Maximum Classifier Discrepancy for Unsupervised Domain Adaptation

Kuniaki Saito[1], Kohei Watanabe[1], Yoshitaka Ushiku[1], and Tatsuya Harada[1,2]
[1]The University of Tokyo [2]RIKEN AIP

東京大学 THE UNIVERSITY OF TOKYO

AIP CVPR 2018 SALT LAKE CITY • JUNE 18-22

Introduction

Unsupervised Domain Adaptation (UDA)

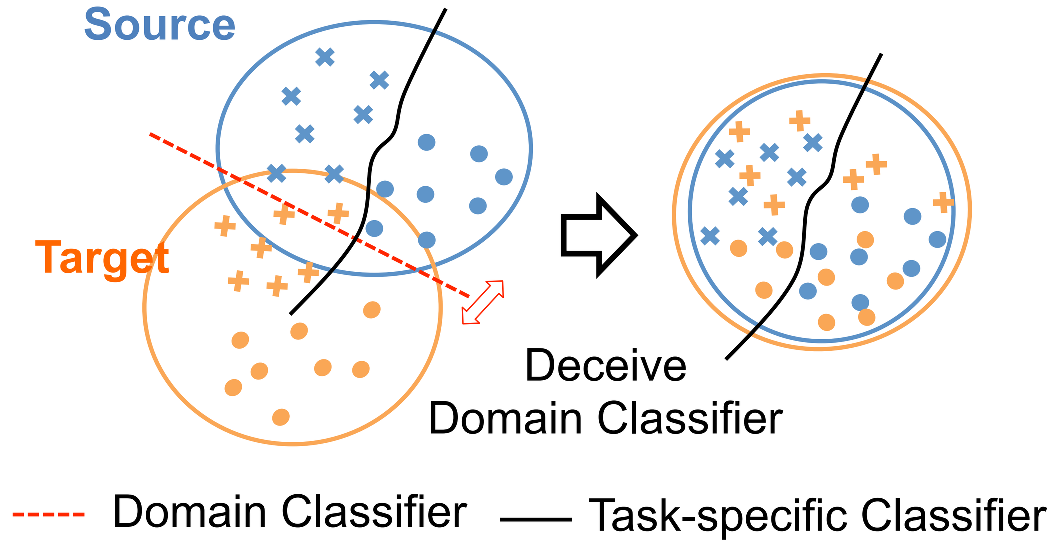
- Transfer knowledge from ***label-rich*** domain (source) to ***unlabeled*** domain (target).
- Goal is to obtain a good classifier for the target domain.

Popular Approach: Aligning features by a domain classifier

- Domain classifier tries to predict domain's label (source or target).
- Feature generator tries to deceive the domain classifier to extract domain-invariant features.

Problem

- Task-specific classifier is not considered to align features.
- Features of target can be generated near the decision boundary.



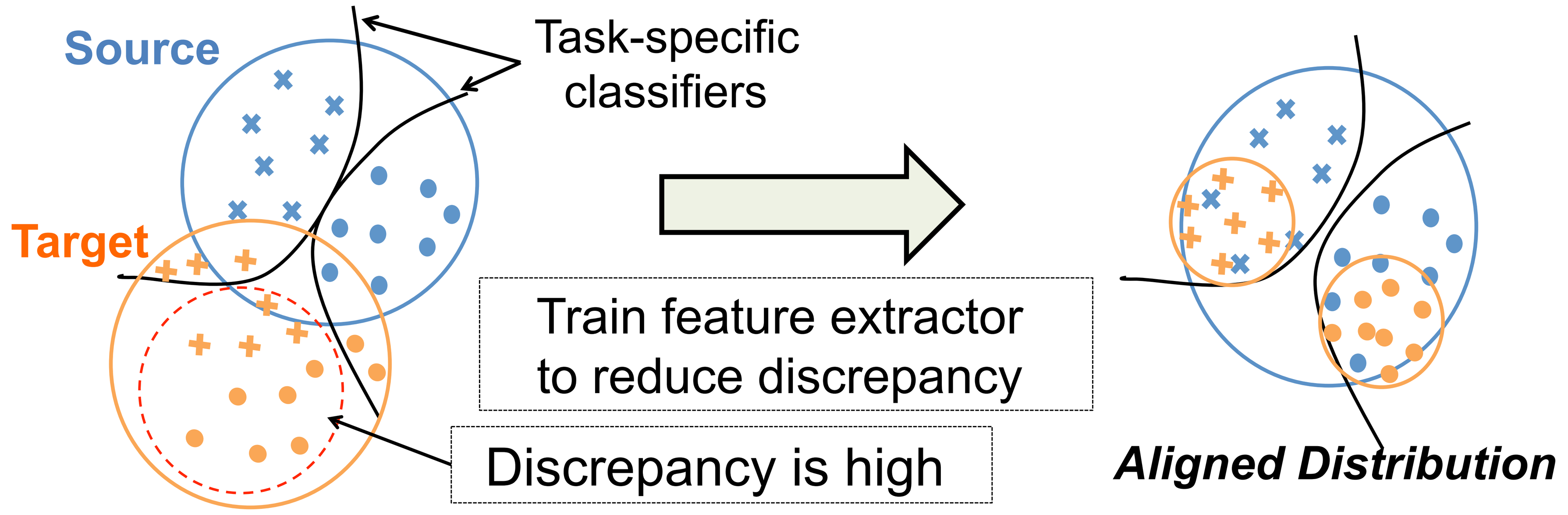
➡ We propose a new method, ***Maximum Classifier Discrepancy***, which utilizes the discrepancy (disagreement) of two classifiers.

Key Idea

Discrepancy : Disagreement of two classifiers' predictions

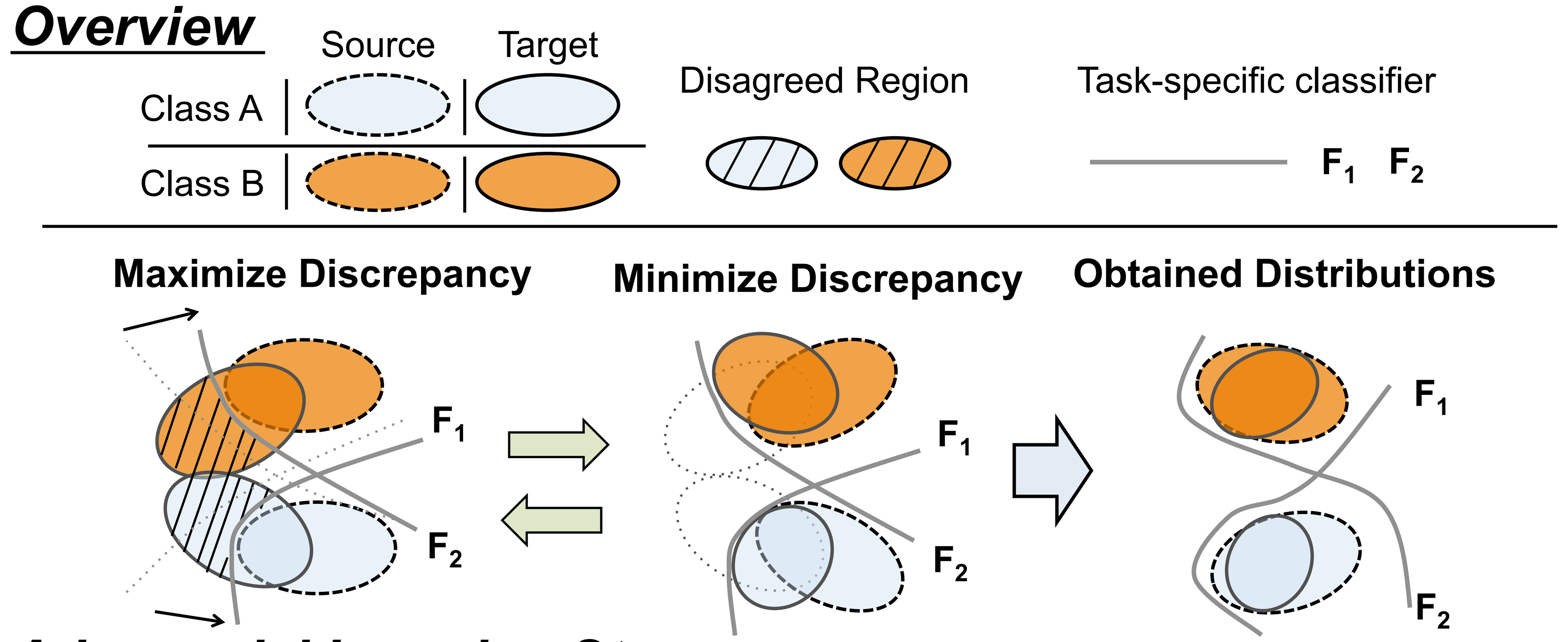
- We have to detect target ones far from the source ones.
- Discrepancy can be large for target ones far from source ones.

➡ Train two classifiers to increase the discrepancy on target.
Train feature extractor to reduce the discrepancy

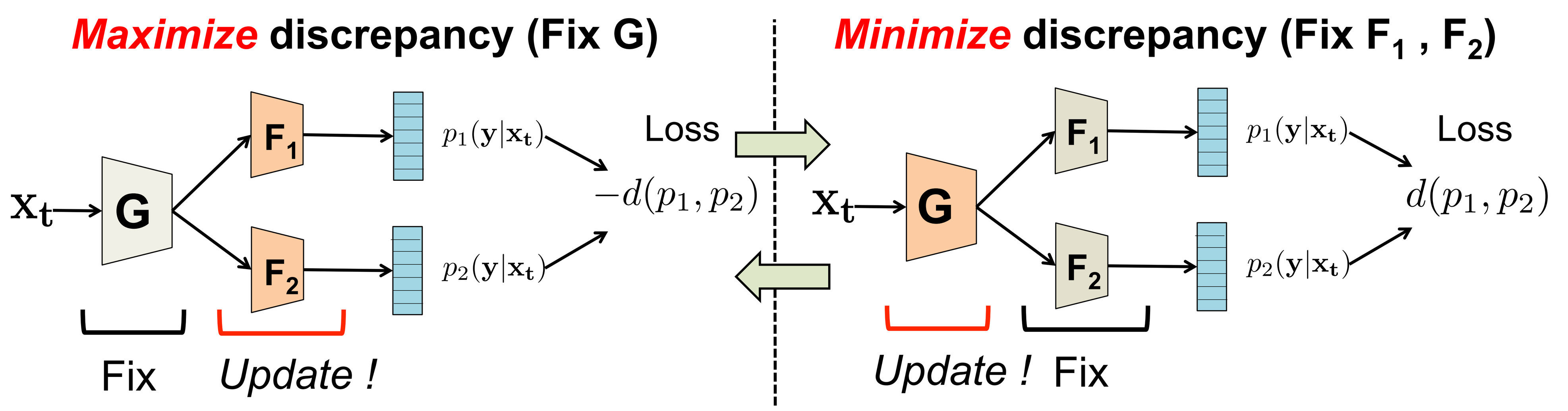


Method

Overview



Adversarial Learning Steps

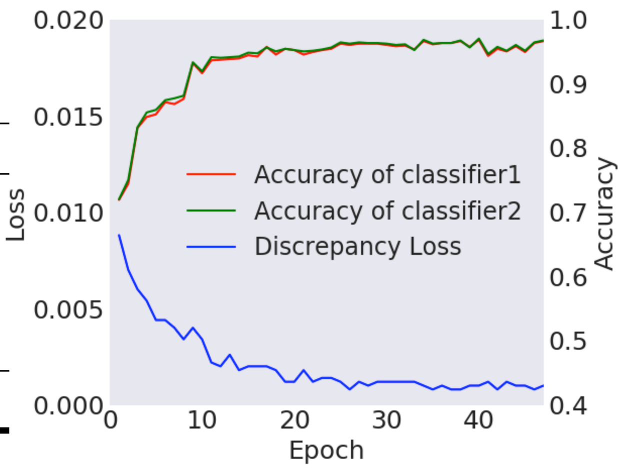


- Loss for source is also used to update all networks.
- We propose three steps training procedure for optimization.
- How many times we update **G** to minimize discrepancy per one iteration is a hyper-parameter.

Experiments1: Digits Image Classification

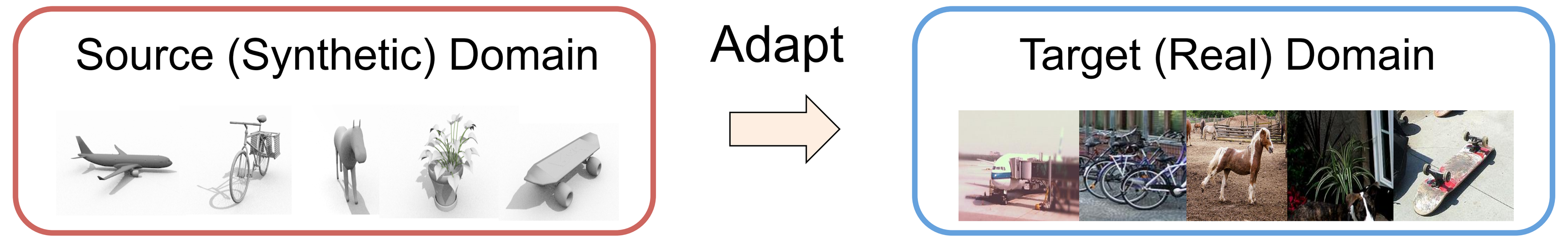
METHOD	SVHN to MNIST	SYNSIG to GTSRB	MNIST to USPS	MNIST* to USPS*	USPS to MNIST
Source Only	67.1	85.1	76.7	79.4	63.4
MMD [1]	71.1	91.1	-	81.1	-
DANN [2]	71.1	88.7	77.1±1.8	85.1	73.0±0.2
DSN [3]	82.7	93.1	91.3	-	-
ADDA [4]	76.0±1.8	-	89.4±0.2	-	90.1±0.8
Ours ($n=4$)	**96.2±0.4**	**94.4±0.3**	**94.2±0.7**	**96.5±0.3**	**94.1±0.3**

SVHN to MNIST



- Our method outperformed domain classifier based methods.
- The accuracy improves with the decrease of discrepancy loss.

Experiments2: Object Classification (Synthetic to Real)

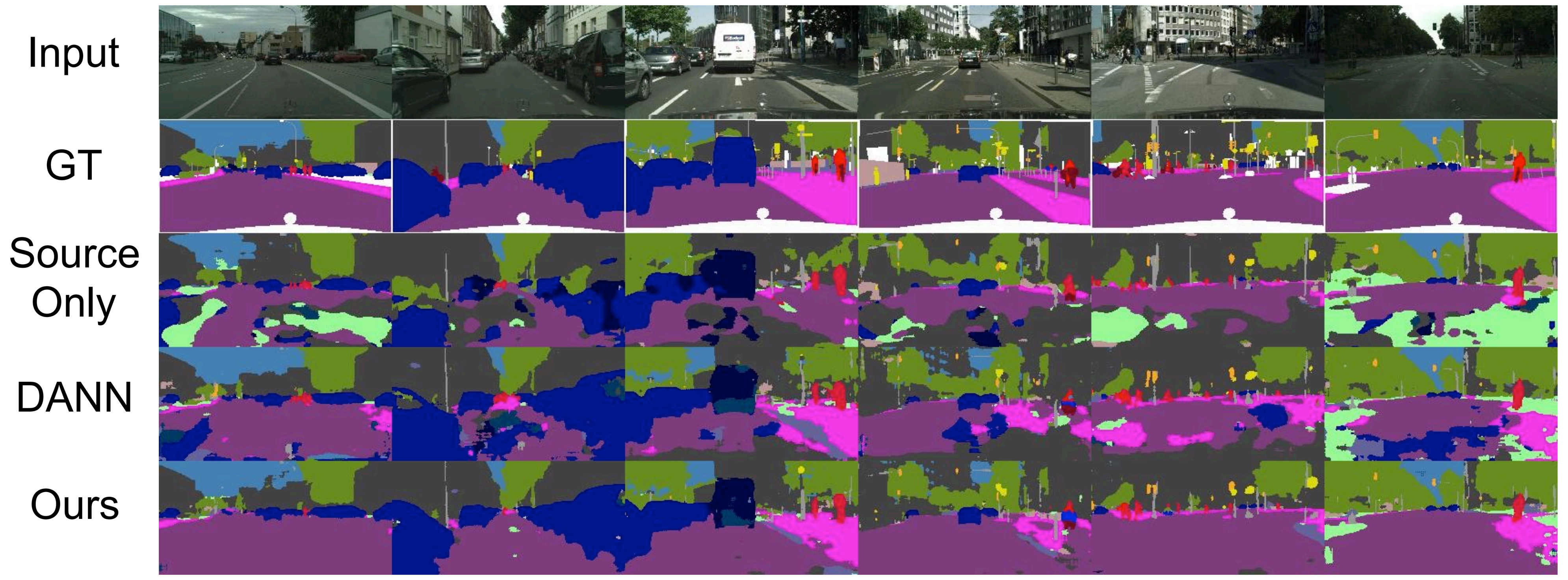


Method	plane	bcycl	bus	car	horse	knife	mcycl	person	plant	sktbrd	train	truck	mean
Source Only	55.1	53.3	61.9	59.1	80.6	17.9	79.7	31.2	81.0	26.5	73.5	8.5	52.4
MMD [1]	**87.1**	63.0	76.5	42.0	**90.3**	42.9	**85.9**	53.1	49.7	36.3	**85.8**	20.7	61.1
DANN [2]	81.9	**77.7**	82.8	44.3	81.2	29.5	65.1	28.6	51.9	**54.6**	82.8	7.8	57.4
Ours ($n=4$)	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**

- Our method is also effective for adaptation from synthetic and real.

Experiments3: Domain Adaptive Semantic Segmentation

Network	method	mIoU	road	sdwk	bldng	wall	fnce	pole	lght	sign	vgttn	trrn	sky	prsn	rder	car	trck	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 [5]	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) [6]	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 (n=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 [2]	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 (n=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**



- Source: GTA5, Target: CityScape, finetune VGG and DRN network
- Significant improvement in mIoU was observed.

Summary

- A new adversarial learning for UDA using task-specific classifiers
- Method effective for image classification and semantic segmentation
- The effectiveness is verified through extensive experiments.
- Gradient reversal layer [2] also works in many settings.
- We proposed to sample two networks from a network with dropout [7].

[1] M. Long et al., Learning transferable features with deep adaptation networks. In *ICML*, 2015.
[2] Y.Ganin et al., Unsupervised domain adaptation by backpropagation. In *ICML*, 2015.
[3] K.Bousmalis et al., Domain separation networks. In *NIPS*, 2016.
[4] E. Tzeng et al., Adversarial discriminative domain adaptation. In *CVPR*, 2017.
[5] J. Hoffman et al., FCNs in the wild: Pixel-level adversarial and constraint-based adaptation. *ArXiiv* 2016.
[6] Y. Zhang et al., Curriculum domain adaptation for semantic segmentation of urban scenes. In *ICCV*, 2017.
[7] K. Saito et al., Adversarial Dropout Regularization. In ICLR 2018.