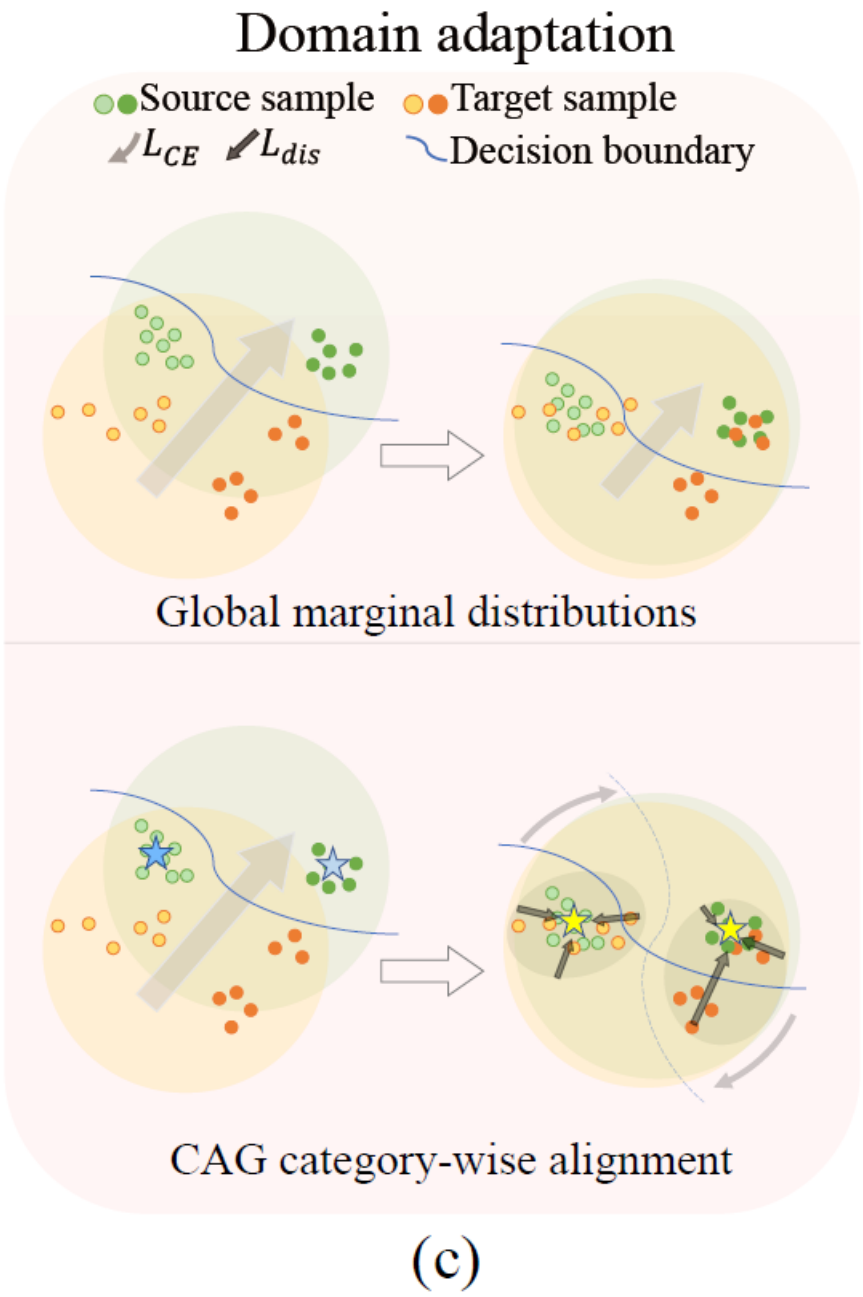
Domain Adaptation for Semantic Segmentation

Yichao Wang
2020.5.22

NIPS-2019

Category Anchor-Guided Unsupervised Domain Adaptation for Semantic Segmentation

Qiming Zhang[*1] **Jing Zhang**[*1] **Wei Liu**[2] **Dacheng Tao**[1]
[1]UBTECH Sydney AI Centre, School of Computer Science, Faculty of Engineering
The University of Sydney, Darlington, NSW 2008, Australia
[2]Tencent AI Lab, China
qzha2506@uni.sydney.edu.au, jing.zhang1@sydney.edu.au
wl2223@columbia.edu, dacheng.tao@sydney.edu.au



(c)

Shortcomings of previous methods

1. **Match the global marginal distribution**

It does not guarantee that samples from different categories in the target domain are properly separated, hence compromising the generalization ability.

2. **Self-supervised Learning**

Error-prone pseudo-labels will mislead the classifier and accumulate errors.

CAG-UDA

This paper propose a novel idea of category anchors, which facilitate both category-wise feature alignment and self training.

It is motivated by the observation that features from the same category tend to be clustered together.

Category anchor construction (CAC)

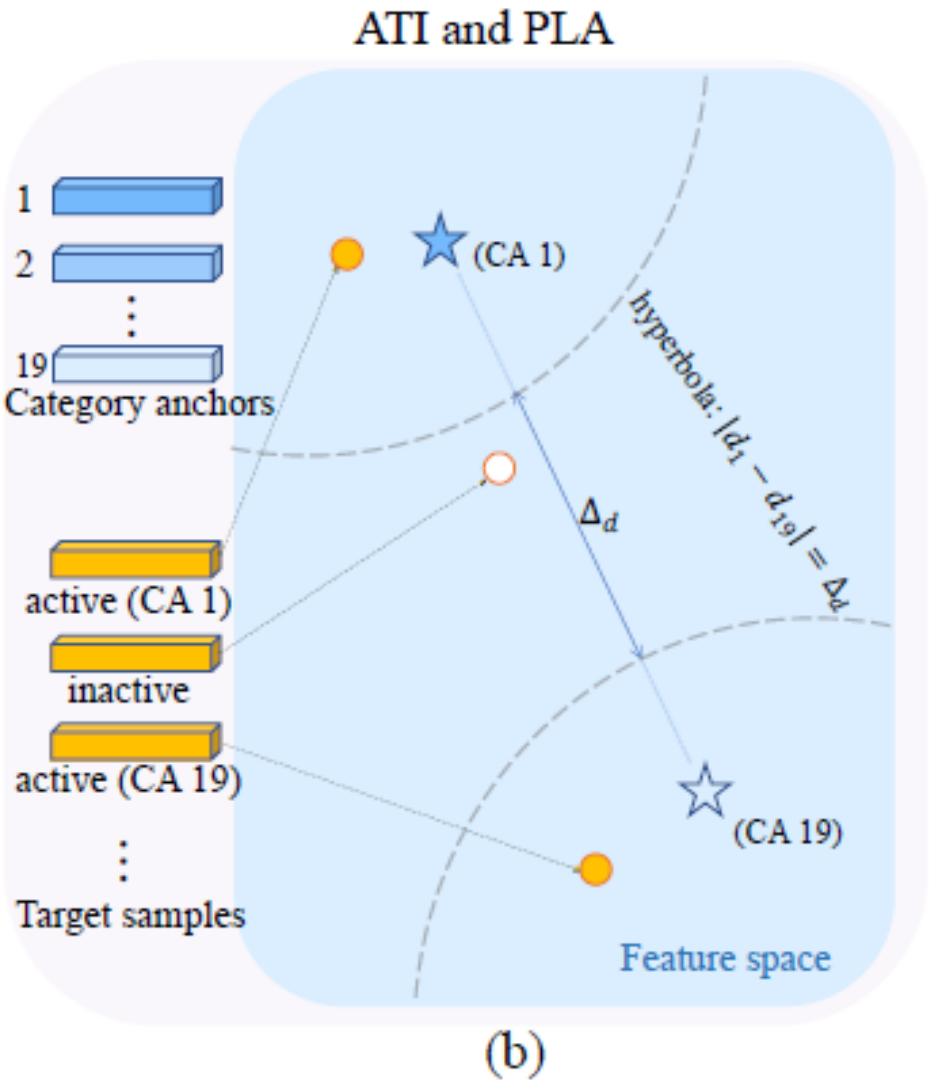
Calculate the centroids of the features of each category in the source domain as a representative of the feature distribution, $i.e.$ the mean.

$$f_c^s = \frac{1}{|\Lambda_c^s|} \sum_{i=1}^{N} \sum_{j}^{H \times W} y_{ijc}^s \left(f_D \left(Enc \left(x_i^s \right) \right) |_j \right),$$

Λ_c^s is the index set of all pixels on the training images in the source domain X_s belonging to the c^{th} category. $|\Lambda_c^s|$ denotes the number of pixels in Λ_c^s

Calculate the category centroids at the beginning of each training stage and then **keep them fixed** during training.

Active target sample identification (ATI)



(b)

The term “active target samples” refers to target samples **near one category anchor and far from the other anchors**, $i.e.$, being activated by one specific category anchor.

Distance between a target feature $f_D\left(Enc\left(x_i^t\right)\right)|_j$ and the c^{th} category anchor as:

$$d_{ijc}^t = \left\| f_c^s - f_D\left(Enc\left(x_i^t\right)\right)|_j \right\|_2,$$

we sort $\{d_{ijc}^t, c = 1, \ldots, C\}$ in an ascending order and compare the shortest distance $d_{ijc^*}^t$ with the second shortest $d_{ijc'}^t$.
If their difference is larger than a predefined threshold Δd, we identify this target sample as active one. $i.e.$

$$a_{ij}^t = \begin{cases} 1, & d_{ijc'}^t - d_{ijc^*}^t > \triangle_d, \\ 0, & otherwise, \end{cases}$$

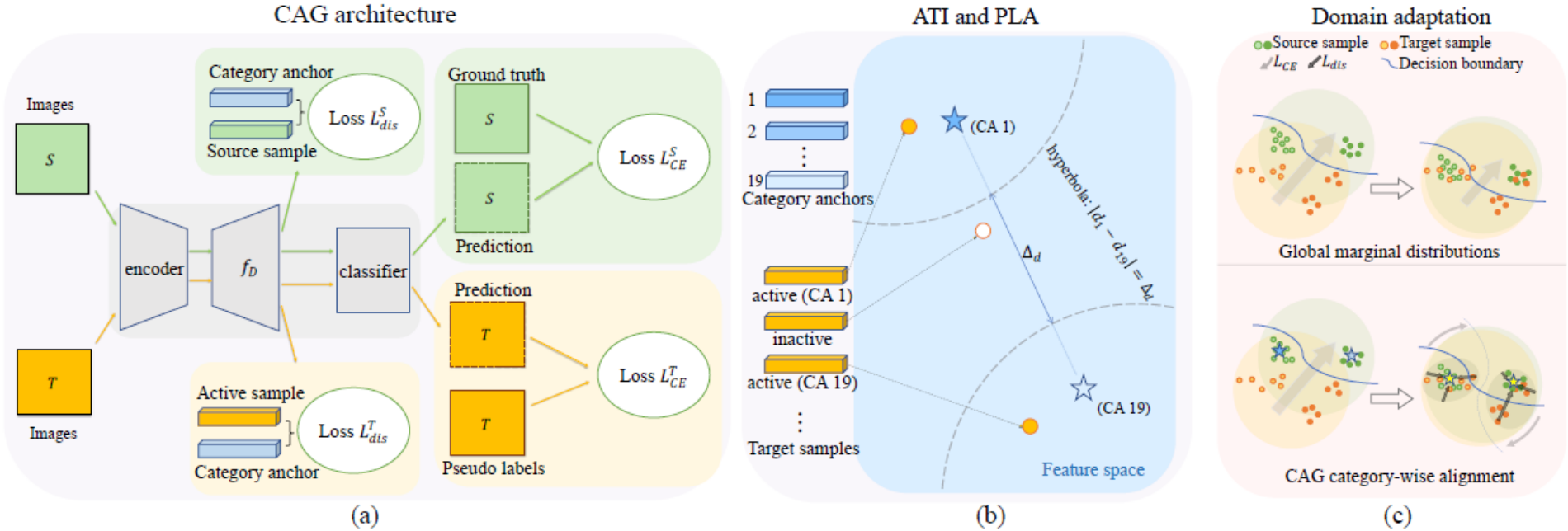


Figure 1: An illustration of the proposed category anchor-guided UDA model for semantic segmentation. (a) The architecture of the proposed CAG-UDA model consists of an encoder, a feature transformer (f_D), and a classifier. The green part denotes the source domain flow while the orange parts represent the target domain flow. (b) The illustration of the process of active target sample identification and pseudo label assignment described in Section 3.2. (c) The illustration of the proposed category-wise feature alignment with the anchor-based pixel-level distance loss L_{dis} and cross-entropy loss L_{CE} described in Section 3.3. Best viewed in color.

Objective Functions

$$L = L^s_{CE} + \lambda_1 \left(L^s_{dis} + L^t_{dis} \right) + \lambda_2 \left(L^t_{CE} + L^{tP}_{CE} \right),$$

Stagewise Training Procedure

1. pretrain the segmentation model on the source domain.
2. leverage the global feature alignment method to warm up the training process and obtain a well-initialized model.
3. train the CAG-UDA model with the proposed losses for several stages.

At the beginning of each stage, we calculate the CAs, identify the active target samples, and assign pseudo-labels to them.

Algorithm 1 Stagewise training the CAG-UDA model

Input: training dataset: (X_s, Y_s, X_t), maximum stages: K, maximum iterations: L, distance threshold: $\triangle_d$.

Output: M_K and $(\hat{Y}_s, \hat{Y}_t)$.

Pretraining: $M_0^p \leftarrow (X_s, Y_s)$ according to [4];
Warm-up: $M_0 \leftarrow (X_s, Y_s)$ and M_0^p according to [14];
for $k \leftarrow 1$ *to* K **do**
    CAC: $\{f_c^s\} \leftarrow M_{k-1}$ and (X_s, Y_s) according to Eq. (4);
    ATI: $\{d_{ijc}^t\}, \{a_{ij}^t\} \leftarrow M_{k-1}, (X_s, Y_s, X_t), \{f_c^s\}$ and $\triangle_d$ according to Eq. (5) and Eq. (6);
    PLA: $\{\hat{y}_{ijc^*}^t\} \leftarrow \{d_{ijc}^t\}, \triangle_d$ according to Eq. (7);
    for $n \leftarrow 1$ *to* L **do**
        SGD: training M_{k-1} on $(X_s, Y_s, X_t, \{\hat{y}_{ijc^*}^t\}, \{f_c^s\}, \{a_{ij}^t\})$ according to Eq.(12);
    end for
    $M_k \leftarrow M_{k-1}$
end for
Prediction: $(\hat{Y}_s, \hat{Y}_t) \leftarrow (X_s, X_t)$ and M_K.

Experiment

Table 1: Results of the CAG-UDA model and SOTA methods (GTA5→Cityscapes).

	road	sidewalk	building	wall	fence	pole	light	sign	vege.	terrace	sky	person	rider	car	truck	bus	train	motor	bike	mIoU
Source only	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
AdaptSegNet[36]	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
Source only	69.9	22.3	75.6	15.8	20.1	18.8	28.2	17.1	75.6	8.00	73.5	55.0	2.9	66.9	34.4	30.8	0.00	18.4	0.00	33.3
DCAN[40]	85.0	30.8	81.3	25.8	21.2	22.2	25.4	26.6	83.4	36.7	76.2	58.9	24.9	80.7	29.5	42.9	2.50	26.9	11.6	41.7
Source only	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
CLAN[26]	87.0	27.1	79.6	27.3	23.3	28.3	35.5	24.2	83.6	27.4	74.2	58.6	28.0	76.2	33.1	36.7	6.7	**31.9**	31.4	43.2
AdvEnt[39]	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
DISE[2]	**91.5**	47.5	82.5	31.3	25.6	33.0	33.7	25.8	82.7	28.8	82.7	62.4	30.8	**85.2**	27.7	34.5	6.4	25.2	24.4	45.4
Cycada[13, 21]	86.7	35.6	80.1	19.8	17.5	38.0	**39.9**	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
Source only	69.0	12.7	69.5	9.9	19.5	22.8	31.7	15.3	73.9	11.3	67.2	54.7	23.9	53.4	29.7	4.6	11.6	26.1	32.5	33.6
BLF[21]	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
Source only	69.8	25.4	74.7	11.3	18.3	24.2	35.6	23.3	72.0	14.4	65.3	58.7	29.0	53.1	14.3	19.2	7.9	15.1	16.3	34.1
CAG-UDA	90.4	**51.6**	83.8	34.2	**27.8**	**38.4**	25.3	**48.4**	**85.4**	38.2	78.1	58.6	**34.6**	84.7	21.9	42.7	**41.1**	29.3	**37.2**	**50.2**

Table 2: Results of the CAG-UDA model on the testing set (GTA5→Cityscapes).

	road	sidewalk	building	wall	fence	pole	light	sign	vege.	terrace	sky	person	rider	car	truck	bus	train	motor	bike	mIoU
CAG-UDA	93.2	57.0	85.6	35.7	25.1	37.5	30.8	45.3	87.1	50.1	89.4	62.7	40.8	87.8	18.0	32.4	34.5	34.4	35.4	51.7

Ablation Study

Table 4: Results of ablation study (GTA5→Cityscapes).

	road	side.	buil.	wall	fenc.	pole	light	sign	vege.	terr.	sky	person	rider	car	truck	bus	train	motor	bike	mIoU	gain
Source only	69.8	25.4	74.7	11.3	18.3	24.2	35.6	23.3	72.0	14.4	65.3	58.7	29.0	53.1	14.3	19.2	7.9	15.1	16.3	34.1	-
Warm-up	88.4	45.2	82.0	30.1	22.0	35.4	36.7	23.7	82.7	27.6	70.8	51.4	26.9	81.5	14.5	25.0	21.4	13.0	7.9	41.4	7.3
$+L_{CE}^{tP}$	88.8	45.5	83.7	33.2	21.4	39.5	**40.0**	25.9	83.9	33.8	74.3	58.2	24.9	84.8	19.3	32.8	22.6	15.0	14.7	44.3	10.2
$+L_{CE}^{t}$	88.3	46.9	81.5	28.7	27.7	38.9	27.0	40.4	83.7	31.2	74.9	61.8	30.2	84.0	15.9	36.7	23.4	23.3	31.7	46.1	12.0
$+L_{dis}^{sP}+L_{dis}^{tP}$	89.4	40.1	81.8	31.0	22.6	39.9	41.2	23.2	83.0	28.3	68.5	54.5	23.8	85.7	21.5	25.6	0.7	13.9	8.5	41.2	7.1
$+L_{dis}^{s}+L_{dis}^{t}$	88.9	41.7	82.0	31.7	22.5	39.7	41.2	23.5	82.7	27.0	70.0	57.8	25.7	**85.8**	**21.9**	27.7	1.1	18.0	11.1	42.1	8.0
$+L_{dis}^{s}+L_{dis}^{t}+L_{CE}^{t}$	88.1	46.6	82.1	30.2	**28.4**	39.7	31.3	38.8	83.6	30.7	75.1	61.9	28.5	84.3	16.3	36.3	29.1	25.0	29.4	46.6	12.5
$+L_{CE}^{t}+L_{CE}^{tP}$	88.9	47.1	83.0	31.0	27.3	39.7	31.0	36.0	84.3	32.6	75.1	62.0	29.4	84.6	16.6	35.7	27.2	19.2	28.4	46.3	12.2
CAG-UDA (Stage 1)	88.8	47.5	83.6	31.7	**29.1**	39.7	34.4	35.6	84.4	33.0	76.8	**62.1**	28.2	84.5	17.2	35.2	32.0	25.8	27.6	47.2	13.1
CAG-UDA (Stage 2)	**90.4**	50.6	84.0	33.5	28.3	**39.9**	31.6	42.4	85.1	35.2	77.3	61.5	34.2	84.9	19.4	41.7	41.0	27.3	32.0	49.5	15.4
CAG-UDA (Stage 3)	**90.4**	**51.6**	**83.8**	**34.2**	27.8	38.4	25.3	**48.4**	**85.4**	**38.2**	**78.1**	58.6	**34.6**	84.7	**21.9**	**42.7**	**41.1**	**29.3**	**37.2**	**50.2**	**16.1**

Objective Functions

$$L = L_{CE}^{s} + \lambda_1 \left(L_{dis}^{s} + L_{dis}^{t} \right) + \lambda_2 \left(L_{CE}^{t} + L_{CE}^{tP} \right),$$

Rethinking

1. Fixed Anchor 设置是否合理？

   因为每次使用的都是上个stage产生的anchor，在下个stage训练时 fix住。但是网络是一直在更新的，得到的feature的 分布也是变化着的，在训练时fix住anchor是不是没办法反映出模型的更新后feature的分布?

2. 固定的Δd设置是否合理?

   假设三个class Δ, O, ★ 中Δ, O比较近，Δ,★两个离得比较远。

   * Δd比较小可以满足Δ, O，对于Δ,★会引入过多的noise，

   因为会引入比较多decision boundary附近的feature。

   * Δd比较大可以满足Δ,★， 对于Δ, O会导致active 的feature过少。

3. 忽略了intra-class 分布?

   每个class本身也有分布，所有class feature都映射到feature space的一个点，是不是没有必要的?

CVPR-2020

Unsupervised Intra-domain Adaptation for Semantic Segmentation through Self-Supervision

Fei Pan Inkyu Shin Francois Rameau Seokju Lee In So Kweon
KAIST, South Korea
{feipan, dlsrbgg33, frameau, seokju91, iskweon77}@kaist.ac.kr

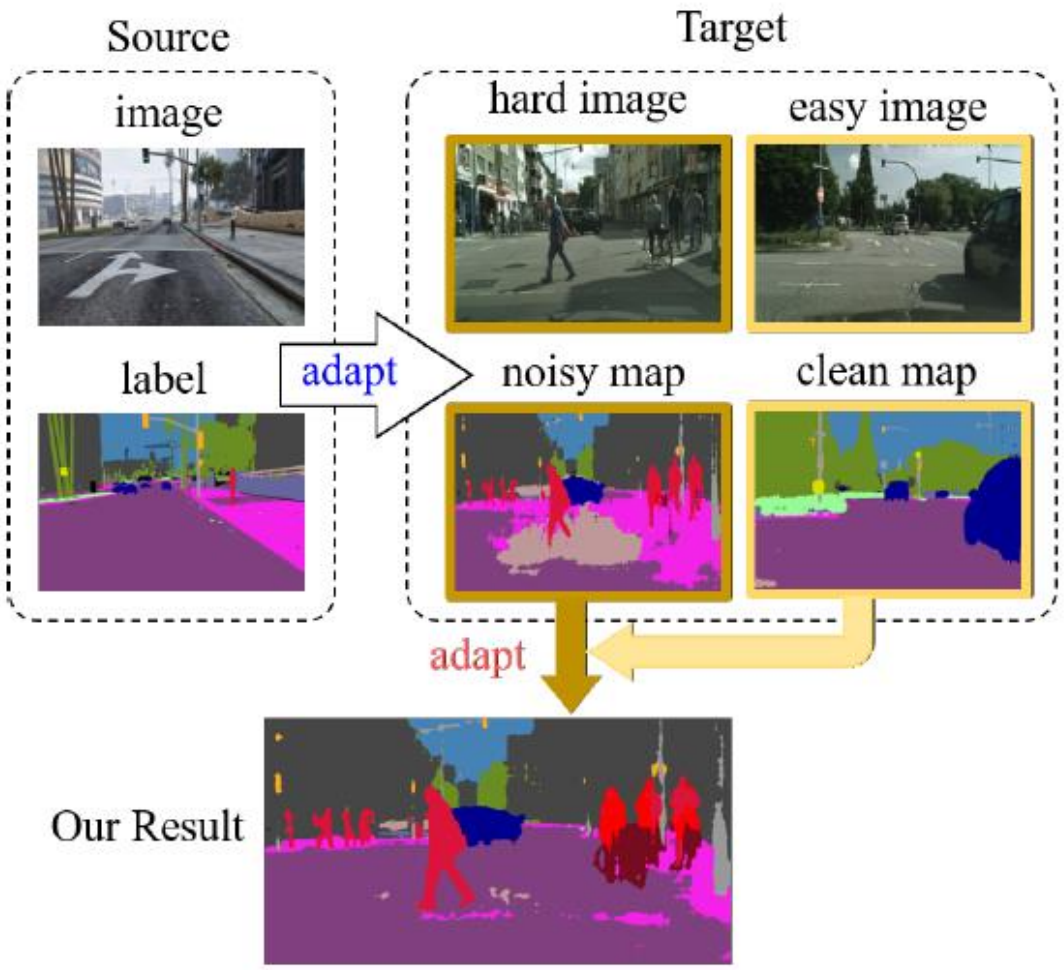
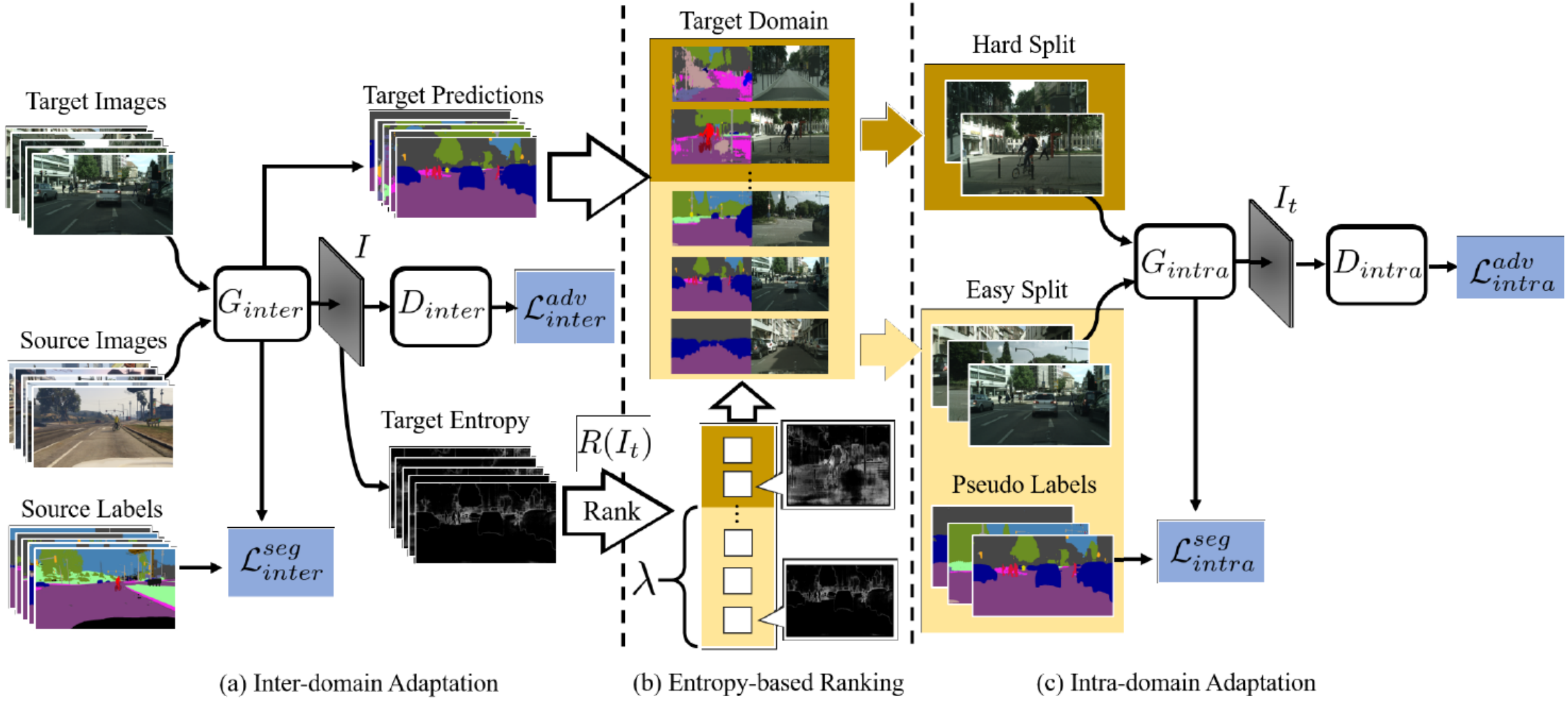


Figure 1: We propose a two-step self-supervised domain adaptation technique for semantic segmentation. Previous works solely adapt the segmentation model from the source domain to the target domain. Our work also consider adapting from the clean map to the noisy map within the target domain.

Target data collected from the real world have diverse scene distributions; these distributions are caused by various factors such as moving objects, weather conditions, which leads to a large gap in the target (**intra-domain gap**).

This paper present a two-step domain adaptation approach to minimize the **inter-domain** and **intra-domain** gap.



Model consists of three parts :

1. An **inter-domain adaptation module** to close inter-domain gap between the labeled source data and the unlabeled target data.
2. An **entropy-based ranking system** to separate target data into the an easy and hard split.
3. An **intra-domain adaptation module** to close intra-domain gap between the easy and hard split (using pseudo labels from the easy subdomain).

Experiment

(a) GTA5 → Cityscapes

Method	road	sidewalk	building	wall	fence	pole	light	sign	veg	terrain	sky	person	rider	car	truck	bus	train	mbike	bike	mIoU
Without adaptation [26]	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
ROAD [5]	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 [26]	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	42.4
MinEnt [29]	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 [29]	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	**90.6**	**37.1**	**82.6**	**30.1**	19.1	**29.5**	32.4	20.6	**85.7**	**40.5**	**79.7**	**58.7**	**31.1**	**86.3**	31.5	**48.3**	0.0	30.2	35.8	**46.3**

(b) SYNTHIA → Cityscapes

Method	road	sidewalk	building	wall*	fence*	pole*	light	sign	veg	sky	person	rider	car	bus	mbike	bike	mIoU	mIoU*
Without adaptation [26]	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
AdaptSegNet [26]	81.7	39.1	78.4	**11.1**	0.3	25.8	6.8	9.0	79.1	80.8	54.8	21.0	66.8	34.7	13.8	29.9	39.6	45.8
MinEnt [29]	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
AdvEnt [29]	**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	84.3	37.7	79.5	5.3	0.4	24.9	**9.2**	8.4	80.0	**84.1**	**57.2**	23.0	**78.0**	**38.1**	**20.3**	**36.5**	**41.7**	**48.9**

(c) Synscapes → Cityscapes

Method	road	sidewalk	building	wall	fence	pole	light	sign	veg	terrain	sky	person	rider	car	truck	bus	train	mbike	bike	mIoU
Without adaptation	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	78.4	20.4	20.5	7.8	27.3	52.5	45.3
AdaptSegNet [26]	**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
Ours	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**

Ablation Study

Table 2: The ablation study on hyperparameter λ for separating the target domain into the easy and the hard split.

GTA5 → Cityscapes						
λ	0.0	0.5	0.6	0.67	0.7	1.0
mIoU	43.8	45.2	46.0	**46.3**	45.6	45.5

Table 3: The self-training and intra-domain adaptation gain on GTA5 → Cityscapes.

Model	mIoU
AdvEnt [25]	43.8
AdvEnt + intra-domain adaptation	45.1
AdvEnt + self-training ($\lambda = 1.0$)	45.5
Ours	46.3
Ours + entropy normalization	**47.0**

Thank you