5. GTA5 [12] is a synthetic urban scene image dataset, rendered from game engine. The scene of the GTA5 images is based on the city of Los Angeles. The GTA5 dataset covers 24966 densely labeled images, the annotation of which is compatible with that of Cityscapes. In OCDA benchmark, GTA5 $\rightarrow$ BDD100K, the GTA5 images, with the ground truth label, serve as source domain.

SYNTHIA-SF. SYNTHIA-SF [4] is a synthetically rendered image dataset from virtual city. There are 2224 images in the SYNTHIA-SF dataset, featuring different scenarios and traffic conditions. The images are densely labeled and the labels are compatible with Cityscapes. In our OCDA benchmark, SYNTHIA-SF $\rightarrow$ BDD100K, the SYNTHIA-SF dataset and the associated ground truth label serve as the source domain.

BDD100K. BDD100K [16] is a real urban scene image dataset, mainly taken from US cities. And the images in BDD100K dataset are diverse in different aspects such as weather and environment. We adopt the C-driving subset of BDD100K proposed in [8], which is composed of rainy, snowy, cloudy and overcast images. During training stage, 14697 images, without the ground truth label, are used as the unlabeled compound target domain, including rainy, snowy and cloudy weather images. All different weather images are mixed and not assigned the weather information. During the testing stage, 803 images covering rainy, snowy and cloudy weather, with ground truth semantic annotation, are used as the validation set of the compound target domain, for evaluating the adaptation performance of the model. Besides, during the testing stage, 627 images with the ground truth semantic label, containing overcast weather, are taken as the validation set of the open domain, for evaluating the generalization performance of the model. The semantic label of the BDD100K dataset is compatible with that of Cityscapes.

Cityscapes. Cityscapes [3] is a real street scene image dataset, collected from different European cities. In our OCDA benchmark, during the testing stage, the validation set of Cityscapes, covering 500 densely labeled images, is used as one of the extended open domains to evaluate the generalization ability of the model.

KITTI. KITTI [1] covers the real urban scene images, taken from the mid-size European city, Karlsruhe. In our OCDA benchmark, the validation set of KITTI, including 200 densely labeled images, is used as one of the extended open domains for generalization ability evaluation during the testing stage. The ground truth label of KITTI dataset is compatible with that of Cityscapes.

WildDash. WildDash [17] is a dataset covering images from diverse driving scenarios under the real-world conditions. The images in WildDash possess the diversity in different aspects, such as the time, weather, data sources and camera characteristics. In our OCDA benchmark, during the testing stage, the validation set of WildDash, containing 70 Cityscapes annotation compatible images, serves as one of the extended open domains for measuring the generalization performance of the model.

S3. Additional Experimental Results

In Sec. 4 of the main paper, we provide the quantitative and qualitative experimental results of our MOCDA model on the OCDA benchmark. Here we provide the detailed quantitative experimental results, and additional qualitative experimental results.

Quantitative results. In Table 1 and Table 4 of the main paper, the quantitative experimental results of our MOCDA model are reported on the mean IoU, for the OCDA task. Correspondingly, in Table S1 and Table S2, the more detailed per-class IoU results, on the compound target domain and the open domain, are shown. Additionally, the quantitative experimental results on different weather images are reported in Table S3. The detailed quantitative experimental results further verify the effectiveness of our MOCDA model for the OCDA task, on both of the compound target domain and the open domain.

Qualitative results. In Fig. 4 of the main paper, we show the qualitative experimental results of our MOCDA model for the OCDA task, on the compound target domain, the open domain and the extended open domains. In Fig. S1, we show more qualitative comparison between our MOCDA model and other methods, on the compound target domain (rainy, snowy and cloudy images), and the open domain (overcast images). It further proves the validity of our MOCDA model for the OCDA task, on both of the compound target domain and the open domain. In Fig. S2, we provide additional qualitative comparison between our MOCDA model with or without online update, on the extended open domains. As shown in Fig. S2, the online update introduces obvious benefit for improving the generalization of the MOCDA model to the extended open domains.

Online Update v.s. Transductive Learning. The online update during testing stage is different from the traditional transductive learning setting. In transductive setting, the testing set is provided as unlabeled training data during the training phase, while our online update performs one gradient step *_after_* predicting a mini-batch of the open domain and extended open domain samples during testing stage. Such update is performed in an unsupervised manner. The ability of online update during testing stage is the advantage of MAML on fast adaptation. For fair comparison, on the benchmark GTA $\rightarrow$ BDD100K, we also conducted an experiment by using online update for the AdaptSegNet [14] during the testing phase (*i.e.,* performing one gradient step adversarial training), and the results are 16.0%, 21.1% and 16.4% on three extended open domains, which is much lower than the AdaptSegNet performance without online update, 22.0%, 23.4% and 17.5%.

GTA5→BDD100K																				
Domain	Method	road	sidewalk	building	wall	fence	pole	traffic light	traffic sign	vegetation	terrian	sky	person	rider	car	truck	bus	train	motorbike	mIoU
Target	Source*	32.1	12.4	47.1	3.9	22.6	17.6	9.9	4.7	52.0	13.9	74.6	24.3	0.0	38.0	10.0	10.4	0.0	0.0	19.7
	AdaptSegNet*[14]	46.9	14.0	60.2	5.9	20.4	**18.3**	9.0	4.6	48.9	14.1	78.2	24.6	0.0	48.7	13.1	16.5	0.0	0.0	22.3
	Ours(Split)	71.6	13.4	**63.7**	8.2	19.9	18.2	6.8	5.6	57.3	16.5	80.9	22.7	0.0	57.4	18.7	21.2	0.0	0.0	25.4
	Ours (Fuse)	**73.9**	**20.6**	58.2	**8.5**	**22.8**	17.9	**10.4**	**7.1**	**61.9**	**20.1**	**84.8**	**26.1**	**2.3**	**61.3**	**19.8**	**26.4**	0.0	**3.7**	**27.7**
Open†	Source*	28.7	20.3	50.3	6.3	25.1	20.6	8.7	12.3	62.0	20.3	79.4	33.4	4.6	38.8	10.4	7.0	0.0	0.3	22.5
	AdaptSegNet*[14]	58.7	22.9	64.1	10.4	24.0	**21.8**	8.1	10.8	62.8	22.4	84.9	**35.5**	8.8	53.2	15.5	10.1	0.5	0.5	27.1
	Ours (Split)	76.5	22.0	**68.6**	**15.8**	22.6	21.6	6.0	6.8	64.8	24.3	86.6	35.2	8.1	63.1	**26.1**	11.7	0.1	0.0	29.5
	Ours (Fuse)	**80.1**	**28.6**	66.0	13.0	**26.6**	20.9	**8.9**	**15.5**	**67.0**	**25.1**	**87.7**	33.2	**9.5**	**69.2**	23.0	**18.3**	**2.2**	**2.0**	**31.4**

Table S1: Per-Class IoU on the compound target domain and open domain of the OCDA benchmark: GTA5 → BDD100K. ∗ represents our reproduced result of the experiments in [8]. The results are reported over 19 classes. The 'bicycle' class is not listed due to the result is close to zero. The best results are denoted in bold. Open † is open domain covering the BDD100K overcast images.

SYNTHIA-SF→BDD100K													
Domain	Method	road	sidewalk	building	wall	fence	pole	traffic light	vegetation	sky	person	car	mIoU
Target	Source	3.1	6.8	42.7	0.0	0.0	10.2	1.1	**39.6**	69.2	9.7	28.2	19.2
	MinEnt[15]	**67.2**	1.8	50.7	0.0	0.0	4.4	1.3	11.7	71.8	8.7	45.7	23.9
	AdaptSegNet[14]	63.1	11.9	46.5	0.1	0.0	10.5	3.1	22.2	**78.7**	17.8	54.1	28.0
	Ours(Split)	59.8	15.5	**52.8**	0.2	0.0	**13.6**	2.3	28.4	73.3	19.2	55.1	29.1
	Ours (Fuse)	61.3	**17.3**	49.7	**1.0**	**0.1**	11.1	**5.9**	37.5	72.6	**21.5**	**56.3**	**30.4**
Open†	Source	1.9	9.0	43.4	0.0	0.0	11.1	1.2	**45.1**	74.7	13.0	27.2	20.6
	MinEnt[15]	68.9	2.5	51.6	0.0	0.0	5.7	1.4	14.2	77.2	11.7	49.3	25.7
	AdaptSegNet[14]	**69.4**	14.4	48.7	0.0	0.0	11.8	2.3	23.0	**82.4**	21.7	59.0	30.3
	Ours (Split)	65.3	22.4	**54.6**	0.2	0.0	**15.1**	2.0	29.3	78.7	24.0	57.8	31.8
	Ours (Fuse)	65.5	**24.7**	50.0	**1.0**	**0.2**	12.0	**5.3**	36.7	76.2	**26.6**	**60.7**	**32.6**

Table S2: Per-Class IoU on the compound target domain and open domain of the OCDA benchmark: SYNTHIA-SF → BDD100K. The results are reported over 11 classes. The best results are denoted in bold. Open † is open domain covering the BDD100K overcast images.

S4. Additional Visualization

Hypernetwork prediction. In Sec. 4.2 of the main paper, we use the ablation study and the variants of our model to prove the validity of the hypernetwork in our MOCDA model. Here we provide additional t-SNE [9] visualization of our hypernetwork prediction. As shown in Fig. S3, for the image samples from different sub-target domains, our hypernetwork prediction possesses different feature attributes, even though we do not explicitly provide the sub-target domain information in this process. It proves that our hypernetwork is able to adaptively adjust the prediction, conditioned on the style code of the image samples.

Style code. In Sec. 4 of the main paper, besides the open domain from BDD100K dataset adopted by [8], we introduce the extended open domains, which have much larger domain gap to the compound target domain than the open domain from BDD100K dataset, to further measure the generalization ability of the model trained for OCDA task. Here we provide the style code t-SNE [9] visualization of the compound target domain, the open domain and the extended open domains. As shown in Fig. S4, it can be observed that the domain gap between the open domain and the compound target domain from BDD100K dataset is narrow due to the similar style. Instead, our introduced extended open domains, Cityscapes, KITTI and WildDash dataset, have much larger domain gap from the compound target domain. And the style code extracted by our MOCDA model can effectively reflect the domain gap. It demonstrates the effectiveness of the style code extracted in our MOCDA model, and proves the rationality of our introduced extended open domains for further evaluating the generalization performance of the model to the unseen domains.

References

[1] Hassan Abu Alhaija, Siva Karthik Mustikovela, Lars Mescheder, Andreas Geiger, and Carsten Rother. Augmented reality meets computer vision: Efficient data genera-

GTA5→BDD100K																				
Weather	Method	road	sidewalk	building	wall	fence	pole	traffic light	traffic sign	vegetation	terrian	sky	person	rider	car	truck	bus	train	motorbike	mIoU
Rainy	Source[8]	48.3	3.4	39.7	0.6	12.2	10.1	**5.6**	5.1	44.3	17.4	65.4	12.1	0.4	34.5	7.2	0.1	0.0	0.5	16.2
	AdaptSegNet[14, 8]	58.6	17.8	46.4	2.1	**19.6**	15.6	5.0	7.7	55.6	**20.7**	65.9	17.3	0.0	41.3	7.4	3.1	0.0	0.0	20.2
	CBST[18, 8]	59.4	13.2	47.2	2.4	12.1	14.1	3.5	8.6	53.8	13.1	**80.3**	13.7	**17.2**	**49.9**	8.9	0.0	0.0	**6.6**	21.3
	IBN-Net[10, 8]	58.1	19.5	51.0	4.3	16.9	**18.8**	4.6	**9.2**	44.5	11.0	69.9	20.0	0.0	39.9	8.4	15.3	0.0	0.0	20.6
	OCDA[8]	63.0	15.4	**54.2**	2.5	16.1	16.0	**5.6**	5.2	54.1	14.9	75.2	18.5	0.0	43.2	9.4	24.6	0.0	0.0	22.0
	Ours	**66.8**	**22.0**	52.4	**6.7**	16.7	16.9	5.3	3.5	**60.4**	17.2	80.1	**21.8**	0.1	46.4	**17.9**	**29.4**	0.0	0.0	**24.4**
Snowy	Source[8]	50.8	4.7	45.1	5.9	**24.0**	8.5	10.8	8.7	35.9	9.4	60.5	17.3	0.0	47.7	9.7	3.2	0.0	0.7	18.0
	AdaptSegNet[14, 8]	59.9	13.3	52.7	3.4	15.9	14.2	12.2	7.2	51.0	**10.8**	72.3	21.9	0.0	55.0	11.3	1.7	0.0	0.0	21.2
	CBST[18, 8]	59.6	11.8	57.2	2.5	19.3	13.3	7.0	**9.6**	41.9	7.3	70.5	18.5	0.0	61.7	8.7	1.8	0.0	0.2	20.6
	IBN-Net[10, 8]	61.3	13.5	57.6	3.3	14.8	**17.7**	10.9	6.8	39.0	6.9	71.6	22.6	0.0	56.1	13.8	20.4	0.0	0.0	21.9
	OCDA[8]	68.0	10.9	61.0	2.3	23.4	15.8	12.3	6.9	48.1	9.9	74.3	19.5	0.0	58.7	10.0	13.8	0.0	0.1	22.9
	Ours	**71.8**	**16.9**	**61.1**	**6.5**	21.4	16.3	**17.0**	7.5	**52.9**	8.7	**79.7**	**29.2**	**0.5**	**62.7**	**18.9**	**29.4**	0.0	**22.6**	**27.5**
Cloudy	Source[8]	47.0	8.8	33.6	4.5	20.6	11.4	**13.5**	8.8	55.4	25.2	78.9	20.3	0.0	53.3	10.7	4.6	0.0	0.0	20.9
	AdaptSegNet[14, 8]	51.8	15.7	46.0	5.4	25.8	18.0	12.0	6.4	64.4	26.4	82.9	24.9	0.0	58.4	10.5	4.4	0.0	0.0	23.8
	CBST[18, 8]	56.8	21.5	45.9	5.7	19.5	17.2	10.3	8.6	62.2	24.3	**89.4**	20.0	0.0	58.0	14.6	0.1	0.0	0.1	23.9
	IBN-Net[10, 8]	60.8	18.1	50.5	8.2	25.6	**20.4**	12.0	**11.3**	59.3	24.7	84.8	24.1	**12.1**	59.3	13.7	9.0	0.0	1.2	26.1
	OCDA[8]	69.3	20.1	55.3	7.3	24.2	18.3	12.0	7.9	64.2	**27.4**	88.2	24.7	0.0	62.8	13.6	18.2	0.0	0.0	27.0
	Ours	**79.6**	**21.7**	**61.4**	**11.0**	**27.6**	19.4	13.4	8.3	**69.0**	26.4	89.1	**25.0**	3.2	**69.5**	**22.7**	**21.5**	0.0	**3.5**	**30.1**
Overcast	Source[8]	46.6	9.5	38.5	2.7	19.8	12.9	**9.2**	17.5	52.7	19.9	76.8	20.9	1.4	53.8	10.8	8.4	0.0	1.8	21.2
	AdaptSegNet[14, 8]	59.5	24.0	49.4	6.3	23.3	19.8	8.0	14.4	61.5	22.9	74.8	29.9	0.3	59.8	12.8	9.7	0.0	0.0	25.1
	CBST[18, 8]	58.9	26.8	51.6	6.5	17.8	17.9	5.9	**17.9**	60.9	21.7	**87.9**	22.9	0.0	59.9	11.0	2.1	0.0	0.2	24.7
	IBN-Net[10, 8]	62.9	25.3	55.5	6.5	21.2	**22.3**	7.2	15.3	53.3	16.5	81.6	31.1	2.4	59.1	10.3	14.2	0.0	0.0	25.5
	OCDA[8]	73.5	26.5	62.5	8.6	24.2	20.2	8.5	15.2	61.2	23.0	86.3	27.3	0.0	64.4	14.3	13.3	0.0	0.0	27.9
	Ours	**80.1**	**28.6**	**66.0**	**13.0**	**26.6**	20.9	8.9	15.5	**67.0**	**25.1**	87.7	**33.2**	**9.5**	**69.2**	**23.0**	**18.3**	**2.2**	**2.0**	**31.4**

Table S3: Per-Class IoU on different weather images of the OCDA benchmark: GTA5 → BDD100K. The rainy, snowy and cloudy weather compose the compound target domain, while the overcast weather is the open domain. The results are reported over 19 classes. The 'bicycle' class is not listed due to the result is close to zero. The best results are denoted in bold.

tion for urban driving scenes. *IJCV*, 126(9):961–972, 2018. 1, 2

[2] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *TPAMI*, 40(4):834–848, 2017. 1

[3] 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 *CVPR*, 2016. 1, 2

[4] Daniel Hernandez-Juarez, Lukas Schneider, Antonio Espinosa, David Vazquez, Antonio M. Lopez, Uwe Franke, Marc Pollefeys, and Juan Carlos Moure. Slanted stixels: Representing san francisco's steepest streets. In *BMVC*, 2017. 1, 2

[5] Xun Huang, Ming-Yu Liu, Serge Belongie, and Jan Kautz. Multimodal unsupervised image-to-image translation. In *ECCV*, 2018. 1

[6] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *ICML*, 2015. 1

[7] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *ICLR*, 2015. 1

[8] Ziwei Liu, Zhongqi Miao, Xingang Pan, Xiaohang Zhan, Dahua Lin, Stella X Yu, and Boqing Gong. Open compound domain adaptation. In *CVPR*, 2020. 2, 3, 4

[9] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *JMLR*, 9(Nov):2579–2605, 2008. 3

[10] Xingang Pan, Ping Luo, Jianping Shi, and Xiaoou Tang. Two at once: Enhancing learning and generalization capacities via ibn-net. In *ECCV*, 2018. 4

[11] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. In *NeurIPS*, 2019. 1

[12] Stephan R Richter, Vibhav Vineet, Stefan Roth, and Vladlen Koltun. Playing for data: Ground truth from computer games. In *ECCV*, 2016. 1

[13] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 1

[14] Y.-H. Tsai, W.-C. Hung, S. Schulter, K. Sohn, M.-H. Yang, and M. Chandraker. Learning to adapt structured output space for semantic segmentation. In *CVPR*, 2018. 1, 2, 3, 4

[15] Tuan-Hung Vu, Himalaya Jain, Maxime Bucher, Matthieu Cord, and Patrick Pérez. Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation. In *CVPR*, 2019. 3

[16] Fisher Yu, Haofeng Chen, Xin Wang, Wenqi Xian, Yingying Chen, Fangchen Liu, Vashisht Madhavan, and Trevor Darrell. Bdd100k: A diverse driving dataset for heterogeneous multitask learning. In *CVPR*, 2020. 1, 2

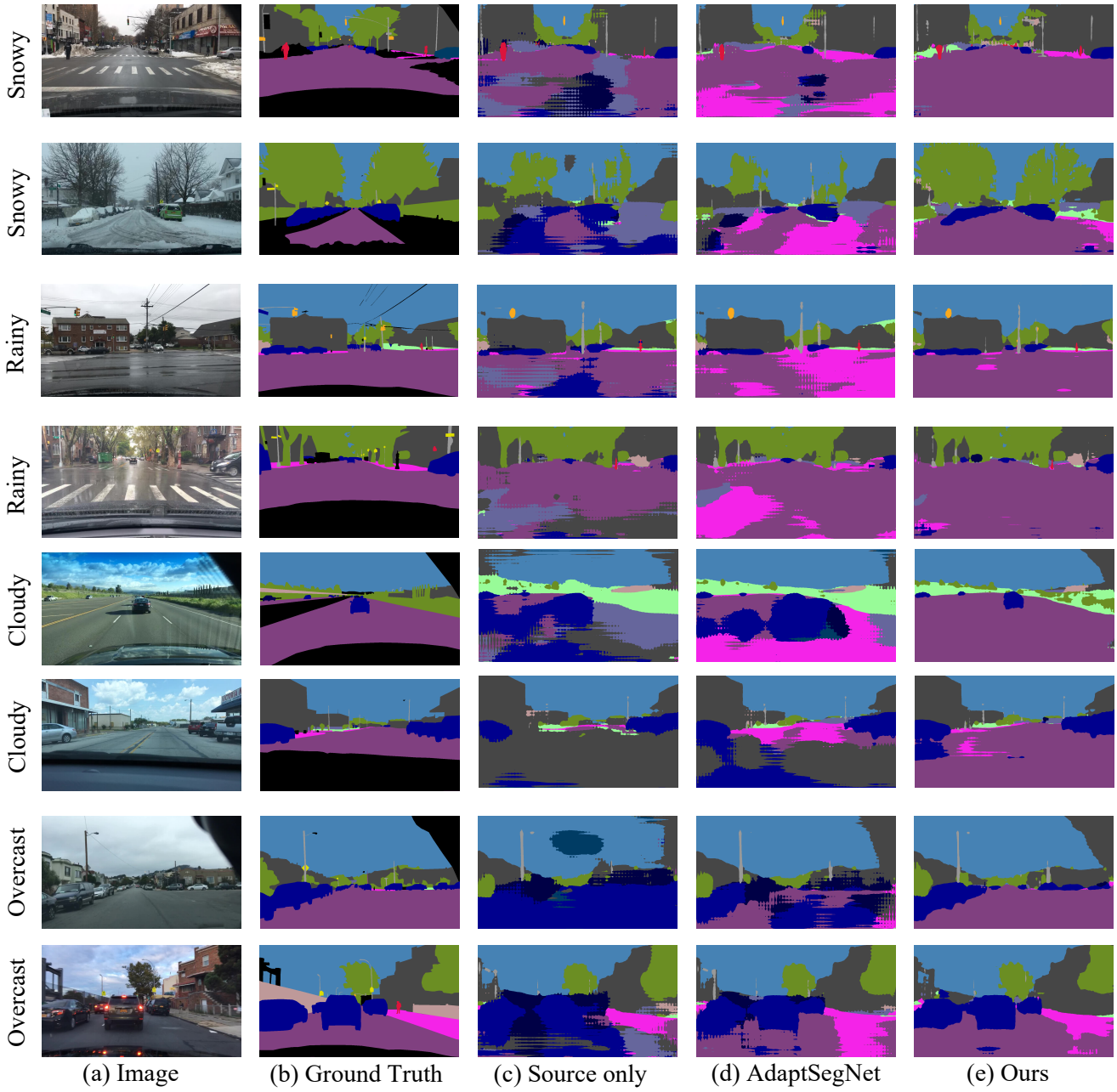

Figure S1: Qualitative semantic segmentation results of the OCDA benchmark: GTA→ BDD100K. The snowy, rainy and cloudy images are from the compound target domain, while the overcast image is from the open domain. It can be observed that our MOCDA model outperforms the source-only baseline and the AdaptSegNet method on both of the compound target domain and the open domain.

[17] Oliver Zendel, Katrin Honauer, Markus Murschitz, Daniel Steininger, and Gustavo Fernandez Dominguez. Wilddash-creating hazard-aware benchmarks. In *ECCV*, 2018. 1, 2

[18] Yang Zou, Zhiding Yu, BVK Vijaya Kumar, and Jinsong Wang. Unsupervised domain adaptation for semantic segmentation via class-balanced self-training. In *ECCV*, 2018.

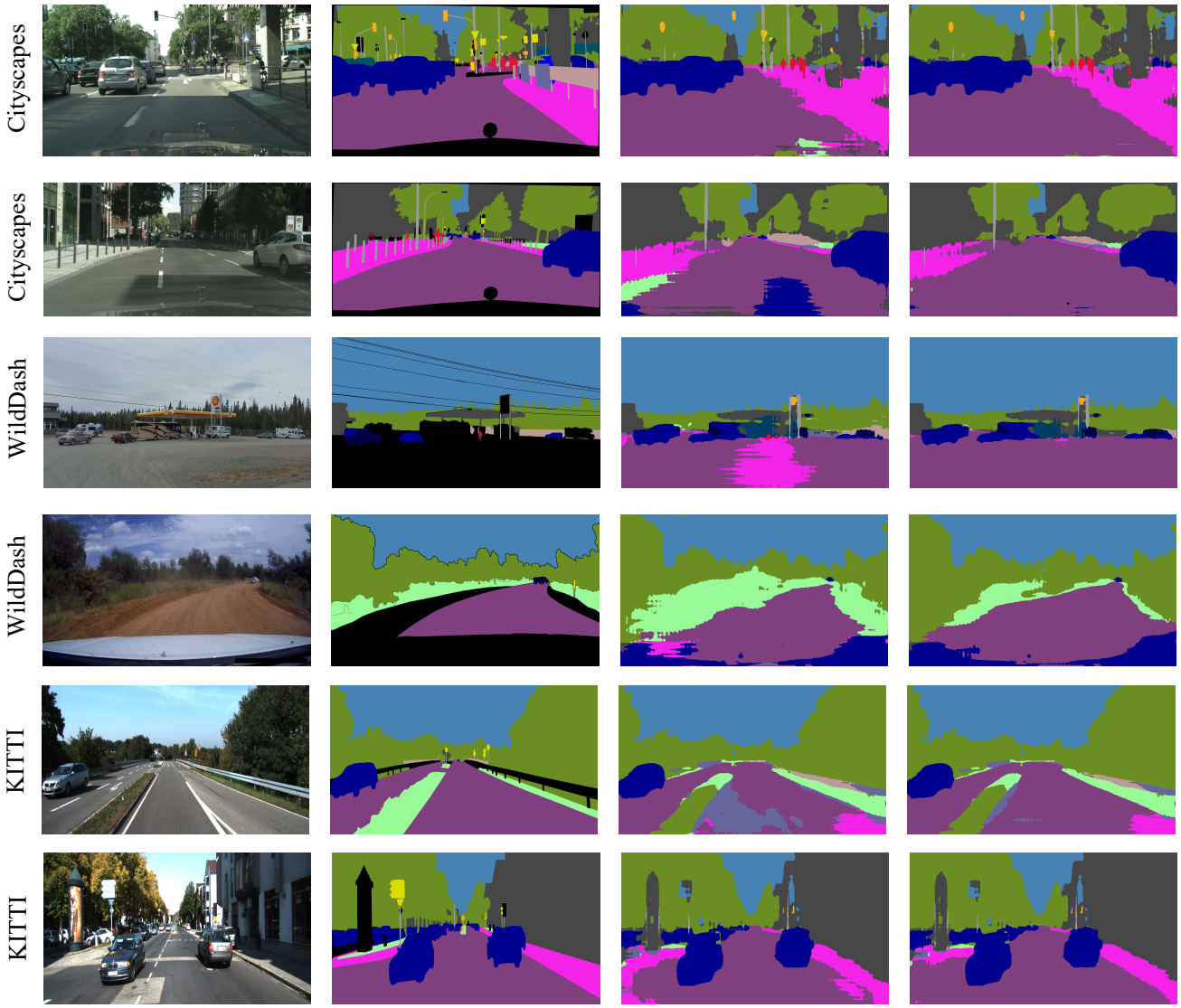

Figure S2: Qualitative semantic segmentation results on the extended open domains of the OCDA benchmark: GTA→ BDD100K. It is observed that the online update shows obvious benefit for the generalization to the extended open domains.

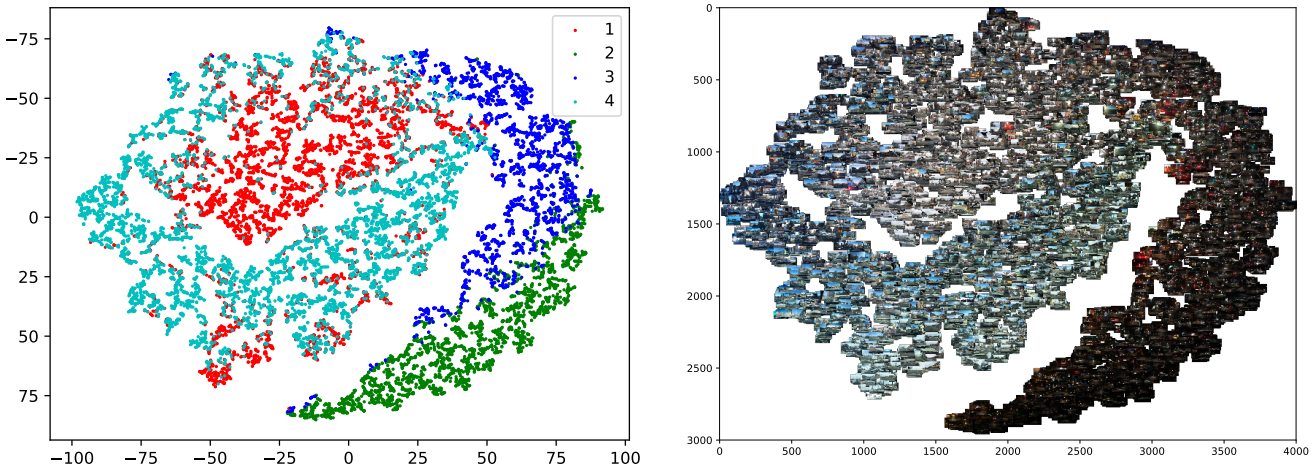

Figure S3: t-SNE visualization of hypernetwork prediction. For image samples belonging to different sub-target domains $1, 2, 3, 4$, our hypernetwork prediction shows different attributes even though we do not explicitly input the sub-target domain information during the fuse module training, which proves the validity of our hypernetwork.

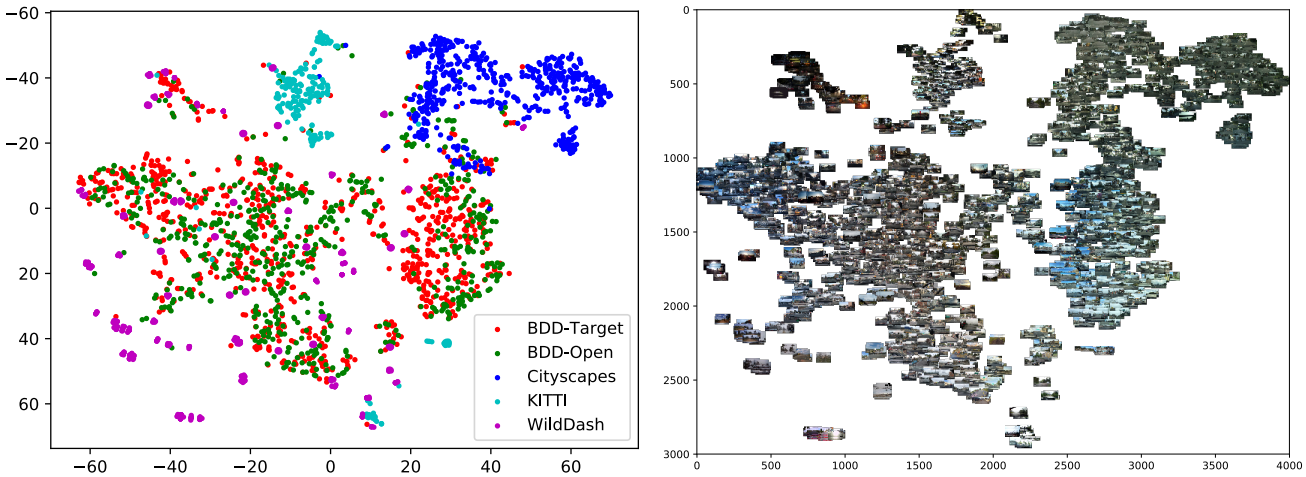

Figure S4: Extended open domains, open domain and target domain style code t-SNE visualization. The domain gap between the BDD100K open domain image and the target domain image (red and green points) is narrow due to the similar style. Our introduced extended open domain Cityscapes, KITTI and WildDash images have much larger domain gap from the BDD100K images. And the style code extracted by our cluster module can effectively reflect the domain gap.