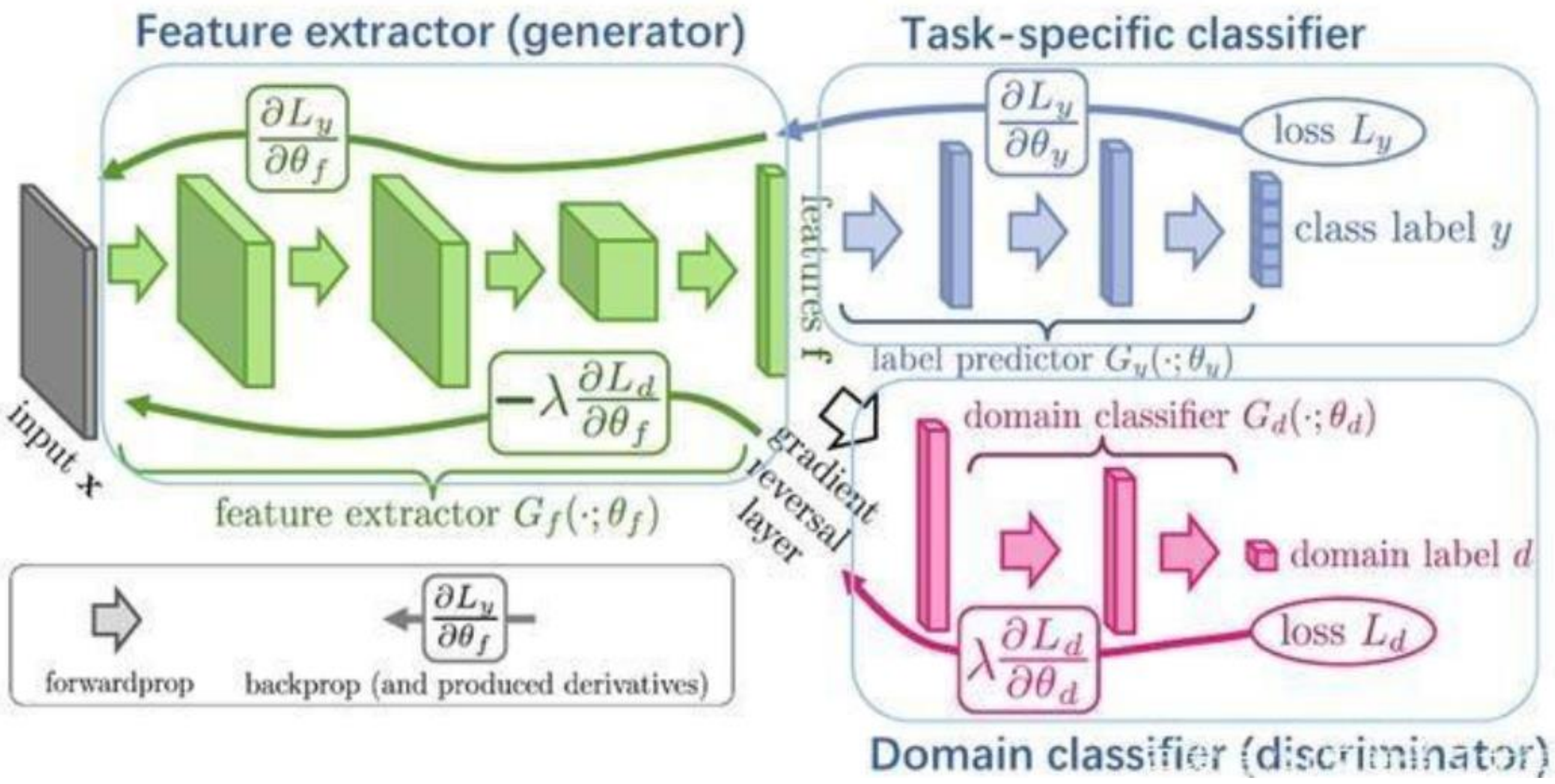
Maximum Classifier Discrepancy for Unsupervised Domain Adaptation

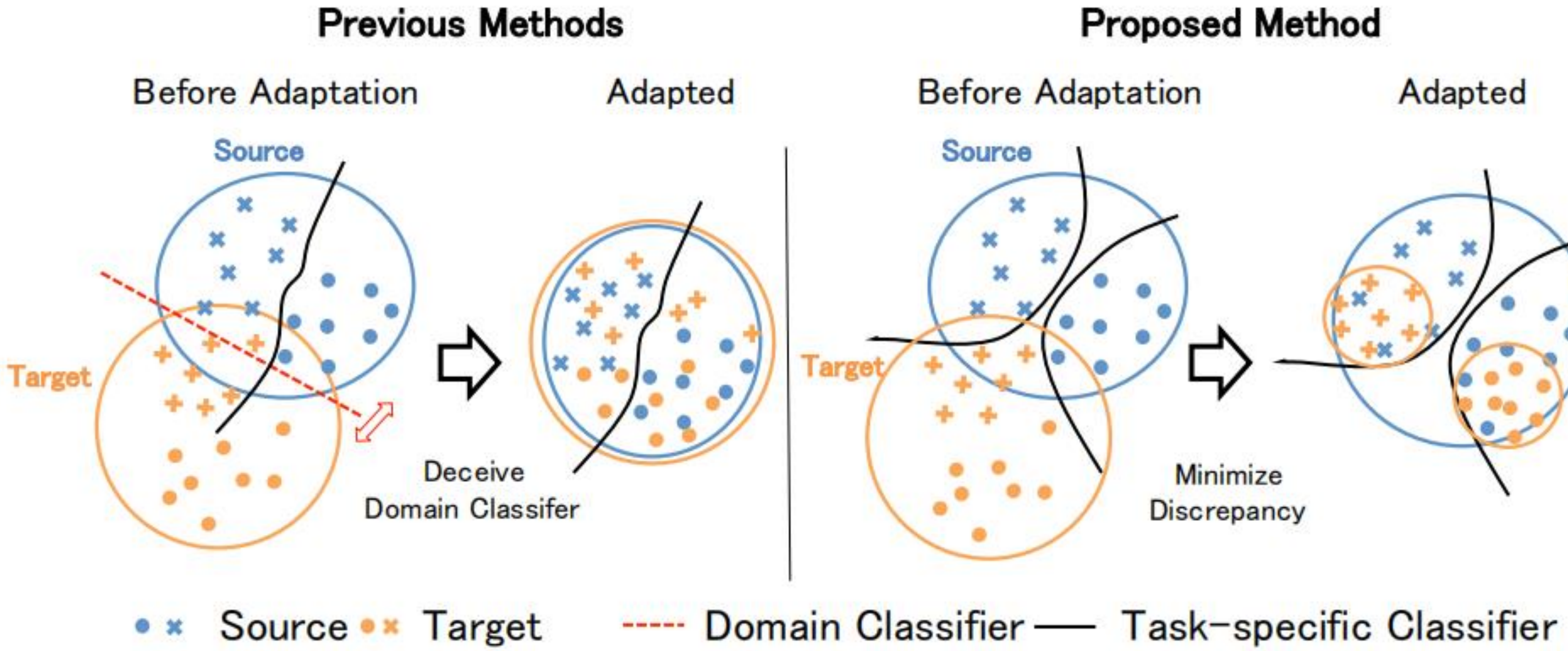
Kuniaki Saito Kohei Watanabe Yoshitaka Ushiku Tatsuya Harada

CVPR 2018

DA aims to train a classifier using source samples that generalize well to the target domain

However, each domain‘s samples have different characteristics, which makes the problem difficult to solve.


Feature extractor (generator)
Task-specific classifier
$\frac{\partial L_y}{\partial \theta_f}$
$\frac{\partial L_y}{\partial \theta_y}$
loss L_y
features f
class label y
label predictor $G_y(\cdot;\theta_y)$
$-\lambda \frac{\partial L_d}{\partial \theta_f}$
input x
feature extractor $G_f(\cdot;\theta_f)$
gradient reversal layer
domain classifier $G_d(\cdot;\theta_d)$
domain label d
loss L_d
$\lambda \frac{\partial L_d}{\partial \theta_d}$
$\frac{\partial L_y}{\partial \theta_f}$
forwardprop
backprop (and produced derivatives)
Domain classifier (discriminator)



Many UDA algorithms, particularly those for training neural networks, attempt to match the distribution of the source features with that of the target without considering the category of the sample

Methods

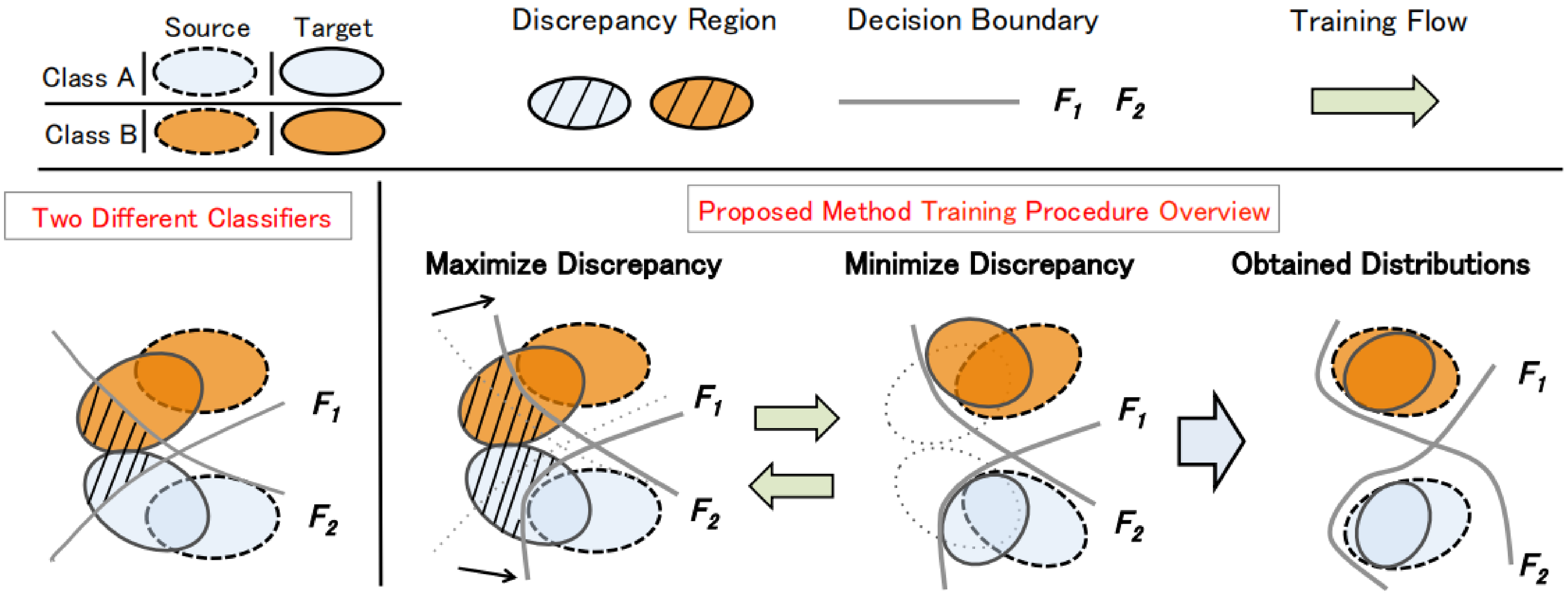


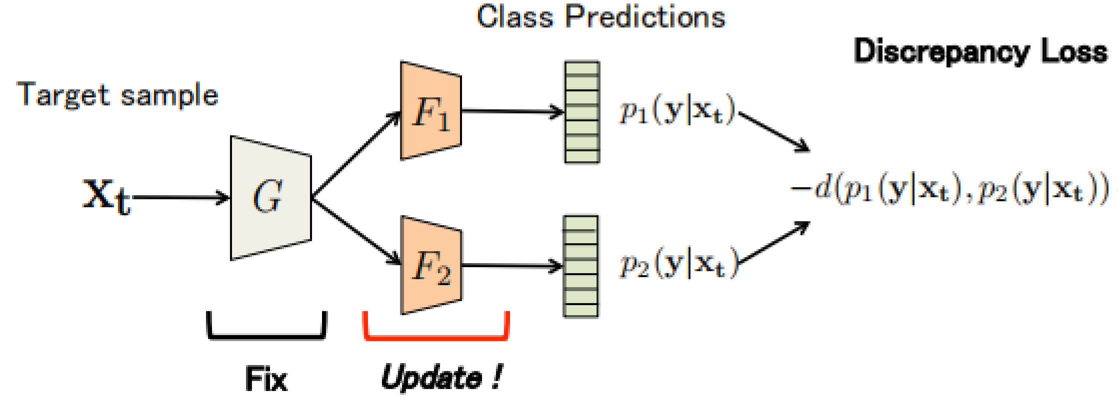
Figure 2. (Best viewed in color.) Example of two classifiers with an overview of the proposed method. Discrepancy refers to the disagreement between the predictions of two classifiers. First, we can see that the target samples outside the support of the source can be measured by two different classifiers (Leftmost, *Two different classifiers*). Second, regarding the training procedure, we solve a minimax problem in which we find two classifiers that *maximize* the discrepancy on the target sample, and then generate features that *minimize* this discrepancy.

Step A： train both classifiers and generator to classify the source samples correctly

$$\min_{G,F_1,F_2} \mathcal{L}(X_s, Y_s).$$

$$\mathcal{L}(X_s, Y_s) = -\mathbb{E}_{(\mathbf{x_s}, y_s)\sim(X_s, Y_s)} \sum_{k=1}^{K} \mathbb{1}_{[k=y_s]} \log p(\mathbf{y}|\mathbf{x}_s)$$

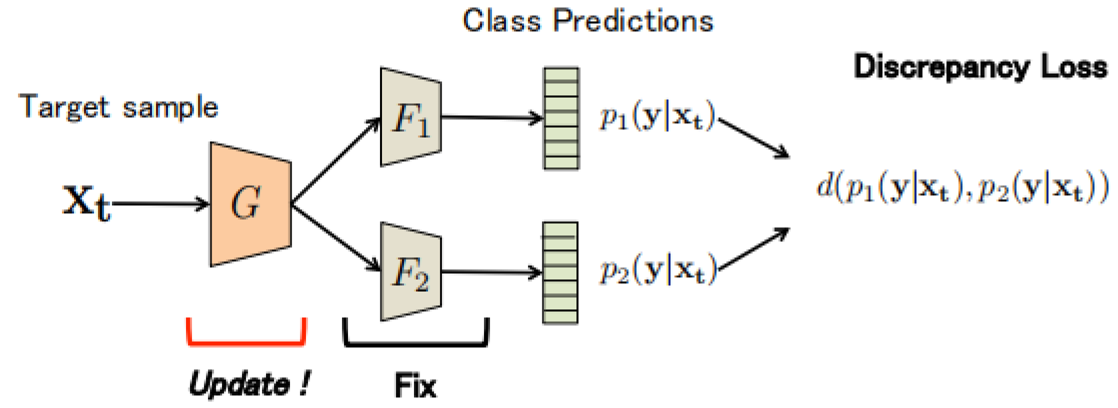
Step B : *Maximize* discrepancy on target (Fix G)



$$\min_{F_1, F_2} \mathcal{L}(X_s, Y_s) - \mathcal{L}_{\mathrm{adv}}(X_t).$$

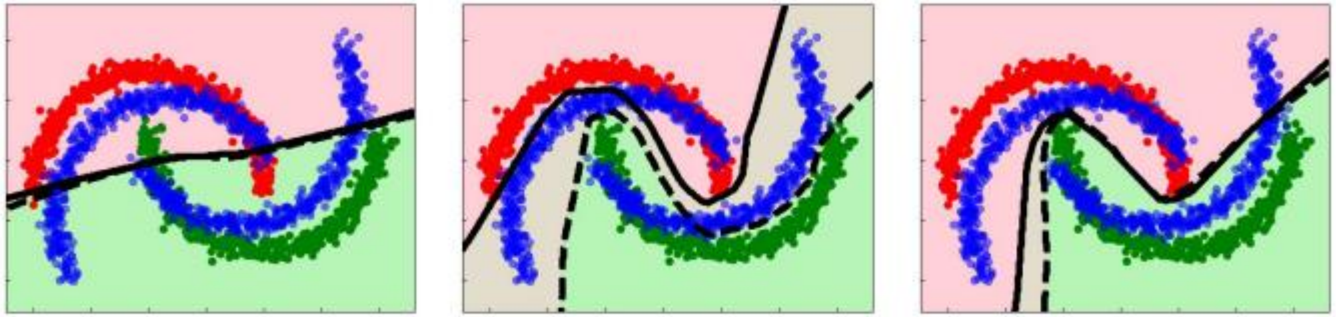
$$\mathcal{L}_{\mathrm{adv}}(X_t) = \mathbb{E}_{\mathbf{x_t} \sim X_t}[d(p_1(\mathbf{y}|\mathbf{x_t}), p_2(\mathbf{y}|\mathbf{x_t}))]$$

Step C : *Minimize* discrepancy on target (Fix F_1 , F_2)



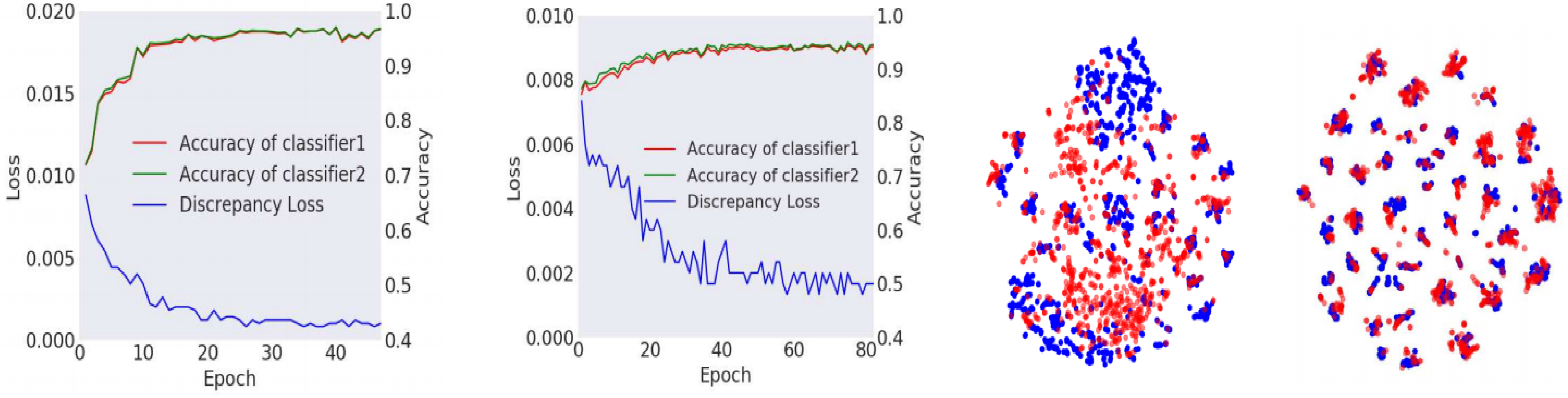
$$\min_{G} \mathcal{L}_{\mathrm{adv}}(X_t).$$

Expriments



(a) Source Only (b) No Step C (c) Proposed

Figure 4. (Best viewed in color.) Red and green points indicate the source samples of class 0 and 1, respectively. Blue points are target samples generated by rotating source samples. The dashed and normal lines are two decision boundaries in our method. The pink and light green regions are where the results of both classifiers are class 0 and 1, respectively. Fig. 4(a) is the model trained only on source samples. Fig. 4(b) is the model trained to increase discrepancy of the two classifiers on target samples without using Step C. Fig. 4(c) shows our proposed method.



(a) SVHN to MNIST

(b) SYN SIGN to GTSRB

(c) Source Only

(d) Adapted (Ours)

Figure 5. (Best viewed in color.) **Left**: Relationship between discrepancy loss (**blue** line) and accuracy (**red** and **green** lines) during training. As discrepancy loss decreased, accuracy improved. **Right**: Visualization of features obtained from last pooling layer of the generator in adaptation from SYN SIGNS to GTSRB using t-SNE [24]. **Red** and **blue** points indicate the target and source samples, respectively. All samples are testing samples. We can see that applying our method makes the target samples discriminative.

METHOD	**SVHN** to **MNIST**	**SYNSIG** to **GTSRB**	**MNIST** to **USPS**	**MNIST*** to **USPS***	**USPS** to **MNIST**
Source Only	67.1	85.1	76.7	79.4	63.4
Distribution Matching based Methods					
MMD † [21]	71.1	91.1	-	81.1	-
DANN † [7]	71.1	88.7	77.1±1.8	85.1	73.0±0.2
DSN † [4]	82.7	93.1	91.3	-	-
ADDA [39]	76.0±1.8	-	89.4±0.2	-	90.1±0.8
CoGAN [19]	-	-	91.2±0.8	-	89.1±0.8
PixelDA [3]	-	-	-	95.9	-
Ours ($n = 2$)	94.2±2.6	93.5±0.4	92.1±0.8	93.1±1.9	90.0±1.4
Ours ($n = 3$)	95.9±0.5	94.0±0.4	93.8±0.8	95.6±0.9	91.8±0.9
Ours ($n = 4$)	**96.2**±0.4	**94.4**±0.3	**94.2**±0.7	**96.5**±0.3	**94.1**±0.3
Other Methods					
ATDA † [32]	86.2	96.2	-	-	-
ASSC [11]	95.7±1.5	82.8±1.3	-	-	-
DRCN [9]	82.0±0.1	-	91.8±0.09	-	73.7±0.04

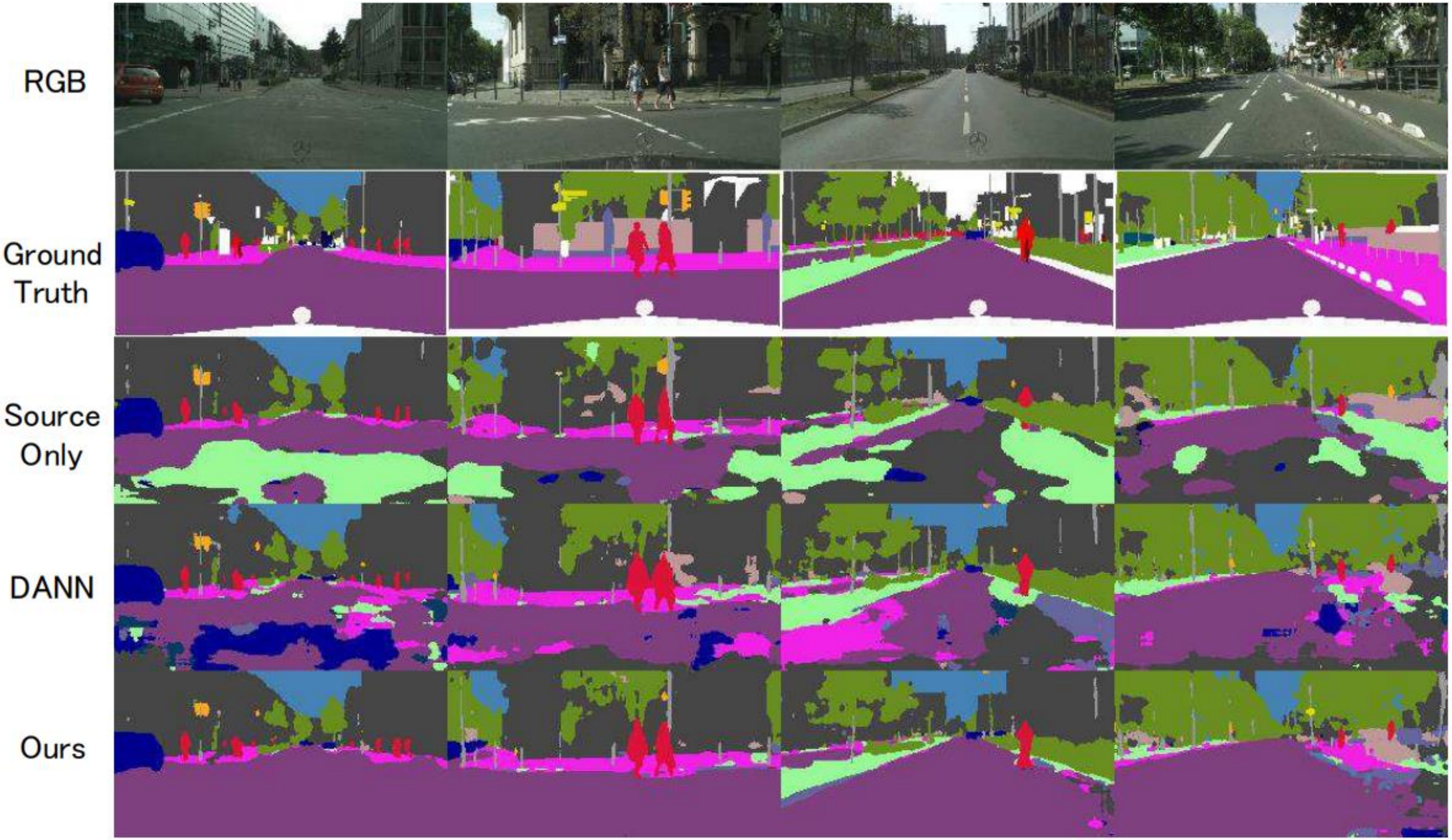


Figure 6. Qualitative results on adaptation from GTA5 to Cityscapes. DRN-105 is used to obtain these results.

Network	method	mIoU	road	sdwk	bldng	wall	fence	pole	light	sign	vgttn	trrn	sky	person	rider	car	truck	bus	train	mcycl	bcycl
VGG-16	Source Only	24.9	25.9	10.9	50.5	3.3	12.2	25.4	28.6	13.0	78.3	7.3	63.9	52.1	7.9	66.3	5.2	7.8	0.9	13.7	0.7
	FCN Wld [13]	27.1	70.4	32.4	62.1	14.9	5.4	10.9	14.2	2.7	79.2	21.3	64.6	44.1	4.2	70.4	8.0	7.3	0.0	3.5	0.0
	CDA (I) [42]	23.1	26.4	10.8	69.7	10.2	9.4	20.2	13.6	14.0	56.9	2.8	63.8	31.8	10.6	60.5	10.9	3.4	**10.9**	3.8	9.5
	Ours (k=2)	28.0	87.4	15.4	75.5	17.4	9.9	16.2	11.9	0.6	80.6	28.1	60.2	32.5	0.9	75.4	13.6	4.8	0.1	0.7	0.0
	Ours (k=3)	27.3	86.0	10.5	75.1	20.0	2.9	19.4	8.4	0.7	78.4	19.4	74.8	23.2	0.3	74.1	14.3	10.4	0.2	0.1	0.0
	Ours (k=4)	28.8	86.4	8.5	76.1	18.6	9.7	14.9	7.8	0.6	82.8	32.7	71.4	25.2	1.1	76.3	16.1	17.1	1.4	0.2	0.0
DRN-105	Source Only	22.2	36.4	14.2	67.4	16.4	12.0	20.1	8.7	0.7	69.8	13.3	56.9	37.0	0.4	53.6	10.6	3.2	0.2	0.9	0.0
	DANN [7]	32.8	64.3	23.2	73.4	11.3	**18.6**	**29.0**	**31.8**	14.9	82.0	16.8	73.2	53.9	12.4	53.3	20.4	11.0	5.0	18.7	9.8
	Ours (k=2)	**39.7**	90.3	31.0	78.5	19.7	17.3	28.6	30.9	**16.1**	83.7	30.0	69.1	**58.5**	**19.6**	81.5	23.8	**30.0**	5.7	**25.7**	**14.3**
	Ours (k=3)	38.9	**90.8**	**35.6**	**80.5**	22.9	15.5	27.5	24.9	15.1	84.2	31.8	77.4	54.6	17.2	82.0	21.6	29.0	1.3	21.8	5.3
	Ours (k=4)	38.1	89.2	23.2	80.2	**23.6**	18.1	**27.7**	25.0	9.3	**84.4**	**34.6**	**79.5**	53.2	16.0	**84.1**	**26.0**	22.5	5.2	16.7	4.8

Table 3. Adaptation results on the semantic segmentation. We evaluate adaptation from GTA5 to Cityscapes dataset.

Network	method	mIoU	road	sdwlk	bldng	wall	fence	pole	light	sign	vgttn	sky	prsn	ridr	car	bus	mcycl	bcycl
VGG-16	Source Only [42]	22.0	5.6	11.2	59.6	0.8	0.5	21.5	8.0	5.3	72.4	75.6	35.1	9.0	23.6	4.5	0.5	**18.0**
	FCN Wld [13]	20.2	11.5	19.6	30.8	4.4	0.0	20.3	0.1	**11.7**	42.3	68.7	51.2	3.8	54.0	3.2	0.2	0.6
	CDA (I+SP) [42]	29.0	65.2	26.1	74.9	0.1	**0.5**	10.7	3.7	3.0	76.1	70.6	47.1	8.2	43.2	20.7	0.7	13.1
DRN_105	Source Only	23.4	14.9	11.4	58.7	1.9	0.0	24.1	1.2	6.0	68.8	76.0	**54.3**	7.1	34.2	15.0	0.8	0.0
	DANN [7]	32.5	67.0	29.1	71.5	**14.3**	0.1	28.1	**12.6**	10.3	72.7	76.7	48.3	**12.7**	62.5	11.3	2.7	0.0
	Ours (k=2)	36.3	83.5	40.9	77.6	6.0	0.1	27.9	6.2	6.0	83.1	**83.5**	51.5	11.8	78.9	19.8	4.6	0.0
	Ours (k=3)	**37.3**	84.8	**43.6**	**79.0**	3.9	0.2	**29.1**	7.2	5.5	**83.8**	83.1	51.0	11.7	79.9	27.2	**6.2**	0.0
	Ours (k=4)	37.2	**88.1**	43.2	79.1	2.4	0.1	27.3	7.4	4.9	83.4	81.1	51.3	10.9	**82.1**	**29.0**	5.7	0.0

Table 4. Adaptation results on the semantic segmentation. We evaluate adaptation from Synthia to Cityscapes dataset.