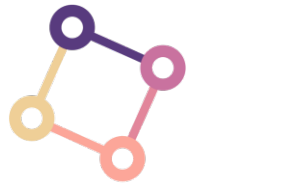
Maximum Classifier Discrepancy for Unsupervised Domain Adaptation

CVPR 2018

이 정 수



DAVIAN

Data and Visual Analytics Lab

Task

What is Domain Adaptation?

source domain

target domain

Source image (GTA5)

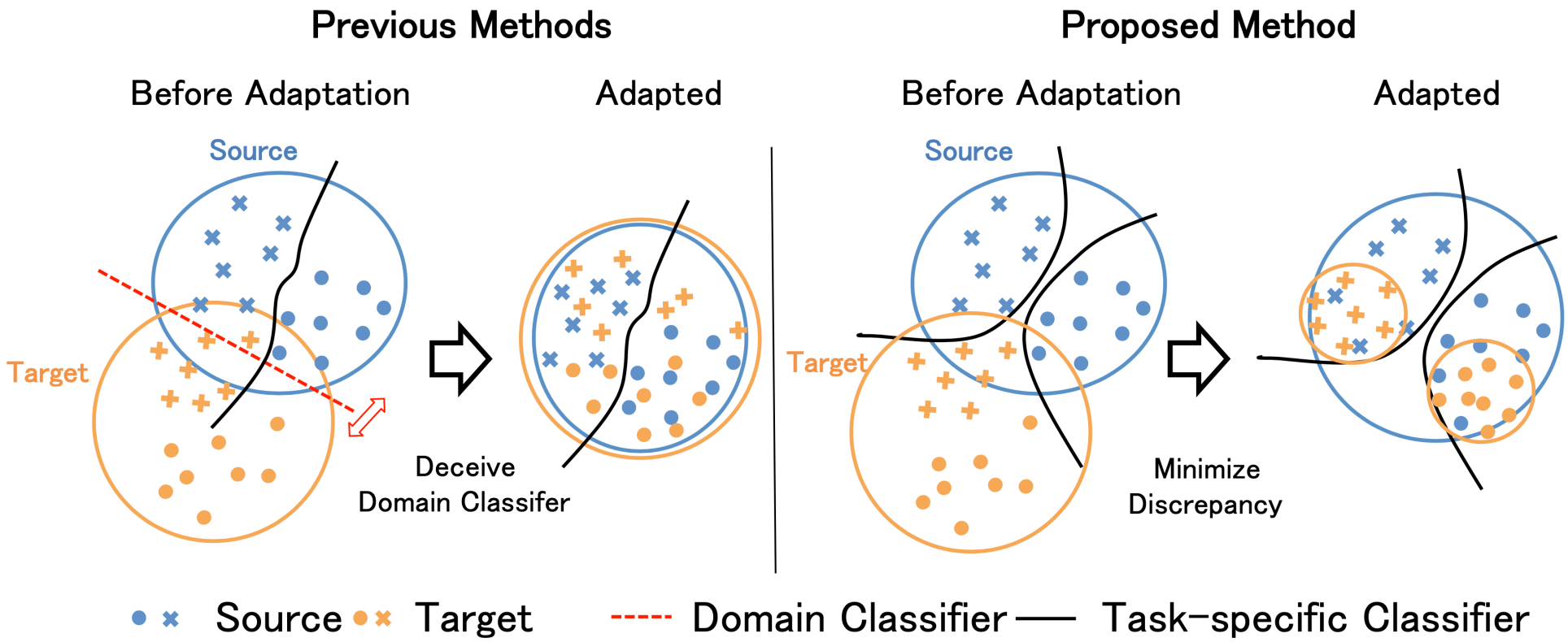
Target image (CityScapes)

Source images (SVHN)

Target images (MNIST)

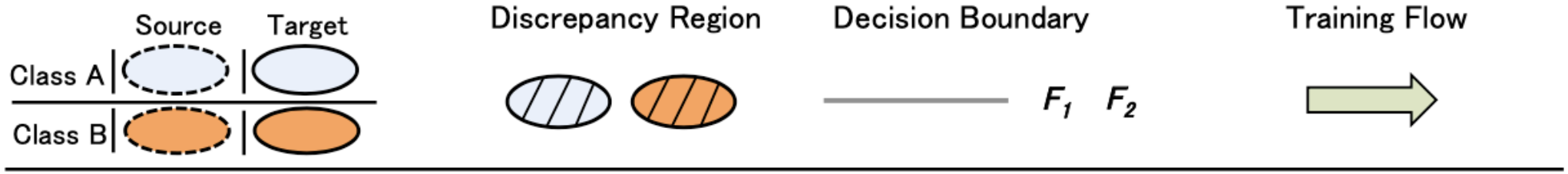
Train: source image, source label, target image
Test: predicted label of target image == target label?

Motivation

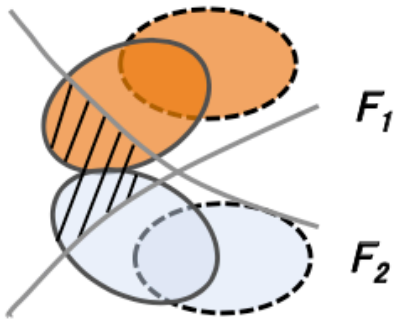


- Previous: only considers domain features, not class-specific features
- Proposed: also consider class-specific features for domain alignment

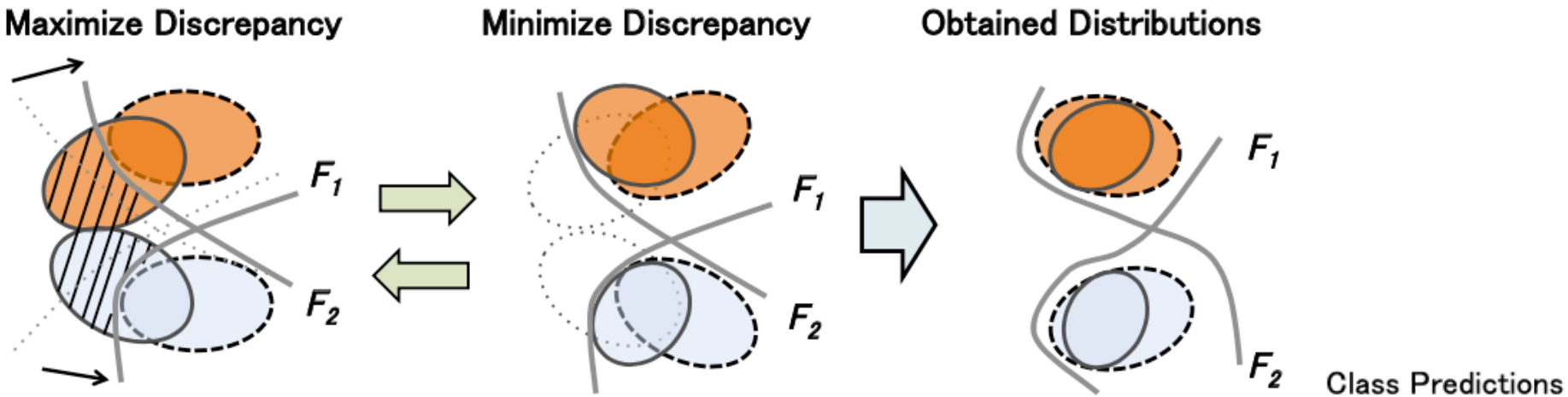
Method

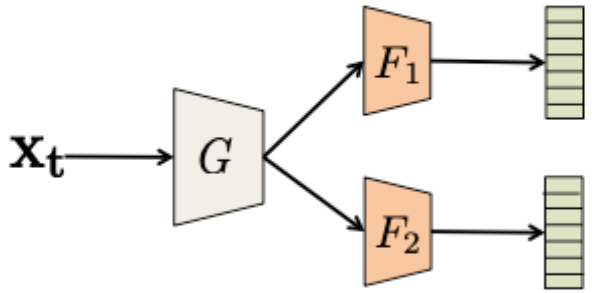


Two Different Classifiers



Proposed Method Training Procedure Overview

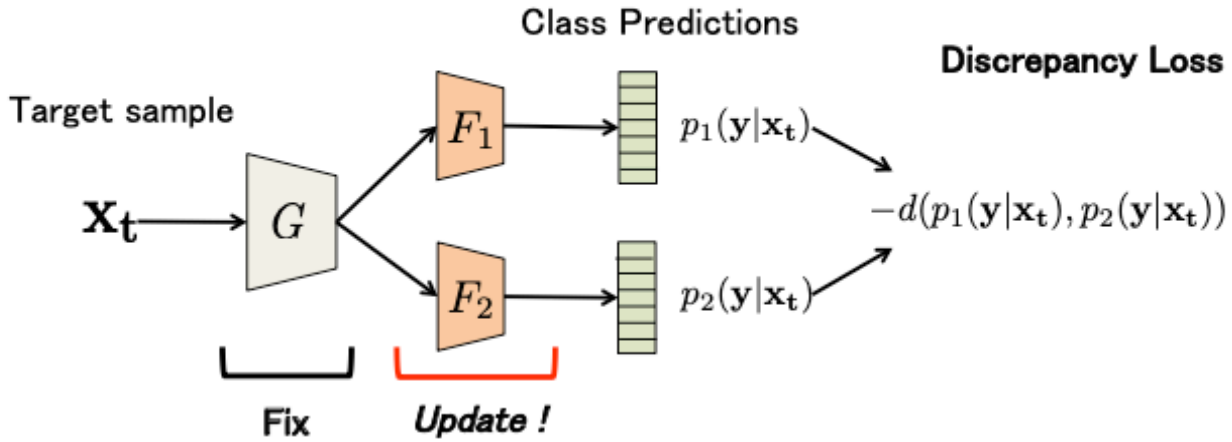


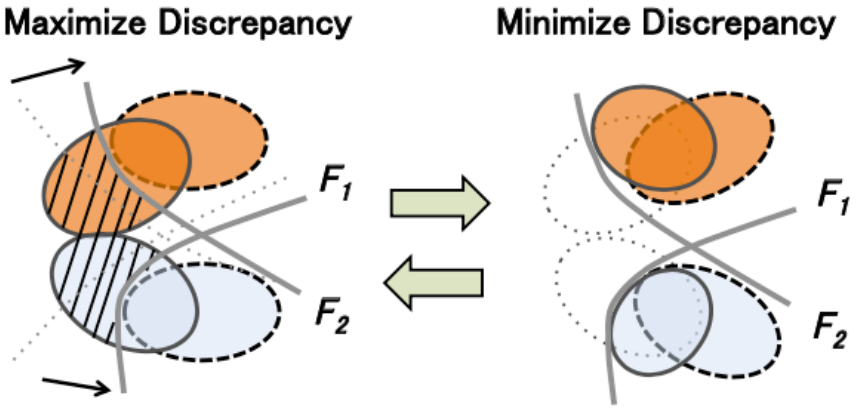


- Max discrepancy: ↑ # of target samples outside source domain representation
  - w/o: two classifiers similar & not detect target samples outside source domain
- Min discrepancy: representation of target samples -> source domain representation

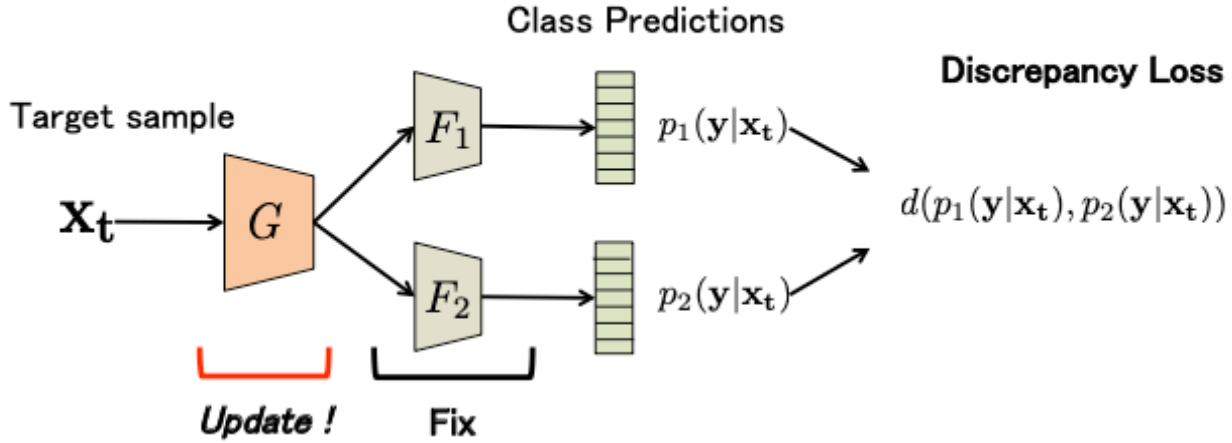
Method

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





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



$$d(p_1, p_2) = \frac{1}{K} \sum_{k=1}^{K} |p_{1k} - p_{2k}|$$

Method

Loss functions in 3 steps

Step A

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

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

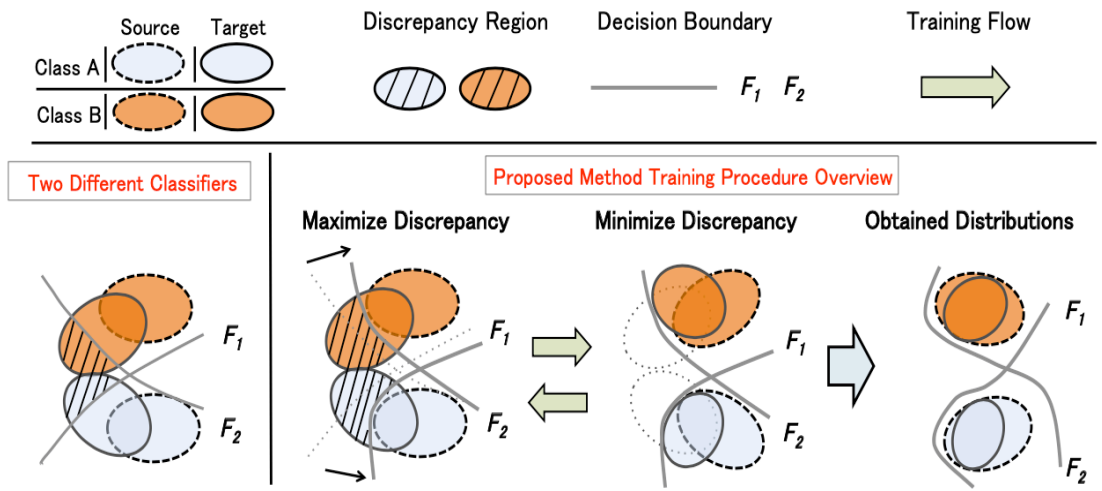
Step B

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

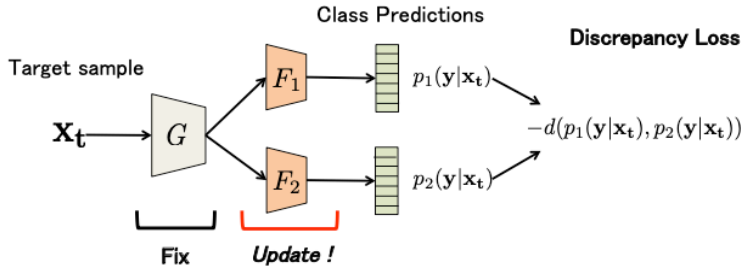
$$\mathcal{L}_{\text{adv}}(X_t) = \mathbb{E}_{\mathbf{x_t}\sim X_t}[d(p_1(\mathbf{y}|\mathbf{x_t}), p_2(\mathbf{y}|\mathbf{x_t}))]$$

Step C

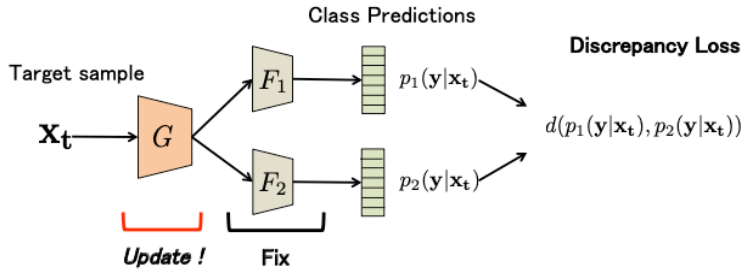
$$\min_{G} \mathcal{L}_{\text{adv}}(X_t)$$



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

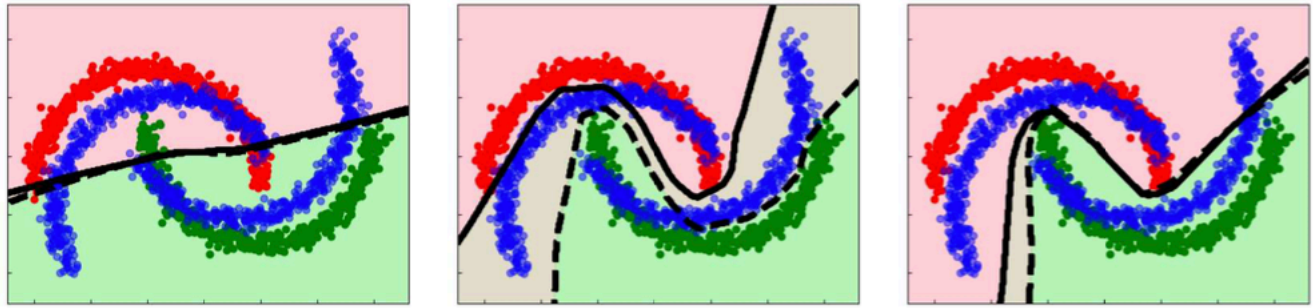


Step C : *Minimize* discrepancy on target (Fix F_1 , F_2)



Results

Toy Datasets



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

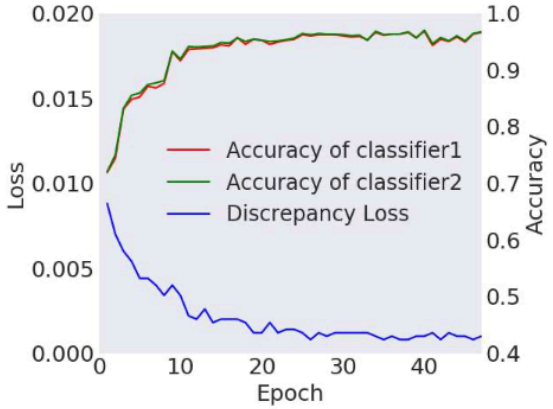
- Red: Source A / Green: Source B
- Blue: target samples (source samples rotated)
- Dashed/normal lines: two decision boundaries (F1, F2)
- Results of both decision boundary 0: pink / 1: light green

Results

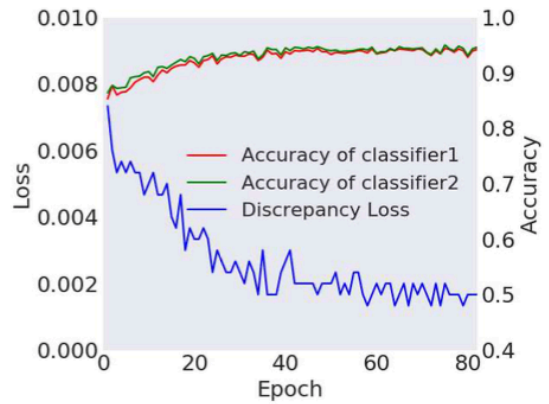
Digits Datasets

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

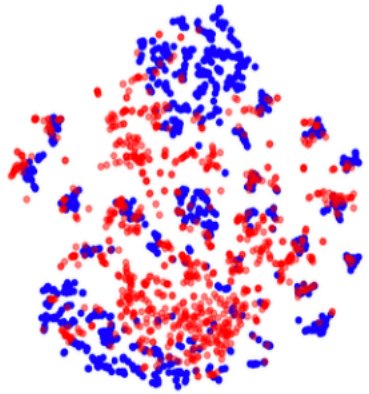
- n: # of times repeat step C for same mini-batch.



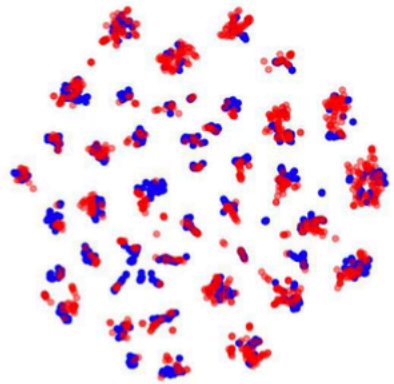
(a) SVHN to MNIST



(b) SYN SIGN to GTSRB



(c) Source Only



(d) Adapted (Ours)

Results

VisDA (Object classification)

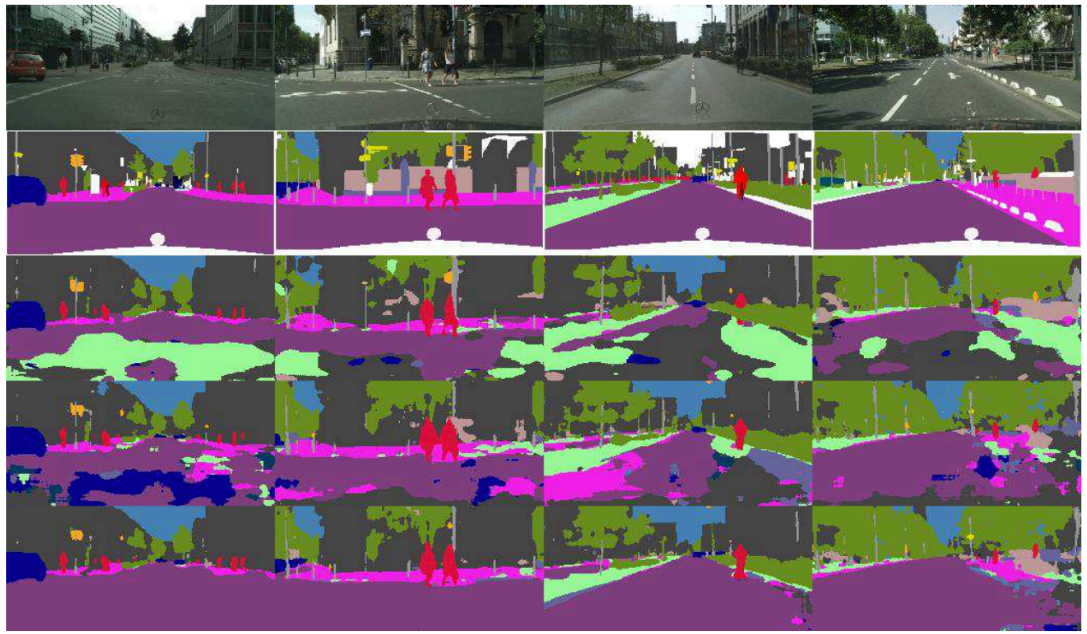
- Source domain: Synthetic objects / Target domain: MSCOCO (real objects)

Method	plane	bcycl	bus	car	horse	knife	mcycl	person	plant	sktbrd	train	truck	mean
Source Only	55.1	53.3	61.9	59.1	80.6	17.9	79.7	31.2	81.0	26.5	73.5	8.5	52.4
MMD [21]	87.1	63.0	76.5	42.0	90.3	42.9	**85.9**	53.1	49.7	36.3	**85.8**	20.7	61.1
DANN [7]	81.9	**77.7**	82.8	44.3	81.2	29.5	65.1	28.6	51.9	**54.6**	82.8	7.8	57.4
Ours ($n = 2$)	81.1	55.3	83.6	**65.7**	87.6	72.7	83.1	73.9	85.3	47.7	73.2	27.1	69.7
Ours ($n = 3$)	**90.3**	49.3	82.1	62.9	**91.8**	69.4	83.8	72.8	79.8	53.3	81.5	**29.7**	70.6
Ours ($n = 4$)	87.0	60.9	**83.7**	64.0	88.9	**79.6**	84.7	**76.9**	**88.6**	40.3	83.0	25.8	**71.9**

Results

Semantic Segmentation

- Source domain: GTA5, Synthia (synthetic) / Target domain: Cityscapes (real objects)



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.

THE END

THANK YOU!