PAPER ANALYSIS

Presented by Yannis He

-

Paper: **Target-targeted Domain Adaptation for Unsupervised Semantic Segmentation**

Conference: ICRA 2021

Authors: Xiaohong Zhang, Haofeng Zhang, Jianfeng Lu, Ling Shao, and Jingyu Yang

- | National Natural Science Foundation of China
- | Nanjing University of Science and Technology
- | Inception Institute of Artificial Intelligence (IIAI), Abu Dhabi

https://ieeexplore.ieee.org/abstract/document/9560785

- Background:
  - Semantic segmentation has attracted increasing attention due to its important role in self-driving, and it is often realized by supervised learning with large number of well labeled maps.
  - The labeled images are hard to be obtained in most circumstances
  - The common way for unsupervised semantic segmentation is usually implemented by transferring the knowledge from source supervised domain to target unsupervised domain
  - Current research encouraging target predictions to be closer to the source ones through a weight-sharing network, and achieve certain performance.
    - Cons:
      - suffer from the domain shift problem that the networks are often trained towards the source domain and lead to performance degradation. → unable to focus on the target domain.
      - adversarial training method induces the segmentation network to be biased towards the source domain.

INTRODUCTION

- Proposal:
  - A target-targeted domain adaptation approach by focusing the training on target domain.
    - Consists of two components:
    1. the Image-to-image Translation (IIT) module to translate the source image to target domain
    2. the Target-targeted Segmentation Adaptation (TSA) module to focus the semantic segmentation on target domain.
  - The IIT module deals with image space alignment while the TSA module bridges the domain gap at the Segmentation map level.
  - In addition, the authors design a closed-loop learning to promote each other by employing feedback from TSA to IIT.

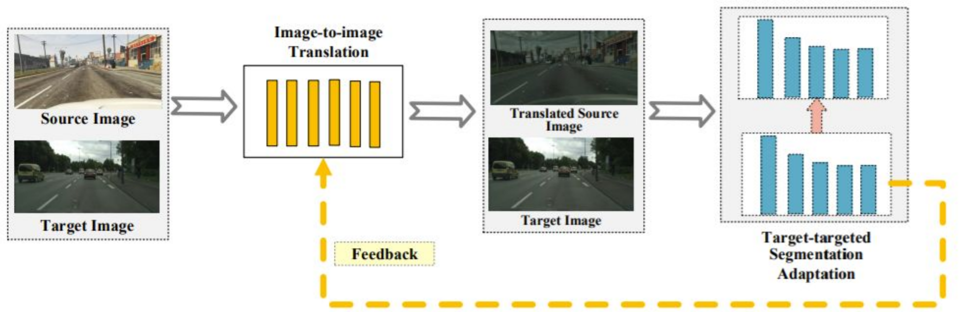


Fig. 1. The brief schematic of our proposed method. This method including two important schemes, one is to translate the source image into target domain and the other is the feedback.

INTRODUCTION

- Result:
  - Extensive experiments are conducted on the benchmarks of GTA5 and SYNTHIA to Cityscapes, the results show that our method can achieve the state-of-the-art performance

METHOD - OVERVIEW

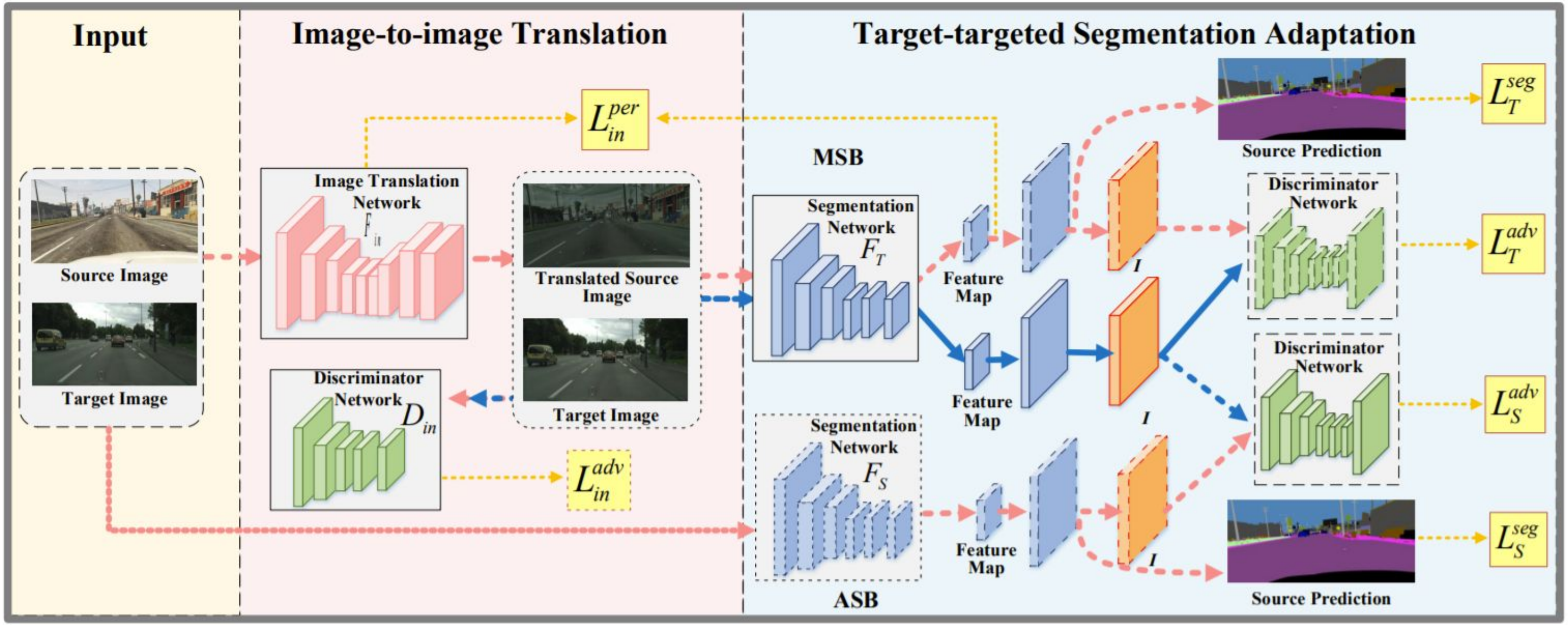
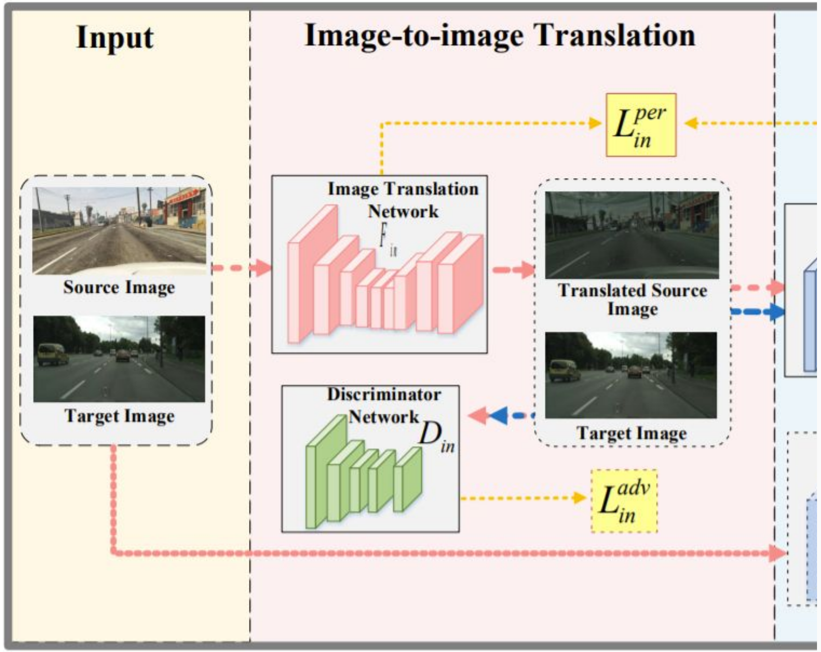


Fig. 2. Algorithmic overview. Our method consist of two modules. First, the Image-to-Image Translation (IIT) module translates the source image to the target domain. Next, the Target-targeted Segmentation Adaptation (TSA) further bridges the domain gap at the output level. We also add a perceptual loss as feedback from TSA to improve IIT, making the whole system a closed loop. The red line indicates the flow of the source image and the blue line indicates the flow of the target image.

METHOD - IMAGE-TO-IMAGE TRANSLATION

- Goal: reduce the discrepancy in appearance between the two domains.

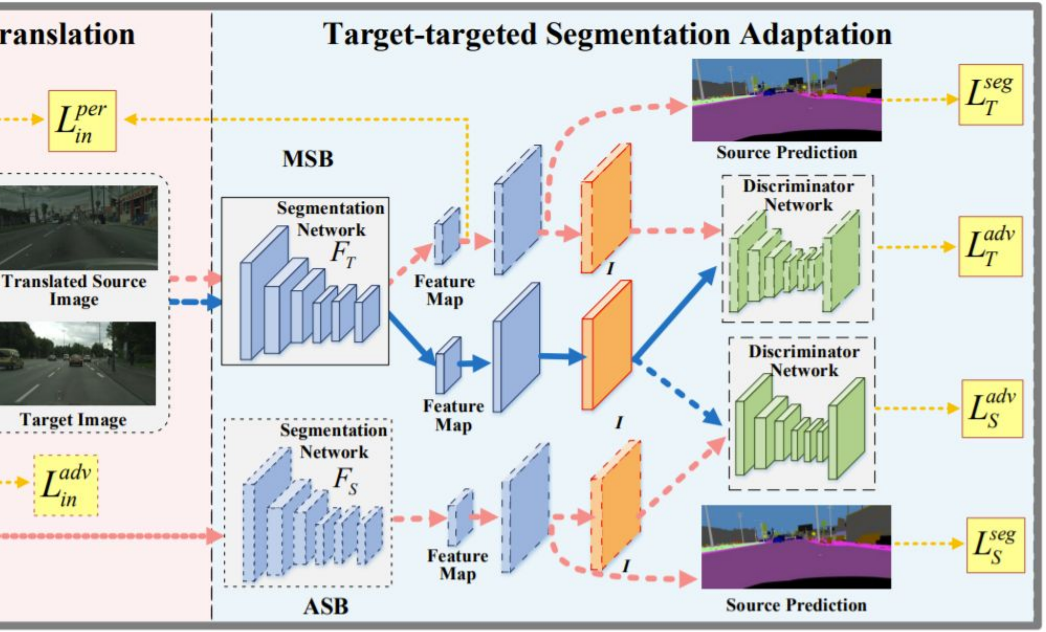


$$L_{in} = L_{in}^{adv}(I_{S'}, I_T) = \sum_{h,w} \log(1 - I_{S'}^{(h,w)}) + \log(I_T^{(h,w)}), \quad (1)$$

where, h, w are the width and height of the images respectively, and $I_{S'} = F_{in}(I_S)$.

- Two branch:
  - Main Segmentation Branch (MSB) and the Auxiliary Segmentation Branch (ASB).

$$L_{\mathrm{T}}^{adv}\left(I_{S'}, I_T\right) = \sum_{h,w} \log\left(1 - F_{\mathrm{T}}\left(I_T^{(h,w)}\right)\right) + \log\left(F_{\mathrm{T}}\left(I_{S'}^{(h,w)}\right)\right). \quad (2)$$



RESULTS

- Target-targeted approach models can achieve state-of-the-art performance in the two UDA benchmarks:
  - GTA5 → Cityscapes:
    - Compared with the segmentation model without adaptation, our method brings +9.0% mIoU improvement with ResNet101 and +10.5% mIoU improvement with VGG16, demonstrating the ability of our method to mitigate the domain gap.
  - SYNTHIA → Cityscapes (more challenging than GTA5 → Cityscapes):
    - our method is superior to the state-of-the-art methods in domain adaptation of segmentation

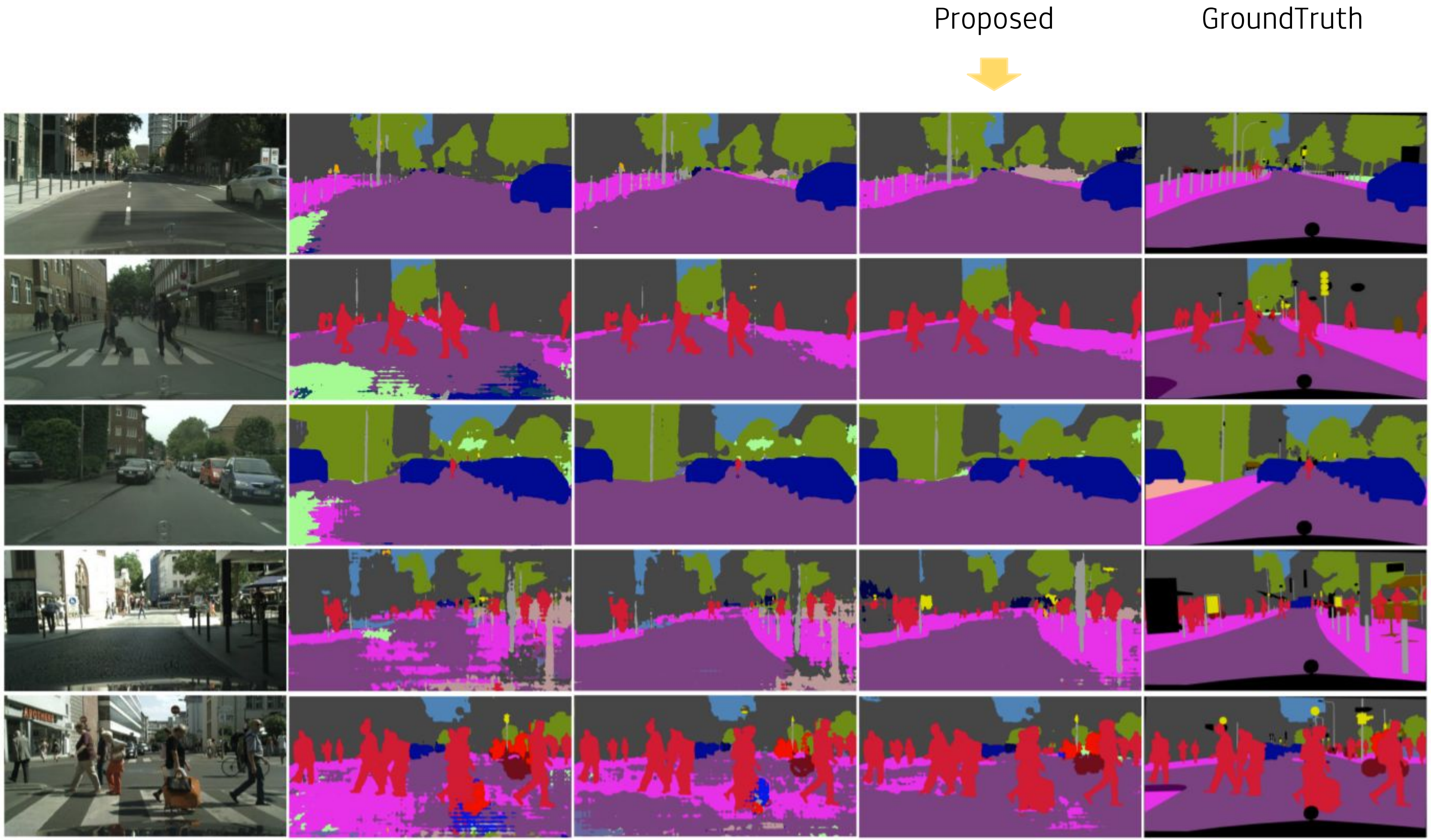


Fig. 3. Visual Comparison on the Cityscapes dataset. Left to right: left images, AdaptSegNet [20], AdvEnt [9], our method and ground truth maps.

TABLE I

RESULTS OF ADAPTING GTA5 TO CITYSCAPES. THE 19 CLASS MIOU (%) IS USED AS THE EVALUATION MATRIX OF SEMANTIC SEGMENTATION PERFORMANCE.

GTA5 $\rightarrow$ Cityscapes																					
Base Model	Method	road	sdwk	bldng	wall	fence	pole	light	sign	vgttn	trrn	sky	person	rider	car	truck	bus	train	mcycl	bcycl	mIoU
ResNet101	Without adaptation [28]	75.8	16.8	77.2	12.5	21	25.5	30.1	20.1	81.3	24.6	70.3	53.8	26.4	49.9	17.2	25.9	**6.5**	25.3	36.0	36.6
	ROAD [16]	76.3	36.1	69.6	28.6	22.4	28.6	29.3	14.8	82.3	35.3	72.9	54.4	17.8	78.9	27.7	30.3	4.0	24.9	12.6	39.4
	AdaptSegNet [20]	86.5	36.0	79.9	23.4	23.3	23.9	35.2	14.8	83.4	33.3	75.6	58.5	27.6	73.7	32.5	35.4	3.9	30.1	28.1	42.0
	CyCADA [30]	86.7	35.6	80.1	19.8	17.5	**38.0**	**39.0**	**41.5**	82.7	27.9	73.6	**64.9**	19.0	65.0	12.0	28.6	4.5	31.1	**42.0**	42.7
	MinEnt [9]	84.2	25.2	77.0	17.0	23.3	24.2	33.3	26.4	80.7	32.1	78.7	57.5	**30.0**	77.0	**37.9**	44.3	1.8	**31.4**	36.9	43.1
	AdvEnt [9]	89.9	36.5	81.6	29.2	**25.2**	28.5	32.3	22.4	**83.9**	34.0	77.1	57.4	27.9	**83.7**	29.4	39.1	1.5	28.4	23.3	43.8
	Ours	**91.5**	**42.4**	**84.6**	**29.2**	23.9	29.5	32.5	36.4	83.0	**37.1**	**80.8**	58.4	25.3	82.7	27.7	**46.1**	0	26.3	29.6	**45.6**
VGG16	Without adaptation [28]	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	7.3	0	3.5	0	27.1
	MinEnt [9]	85.1	18.9	76.3	**32.4**	19.7	19.9	21	8.9	76.3	26.2	63.1	42.8	5.9	80.8	20.2	9.8	0	14.8	0.6	32.8
	CyCADA [30]	83.5	38.3	76.4	20.6	16.5	22.2	26.2	**21.9**	80.4	28.7	65.7	49.4	4.2	74.6	16.0	26.6	2.0	8.0	0.0	34.8
	Adapt-SegMap [20]	87.3	29.8	78.6	21.1	18.2	**22.5**	21.5	11.0	79.7	29.6	71.3	46.8	6.5	80.1	23.0	26.9	0.0	10.6	0.3	35.0
	AdvEnt [9]	86.9	28.7	78.7	28.5	**25.2**	17.1	20.3	10.9	80	26.4	70.2	47.1	8.4	81.5	**26**	17.2	**18.9**	11.7	1.6	36.1
	BDL [10]	**89.2**	**40.9**	**81.2**	29.1	19.2	14.2	**29.0**	19.6	**83.7**	**35.9**	**80.7**	**54.7**	**23.3**	**82.7**	25.8	**28.0**	2.3	**25.7**	**19.9**	**41.3**
	Ours	88.7	38.6	80.2	26	21.5	22.3	25	14.7	83.2	32.3	77	53	17.5	81.1	21.3	21.5	0.01	21.5	7.8	38.6

TABLE II

RESULTS OF ADAPTING SYNTHIA TO CITYSCAPES. THE 16/13 CLASS MIOU (%) IS USED AS THE EVALUATION MATRIX OF SEMANTIC SEGMENTATION PERFORMANCE.

SYNTHIA $\rightarrow$ Cityscapes																			
Base Model	Method	road	sdwk	bldng	wall*	fence*	pole*	light	sign	vgttn	sky	person	rider	car	bus	mcycl	bcycl	mIoU	mIoU*
ResNet101	Without adaptation [28]	55.6	23.8	74.6	9.2	0.2	24.4	6.1	12.1	74.8	79.0	55.3	19.1	39.6	23.3	13.7	25.0	33.5	38.6
	MinEnt [9]	73.5	29.2	77.1	7.7	0.2	27.0	7.1	11.4	76.7	82.1	57.2	21.3	69.4	29.2	12.9	27.9	38.1	44.2
	SIBAN [31]	82.5	24.0	79.4	-	-	-	16.5	12.7	79.2	82.8	**58.3**	18.0	**79.3**	25.3	**17.6**	25.9	-	46.3
	AdvEnt [9]	**87.0**	**44.1**	**79.7**	9.6	0.6	24.3	4.8	7.2	80.1	83.6	56.4	**23.7**	72.7	**32.6**	12.8	**33.7**	40.8	**47.6**
	Ours	71.3	26.7	71.4	**15.4**	**1.8**	**30.4**	**29.6**	**32.6**	**80.2**	**84.1**	50.1	18.4	73.1	29.0	11.7	33.1	**41.2**	47.0
VGG16	Without adaptation [28]	11.5	19.6	30.8	-	-	-	0.1	11.7	42.3	68.7	**51.2**	3.8	54.0	3.2	0.2	0.6	-	22.9
	MinEnt [9]	37.8	18.2	65.8	2.0	0.0	15.5	0.0	0.0	76	73.9	45.7	11.3	66.6	13.3	1.5	13.1	27.5	32.5
	ROAD [16]	77.7	30.0	**77.5**	**9.6**	0.3	**25.8**	**10.3**	**15.6**	77.6	**79.8**	44.5	**16.6**	67.8	14.5	**7.0**	**23.8**	36.2	41.7
	AdvEnt [9]	67.9	29.4	71.9	6.3	0.3	19.9	0.6	2.6	74.9	74.9	35.4	9.6	**67.8**	21.4	4.1	15.5	31.4	36.6
	Ours	**82.3**	**40.8**	77.0	8.8	**1.1**	23.5	8.9	15.4	**79.2**	77.8	43.4	14.3	64.0	**26.5**	5.9	17.6	**36.7**	**42.5**

ABLATION

TABLE III

ABLATION STUDY TO DEMONSTRATE THE EFFECT OF EACH MODULE ON GTA5 $\rightarrow$ CITYSCAPES.

Network Setting				mIoU
IIT		TSA		
F_{in}	L_{in}^{per}	MSB	ASB	
		√		35.6
		√	√	37.5
√		√	√	43.6
√	√	√	√	45.6