Multi-Source Domain Adaptation with Collaborative Learning for Semantic Segmentation

Jianzhong He, Xu Jia, Shuaijun Chen, Jianzhuang Liu
1Data Storage and Intelligent Vision Technical Research Dept, Huawei Cloud.
2Noah's Ark Lab, Huawei Technologies. 3Dalian University of Technology
CVPR(2021)

July. 28, 2021

Yujeong Lee (CVLab)

(https://openaccess.thecvf.com/content/CVPR2021/papers/He_Multi-Source_Domain_Adaptation_With_Collaborative_Learning_for_Semantic_Segmentation_CVPR_2021_paper.pdf)

1

요약

Multi-Source Domain Adaptation with Collaborative Learning for Semantic Segmentation

[연구 목적] Multi-source domain adaptation for semantic segmentation

Source domain : GTA5, Synscapes, Synthia

Target domain : Cityscapes

[방법 1] LAB-based image translation

[방법 2] Collaborative learning between labeled source domains

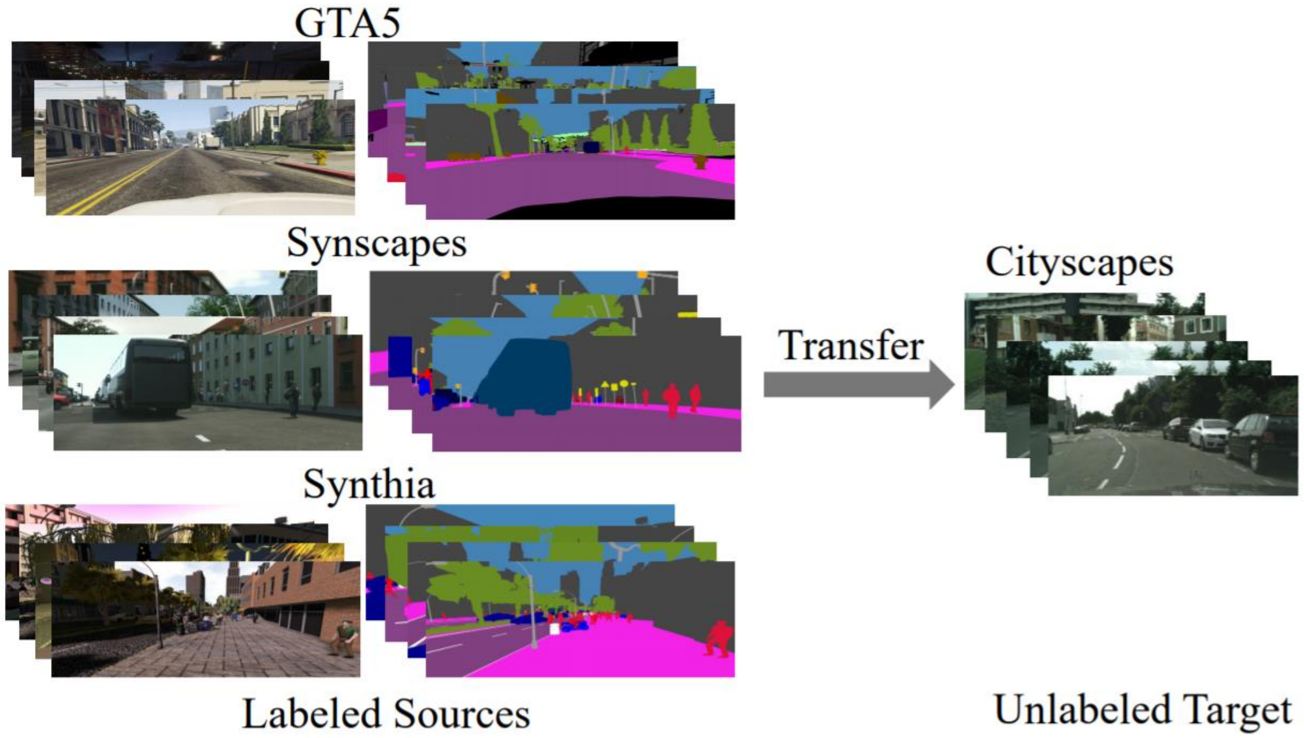
[방법 3] Collaborative learning on unlabeled target domain (Self-training)

[성능] mIoU 59.0 (이전 SOTA : 55.7)

2 방법 1

LAB-based image translation

Translation Method



a RGB image $\mathcal{X}_S^{RGB}$ in source domains

$$\mathcal{X}_S^{LAB} = rgb2lab(\mathcal{X}_S^{RGB})$$

Calculate standard deviation values σ_S the mean μ_S

$$\hat{\mathcal{X}}_S^{LAB} = \frac{(\mathcal{X}_S^{LAB} - \mu_S)}{\sigma_S} * \sigma_T + \mu_T.$$

$$\hat{\mathcal{X}}_S^{RGB} = lab2rgb(\hat{\mathcal{X}}_S^{LAB}),$$

2 방법 1

LAB-based image translation

Results

Table 1. Validity of the proposed image translation method. The performance comparison with the recent single-source UDA methods trained on images that before and after translation.

GTA5→Cityscapes			
Methods	Before	+Trans	Diff.
Direct Transfer	39.53	43.36	↑ 3.83
AdaptSeg [32]	41.32	43.66	↑ 2.43
AdaptSeg-LS [32]	43.11	45.95	↑ 2.84
Advent [35]	44.30	45.96	↑ 1.66

Source Image Trans. on RGB Trans. on LAB Target Image

Figure 3. The qualitative comparison of image translation on different color space.

3 방법 2

Collaborative learning between labeled source domains

"each source domain teach each other to extract essential semantic information across domains"

Assume that there are N different labeled source domains $S = \{S_1, S_2, \cdots, S_N\}$ which are sampled from N different *i.i.d* distributions, and N deep neural networks $\mathcal{M} = \{\mathcal{M}_{S_1}, \mathcal{M}_{S_2}, \cdots, \mathcal{M}_{S_N}\}$ of the same architecture but different weights learned on these source domains. Then, for an model $\mathcal{M}_{S_i}$, the learning process of model $\mathcal{M}_{S_i}$ is supervised by segmentation loss on labeled data from source S_i and collaborative loss on output from source $S_{k,k\neq i}$. That is, for model $\mathcal{M}_{S_i}$, the object function is

$$\mathcal{L}_i = \mathcal{L}^{seg}_{S_i}(\mathcal{F}^{S_i}_{S_i}, \mathcal{Y}_{S_i}) + \lambda^{col}_S \mathcal{L}^{col}_S(\{(\mathcal{F}^{S_k}_{S_i}, \mathcal{F}^{S_k}_{S_k})_{k\neq i}\}), \quad (2)$$

where the loss $\mathcal{L}^{seg}$ is the cross entropy loss, *i.e.*,

$$\mathcal{L}^{seg}_S(\mathcal{F}_S, \mathcal{Y}_S) = -\frac{1}{|\mathcal{X}_S|}\sum_{h,w}\sum_{c\in C} \mathcal{Y}^{(h,w,c)}_S log(\sigma(\mathcal{F}^{(h,w,c)}_S)), \quad (3)$$

and the loss $\mathcal{L}^{col}$ is the average of Kullback-Leibler (KL) divergence loss, *i.e.*,

$$\mathcal{L}^{col}_S(\{(\mathcal{F}^{S_k}_{S_i}, \mathcal{F}^{S_k}_{S_k})_{k\neq i}\}) = \frac{1}{N-1}\sum_{k,k\neq i} \mathcal{L}^{kl}_{k\to i}(\mathcal{F}^{S_k}_{S_k} \| \mathcal{F}^{S_k}_{S_i}), \quad (4)$$

$$\mathcal{L}^{kl}_{k\to i}(\mathcal{F}^{S_k}_{S_k} \| \mathcal{F}^{S_k}_{S_i}) = -\frac{1}{|\mathcal{X}_{S_k}|}\sum \sigma(\mathcal{F}^{S_k}_{S_k}) \log(\frac{\sigma(\mathcal{F}^{S_k}_{S_i})}{\sigma(\mathcal{F}^{S_k}_{S_k})}). \quad (5)$$

방법 3

Collaborative learning on unlabeled target domains

Pseudo label을 만들고 source와 target을 함께 이용하여 모델 학습

[Pseudo label 생성]

Target domain의 image를 N개의 model을 통과시킨다.
각 output(feature)을 모든 모델에 대해 더한다. (ensemble)
softmax(ensembled feature) ⇔ Pseudo label

$$\hat{P} = \sigma(\frac{1}{N}\sum_i \mathcal{F}_{S_i}^T)$$

[Training]

Source와 target domain을 함께 사용함
Loss : cross entropy loss

$$\mathcal{L} = \mathcal{L}_S^{seg} + \lambda_S^{col}\mathcal{L}_S^{col} + \frac{cur\_it}{max\_its}\lambda_T^{seg}\mathcal{L}_T^{seg}$$

Algorithm 1: Pseudo Labels Generation

Data: The probability map $\hat{P} \in \mathcal{R}^{C\times H\times W}$, keep proportion α, maximum thresh τ, the ignore label l_{ig}

Result: one-hot hard pseudo labels $\hat{\mathcal{Y}}$

$\hat{\mathcal{Y}} \leftarrow argmax(\hat{P}, dim = 0), \hat{\mathcal{Y}} \in \mathcal{R}^{H\times W}$

for $c \leftarrow 0$ *to* $C - 1$ **do**

- $\hat{P}_c \leftarrow sort(\hat{P}_{\{c,\cdot,\cdot\}}, order = Descending)$;
- get the number of pixels n_c which are predicted to category c: $n_c \leftarrow sum(\hat{\mathcal{Y}} == c)$;
- get the threshold t that used to filter the prediction: $t \leftarrow min(\hat{P}_c[n_c \times \alpha], \tau)$;
- $mask1 \leftarrow \hat{\mathcal{Y}} == c$;
- $mask2 \leftarrow \hat{P}_{\{c,\cdot,\cdot\}} <= t$;
- $\hat{\mathcal{Y}}[mask1 \,\&\, mask2] \leftarrow l_{ig}$.

end

5 전체 Method

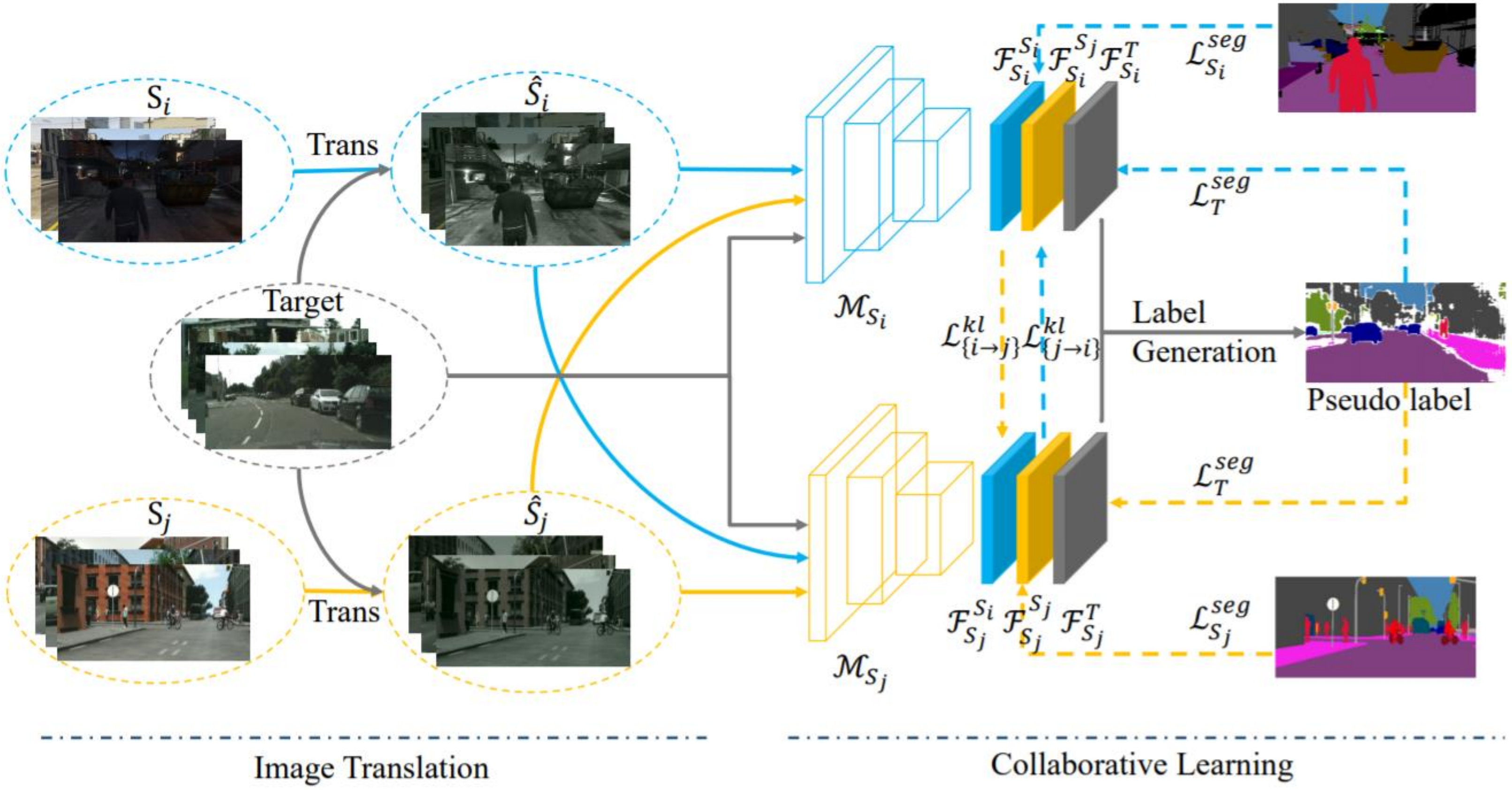


Figure 2. The overall framework of proposed approach consists of three components, including that image-to-image translation based on LAB color space, collaborative learning between source domains and collaborative learning on target domain. The solid arrows represent the forward data flow and different colors indicate different source domains or target domain data flow. The dash arrows represent the supervision to the network outputs. For illustration, we just show the case of two source domains as an example to explain our method.

6 실험 결과

Ablation study

Table 2. The validity of model selection and the proposed collaborative learning on the GTA5 + Synscapes to Cityscapes. (a) shows the performance of each single model and the final ensemble, (b) shows the comparison of proposed collaborative learning between source domains (Co-Learning-Src) with baseline and MLDG [40]. **E:** End-to-End, **S:** Stage-Wise.

(a)

Model	E	S
$\mathcal{M}_{S_{GTA5}}$	56.90	57.72
$\mathcal{M}_{S_{Syns}}$	56.65	57.81
$\mathcal{M}_{Ensemble}$	58.55	59.04

(b)

Methods	mIoU	Diff.
Data Combination	51.56	–
MLDG+TN [40]	52.73	↑ 1.17
Co-Learning-Src	55.79	↑ 4.23

Table 3. Ablation studies of proposed methods. Note that, the performances are achieved by end-to-end training strategy for comparison with simple combination of sources.

GTA5 + Synscapes → Cityscapes				
LAB-based Trans.	Data Comb.	Co-Learning between Src.	Co-Learning on Target	mIoU
	✓			51.59
✓	✓			54.38
✓			✓	54.03
✓		✓		56.03
		✓	✓	57.27
✓		✓	✓	58.55

6 실험 결과

Ablation study

Table 4. The quantitative comparison with the state-of-the-art methods. DT is the abbreviation of direct transfer. G, S and A indicate GTA5, Synscapes and All respectively. Adv, CL, ST and RL indicate Adversarial learning, Curriculum Learning, Self Training and Reconstruction Learning respectively. Ours-E and Ours-S represent end-to-end training and stage-wise training of our proposed method respectively.

Methods	Appr.	Source	road	sidewalk	building	wall	fence	pole	light	sign	veg	terrain	sky	person	rider	car	truck	bus	train	mbike	bike	mIoU
DT [32]	–		81.8	40.6	76.1	23.3	16.8	36.9	36.8	40.1	83.0	34.8	84.9	59.9	37.7	78.5	20.4	20.5	7.8	27.3	52.5	45.3
AdaptSeg [32]	Adv		94.2	60.9	85.1	29.1	25.2	38.6	43.9	40.8	85.2	29.7	88.2	64.4	40.6	85.8	31.5	43.0	28.3	30.5	56.7	52.7
FDA [37]	ST	S	93.6	58.1	84.0	30.4	29.2	39.0	43.1	**51.7**	85.9	28.8	86.9	64.0	**45.7**	84.7	30.4	36.5	28.5	34.4	**62.4**	53.5
Advent [35]	Adv		92.2	51.3	85.0	40.8	31.2	39.0	42.5	42.5	86.5	46.1	84.8	65.2	39.0	87.0	32.6	49.0	29.5	28.6	50.0	53.8
UIA [21]	Adv		94.0	60.0	84.9	29.5	26.2	38.5	41.6	43.7	85.3	31.7	88.2	66.3	44.7	85.7	30.7	53.0	29.5	36.5	60.2	54.2
DT [32]	–		75.8	16.8	77.2	12.5	21.0	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
AdaptSeg [32]	Adv		86.5	25.9	79.8	22.1	20.0	23.6	33.1	21.8	81.8	25.9	75.9	57.3	26.2	76.3	29.8	32.1	7.2	29.5	32.5	41.4
Advent [35]	Adv		89.4	33.1	81.0	26.6	26.8	27.2	33.5	24.7	83.9	36.7	78.8	58.7	30.5	84.8	38.5	44.5	1.7	31.6	32.4	45.5
UIA [21]	Adv		90.6	36.1	82.6	29.5	21.3	27.6	31.4	23.1	85.2	39.3	80.2	59.3	29.4	86.4	33.6	53.9	0.0	32.7	37.6	46.3
PyCDA [14]	CL	G	90.5	36.3	84.4	32.4	28.7	34.6	36.4	31.5	86.8	37.9	78.5	62.3	21.5	85.6	27.9	34.8	18.0	22.9	49.3	47.4
BDL [13]	ST		91.0	44.7	84.2	34.6	27.6	30.2	36.0	36.0	85.0	43.6	83.0	58.6	31.6	83.3	35.3	49.7	3.3	28.8	35.6	48.5
FDA [37]	ST		92.5	53.3	82.4	26.5	27.6	36.4	40.6	38.9	82.3	39.8	78.0	62.6	34.4	84.9	34.1	53.1	16.9	27.7	46.4	50.5
PIT [20]	RL		87.5	43.4	78.8	31.2	30.2	36.3	39.9	42.0	79.2	37.1	79.3	65.4	37.5	83.2	46.0	45.6	25.7	23.5	49.9	50.6
Data Comb.	–		85.1	36.9	84.1	39.0	33.3	38.7	43.1	40.2	84.8	37.1	82.4	65.2	37.8	69.4	43.4	38.8	34.6	33.2	53.1	51.6
AdaptSeg [32]	Adv		89.3	47.3	83.6	40.3	27.8	39.0	44.2	42.5	86.7	45.5	84.5	63.1	38.0	79.4	34.9	48.3	42.1	30.7	52.3	53.7
Advent [35]	Adv		91.8	49.0	84.6	39.4	31.5	39.9	42.9	43.5	86.3	45.1	84.6	65.3	41.0	87.1	37.9	49.2	31.0	30.3	48.8	54.2
MDAN [42]	Adv	A	92.4	56.1	86.8	42.7	32.9	39.3	48.0	40.3	87.2	47.2	**90.5**	64.1	35.9	87.8	33.8	48.6	39.0	27.6	49.2	55.2
MADAN [43]	Adv		94.1	61.0	86.4	43.3	32.1	40.6	49.0	44.4	87.3	47.7	89.4	61.7	36.3	87.5	35.5	45.8	31.0	33.5	52.1	55.7
Ours-E	–		**94.2**	**61.8**	86.7	**47.7**	34.1	39.3	44.6	34.2	87.2	**49.6**	89.7	**65.6**	38.1	**88.2**	**48.1**	**63.0**	41.9	**39.2**	59.2	58.6
Ours-S	–		93.6	59.6	**87.1**	44.9	**36.7**	**42.1**	**49.9**	42.5	**87.7**	47.6	89.9	63.5	40.3	88.2	41.0	58.3	**53.1**	37.9	57.7	**59.0**

6 실험 결과

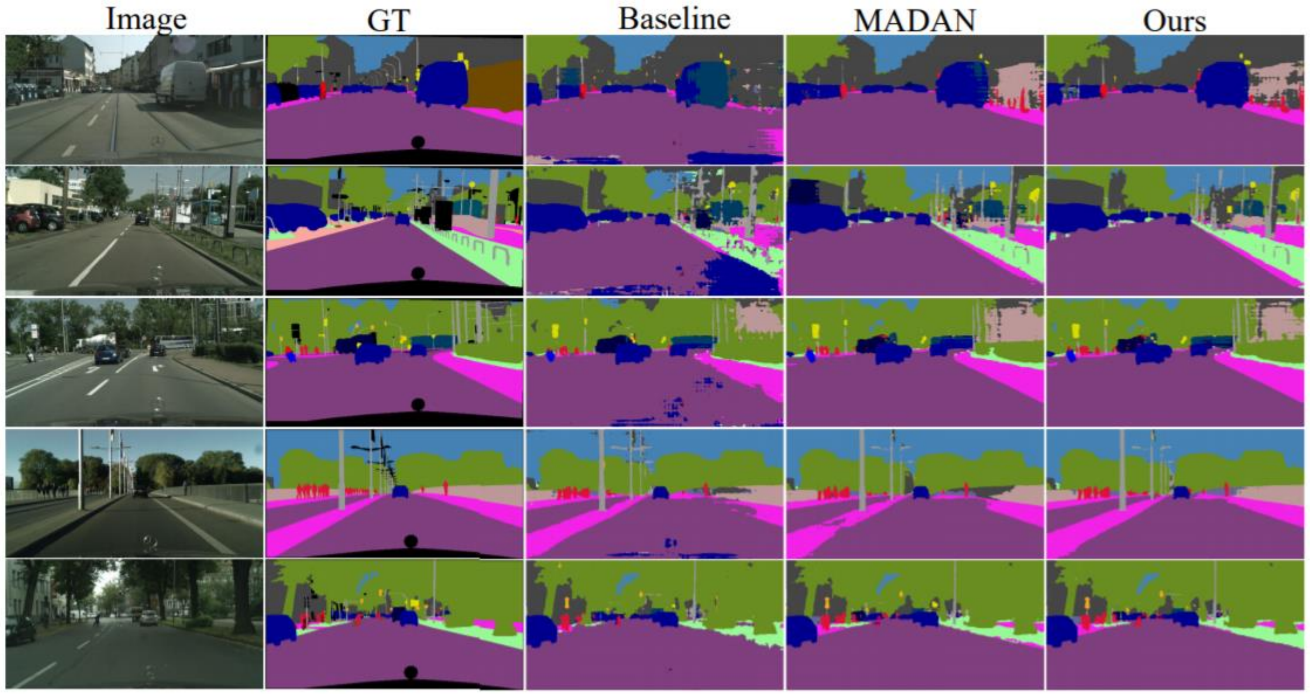


Figure 4. Visual Comparison with baseline and other methods. Left to right: input image from Cityscapes, corresponding ground-truth, segmentation results of baseline that simple combination of source domains, MADAN [43] and our proposed method. Note that, all these results are adapting from GTA5+Synscapes.

Table 5. The performance of our proposed method that uses different source domains for adaptation. **G: GTA5, S: Synscapes, Y: SYNTHIA.** mIoU19, mIoU16 and mIoU13 indicate performance on different number of categories.

	sources	mIoU19	mIoU16	mIoU13
Source-Only	G	39.53	43.28	48.25
	S	44.43	48.74	54.09
	Y	–	32.31	37.41
Multi-Sources	G+S	59.04	61.25	65.87
	G+Y	–	54.03	59.42
	S+Y	–	58.19	63.18
	G+S+Y	–	62.24	67.15