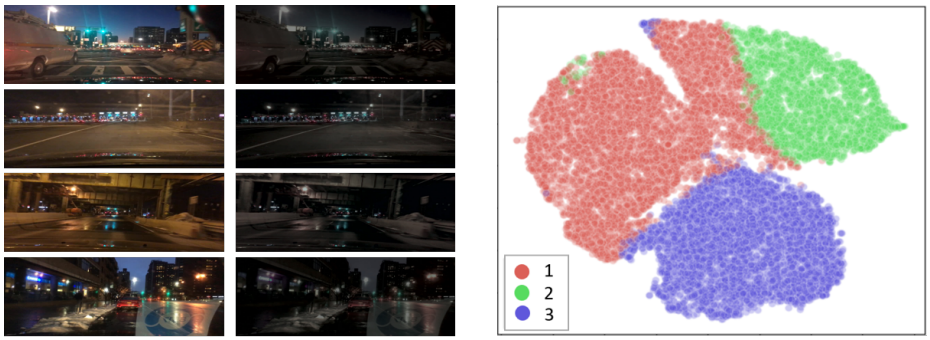
Supplemental Material: ML-BPM: Multi-teacher Learning with Bidirectional Photometric Mixing for Open Compound Domain Adaptation in Semantic Segmentation

Fei Pan[1], Sungsu Hur[1], Seokju Lee[2], Junsik Kim[3], and In So Kweon[1]

[1] KAIST, South Korea. {feipan, sshuh1215, iskweon77}@kaist.ac.kr
[2] KENTECH, South Korea. slee@kentech.ac.kr
[3] Harvard University, USA. mibastro@gmail.com

1 Subdomain Style Purification and the t-SNE Visualization



(a) The samples from subdomain 2 (b) The t-SNE visualization.

Fig. A1: (a) presents the noisy samples from Subdomain 2 of C-Driving dataset before subdomain style purification (before SSP) and after subdomain style purification (after SSP). (b) shows the t-SNE visualization of the concatenated histograms of the C-Driving dataset on LAB color space when $k = 3$.

As mentioned in Section 3.2, it is hard to guarantee that the images from the same target subdomain have the same style. In other words, small domain gaps might still results from the various image styles in each subdomain. We propose subdomain style purification to unify the styles of the target data that belongs to the same subdomain so that the domain gaps in these images could be

further reduced. We provide the visualization of the sample images transformed by subdomain style purification (SSP) from subdomain 2 in Figure A1 (a). Note that the images from *before SSP* in Figure A1(a) has the styles different from the *standard style*, and they are transformed into the standard style with the help of histogram matching on the LAB color space. We further set up $k = 3$ and present the t-SNE visualization of the concatenated histograms of the C-Driving images from LAB color space in Figure A1 (b).

The reason of subdomain style purification (SSP). With the help of automatic domain separation, the number of abnormal samples with different styles is small. Though these abnormal samples might be helpful for the model's generalization, they could also lead to a negative transfer, which further hinders the model from learning domain invariant features in a specific subdomain. With GTA5→C-Driving, we get a 0.5% of mIoU drop on average over all the subdomains without using SSP, as shown in Table 3(b).

2 ACDC Dataset

We also evaluate the proposed approach on another ACDC dataset[24]. ACDC dataset contains real-world images from the road scenes in diverse weather conditions, including fog, nighttime, rain and snow. We consider the $2,800$ images of fog, nighttime and rain from the training split of ACDC as the compound domain; the 400 snow images with pixel-wise annotations of ACDC training split are taken as the open domain. The final performance is evaluated on the validation set of ACDC, which contains 306 images with ground-truth maps.

We present the performance comparison of mean IoU in Table A1. For the compound target domain of ACDC (fog, nighttime, rain), we achieve 32.1% of mean IoU on GTA5→ACDC and 31.9% of mean IoU on SYNTHAI→ACDC, outperforming all the UDA and OCDA approaches in the list. We also evaluate the generalization of our approach compared with other works. After finishing the compound domain adaptation training, all the models are directly tested on the open domain of ACDC (snow). Note that the snow images have never been used in training before. Under the benchmark datasets GTA5→ACDC and SYNTHIA→ACDC, our approach shows 41.6% and 29.1% of mean IoU. This demonstrates that our approach has better generalization ability toward novel domains (snow).

3 The Practicability of Our Approach

Though we use the multi-teacher models for training, our approach still has strong practicability for the two following reasons: these teacher models are trained simultaneously; only a single student model from distillation is needed for inference. The size of the student model is not affected by the number of the subdomains. With the number of the subdomains k^*, the FLOPS and the number of parameters of our *multi-teacher's* model are 327.08×10^9 and $43.8 \times k^* \times 10^6$. After the adaptive knowledge distillation, the FLOPS and number of parameters

Table A1: The performance comparison of mean IoU on the compound target domain (fog, nighttime, and rain) and the open domain (fog) of ACDC. Our approach is compared with the state-of-the-art UDA and OCDA approaches on (a) GTA5→ACDC and (b) SYNTHIA→ACDC benchmark dataset with ResNet-101 as the backbone.

(a) GTA5→ACDC

		Compound																			Open
Method	Type	road	sidewalk	building	wall	fence	pole	light	sign	veg	terrain	sky	person	rider	car	truck	bus	train	mbike	bike	mIoU	mIoU
Source	-	43.6	2.5	46.2	5.2	0.1	30.3	15.3	16.3	56.9	0.0	71.5	16.3	13.7	51.4	0.0	15.1	0.0	1.4	4.2	20.5	27.1
CDAS [13]	OCDA	**53.2**	5.9	56.1	10.1	2.6	22.0	37.1	11.4	53.9	23.5	71.3	27.6	14.6	47.5	16.8	19.5	0.0	3.2	3.8	25.3	29.1
CSFU [8]	OCDA	47.0	4.1	53.0	13.9	1.0	23.2	41.2	18.8	55.8	23.2	72.1	31.5	10.8	69.1	**26.4**	27.8	0.2	1.7	2.6	27.6	30.5
SAC [2]	UDA	42.6	4.2	57.6	11.9	3.8	23.0	49.7	23.8	63.6	31.9	**76.0**	30.3	10.5	65.3	23.6	23.1	0.1	0.7	3.2	28.7	33.6
DACS [24]	UDA	48.9	**9.7**	54.5	16.8	5.7	22.7	42.0	22.9	61.3	29.7	73.7	**32.2**	11.6	63.3	23.2	26.5	0.0	1.2	**5.2**	29.0	34.8
DHA [19]	OCDA	49.8	5.2	**59.1**	10.2	3.1	25.6	47.8	**27.9**	65.1	32.0	75.2	29.0	12.2	61.5	20.5	**32.4**	0.0	1.0	2.0	29.5	37.5
Ours	OCDA	48.4	5.0	58.2	**25.3**	**10.0**	**35.1**	**50.4**	26.7	**66.8**	**33.3**	75.8	32.1	**16.7**	**73.5**	16.8	26.6	**0.2**	**3.9**	4.6	**32.1**	**41.6**

(b) SYNTHIA→ACDC

		Compound																	Open
Method	Type	road	sidewalk	building	wall	fence	pole	light	sign	veg	sky	person	rider	car	bus	mbike	bike	$mIoU^{16}$	$mIoU^{16}$
Source	-	45.2	0.2	36.7	1.7	0.6	25.7	4.0	5.6	46.6	64.3	16.9	11.3	39.6	16.5	0.6	1.9	19.8	20.5
CDAS[13]	OCDA	61.3	0.7	60.1	**11.7**	**1.8**	28.4	18.8	23.5	48.6	28.9	16.5	15.9	69.2	18.4	5.4	**5.6**	25.9	23.3
CSFU[8]	OCDA	62.6	0.3	60.3	8.6	**1.8**	21.3	20.7	29.1	44.5	22.1	**34.5**	19.0	71.1	**23.2**	4.4	4.3	26.7	24.8
SAC[2]	UDA	**69.8**	0.4	56.2	1.7	0.0	20.0	12.6	13.7	52.5	**78.1**	29.1	15.5	68.9	20.9	3.2	1.2	27.7	25.4
DACS[24]	UDA	55.6	1.1	55.7	0.1	0.7	25.8	**31.7**	18.3	65.5	53.7	31.1	16.6	69.2	22.5	2.9	3.1	28.3	27.0
DHA[19]	OCDA	55.5	1.1	57.2	0.7	0.8	26.6	22.7	24.6	**65.8**	58.4	29.6	**23.9**	70.8	19.5	5.4	4.2	29.2	27.3
Ours	OCDA	66.7	**1.7**	**62.4**	10.8	1.4	**30.8**	23.9	**29.2**	62.6	69.0	31.6	14.6	**71.8**	22.9	**6.8**	4.5	**31.9**	**29.1**

Table A2: The evaluation on GTA5→C-Driving.

(a) ImageNet pre-trained VGG-16 Backbone

Method	Compound (C)			Open (O)	Average	
	Rainy	Snowy	Cloudy	Overcast	C	C+O
CDAS [13]	23.8	25.3	29.1	31.0	26.1	27.3
CSFU [8]	24.5	27.5	30.1	31.4	27.7	29.4
DACS [24]	26.8	29.2	35.1	35.9	30.4	31.8
DHA [19]	27.1	30.4	35.5	36.1	32.0	32.3
Ours	34.5	35.8	39.9	40.1	**36.7**	**37.5**

(b) Mixing Algorithm Comparison

Algorithm	BPM (*Ours*)	ClassMix [15]	CutMix [31]	CowMix [7]
mIoU	**40.2**	39.1	37.6	37.4

of our *student* model is 327.08×10^9 and 43.8×10^6.

The VGG-16 backbone and different mixup algorithms. We use VGG-16 backbone network for evaluation. The experimental results on GTA5→C-Driving in Table A2(a) demonstrates the effectiveness of our approach against existing works with ImageNet pre-trained VGG-16 as the backbone. We provide the comparison to existing domain mixup algorithms in the same setting, including ClassMix [15], CutMix [31], and CowMix [7].

The online updating on the open domains. Our online updating is conducted on each sample from the open domain, thus it is still domain generalization at the testing stage. Our student model G_{sd} is trained through the adaptive distillation from all the subdomain's segmentation models $\{G_m\}_{m=1}^{k^*}$ (Eq. (10, 11)). Each G_m is optimized by Eq. (7) with the help of the mean teacher M_m, following the work of DACS[24]. We also used M_m instead of G_m for distillation but do not see significant performance gain.

The reason of using bidirectional mixing. Using the photometric transform Δ (Eq.(6)) on target-to-source mixing, we enforce the consistency of prediction between the target and the mixed image, which are taken as additional augmentation to improve the model's performance (Table.3(a,b)). With the experiment on GTA5→C-Driving, we get 40.1% of mIoU on using pseudo-labels of target data for ClassMix on target-to-source mixing, similar to ours 40.2% (Table 1(a)). Table A2 (b) shows that our BPM outperforms existing mixing algorithms.