Domain Adaptive Semantic Segmentation and Image Classification

Guoliang Kang

Postdoctoral Research Associate

Carnegie Mellon University

kgl.prml@gmail.com

- **Introduction**
- Contrastive Adaptation Network
- Pixel-Level Cycle Association
- Summary

Deep learning for Computer Vision Tasks

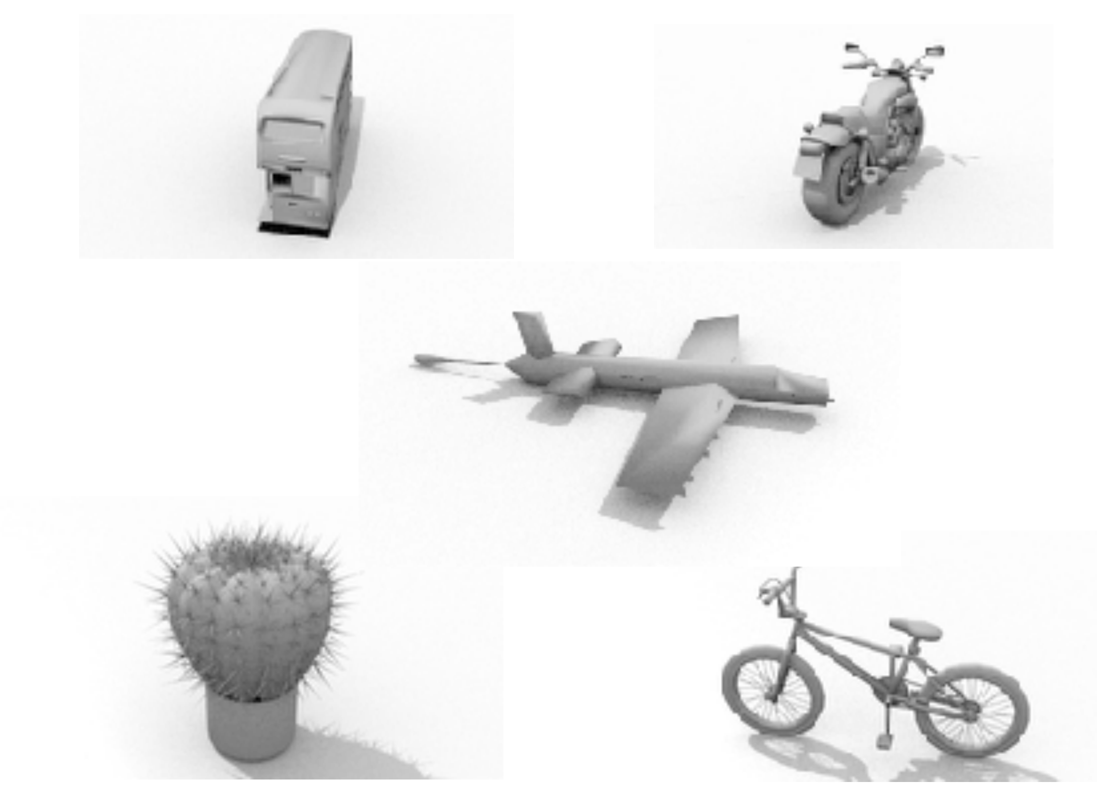
- Image Classification
- Semantic Segmentation
- Object Detection
- Tracking
- ……

ImageNet

Cross-Domain Prediction

- **The distribution of test data is different that of training data**

  Style, layout, shape, context, illumination, etc.



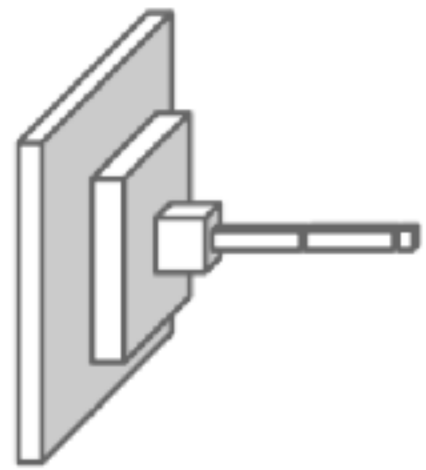
Training data

Test data

Cross-Domain Prediction

- **Performance degenerates due to the domain shift**

Motocycle



Deep Model

Bicycle ✗

Domain Adaptation

The setting of domain adaptation

- Target distribution is different from the source one
- Same task (shared label sets)
- Large amounts of labeled source data and unlabeled target data

Why do we need domain adaptation?

DomainNet

Discriminative Domain-Invariant Feature Learning

Through domain adaptation, we expect the learned features satisfy:

- Domain-Invariant: indistinguishable from features
- Discriminative: good inter-class separability and high intra-class compactness

Discriminative Domain-Invariant Feature Learning

Through domain adaptation, we expect the learned features satisfy:

- Domain-Invariant: indistinguishable from features
- Discriminative: good inter-class separability and high intra-class compactness

Conventional way to learn domain-invariant features

- Ground-truth supervision from source data
- Sharing network parameters

Domain Discrepancy Minimization

image style transfer; adversarial loss; Maximum Mean Discrepancy (MMD); etc.

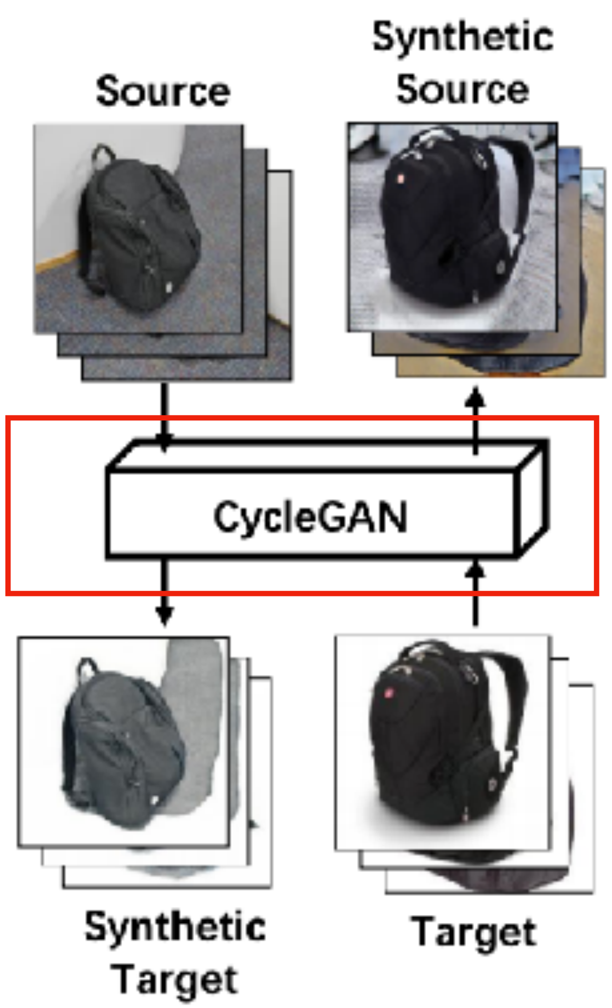
Consistency Regularization

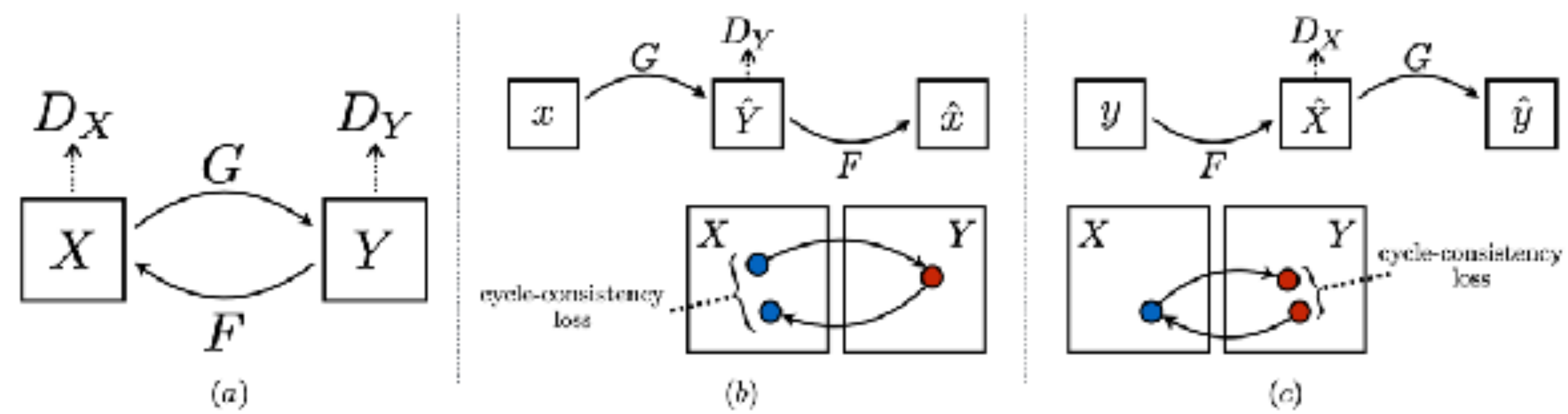
self-ensemble method; attention alignment; etc.

Self-training based methods

Domain Discrepancy Minimization

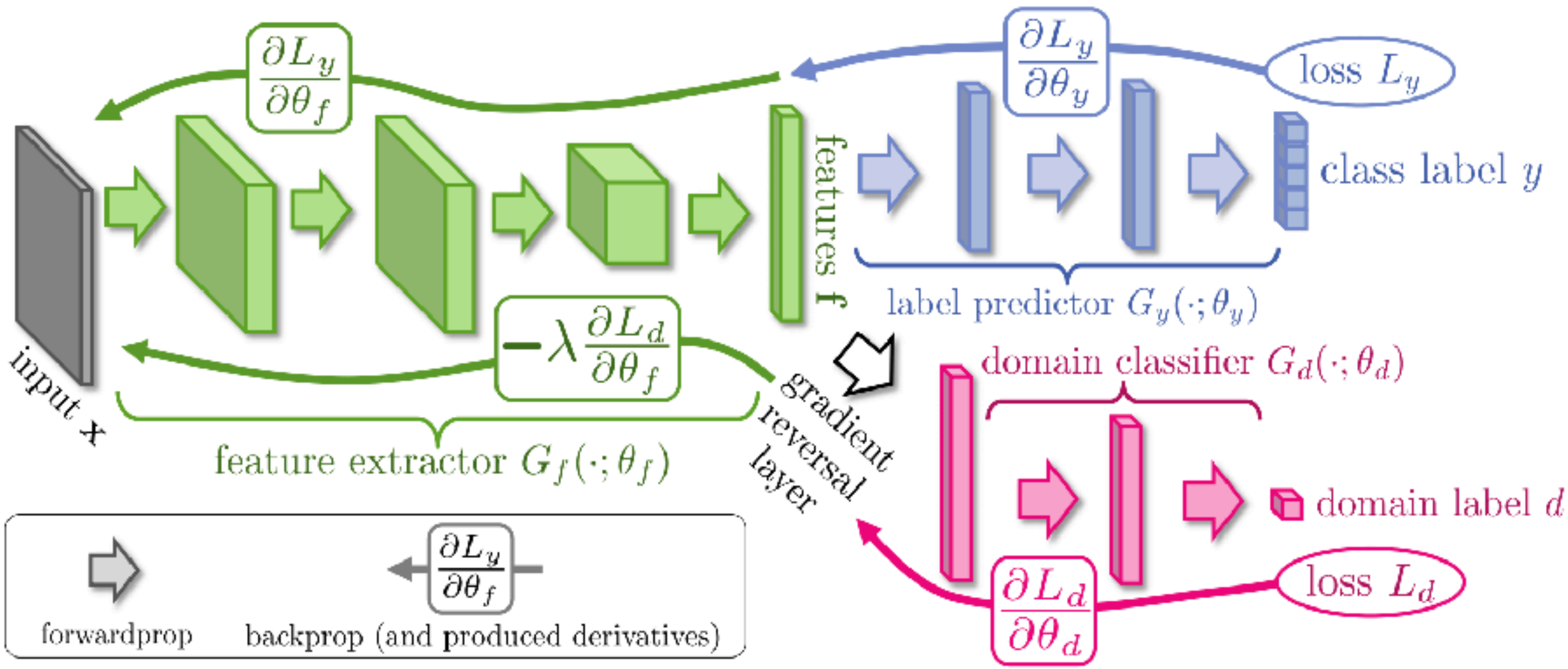
Style Transfer





Domain Discrepancy Minimization

Adversarial Loss / Reverse Gradient

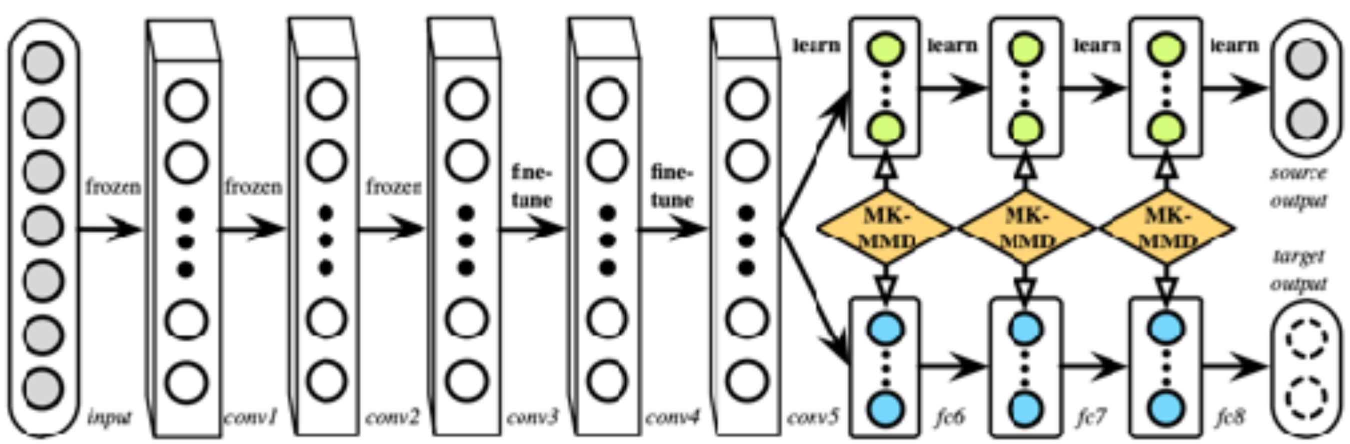


[1] Ganin, Yaroslav, and Victor Lempitsky. "Unsupervised domain adaptation by backpropagation."ICML, 2015.

Domain Discrepancy Minimization

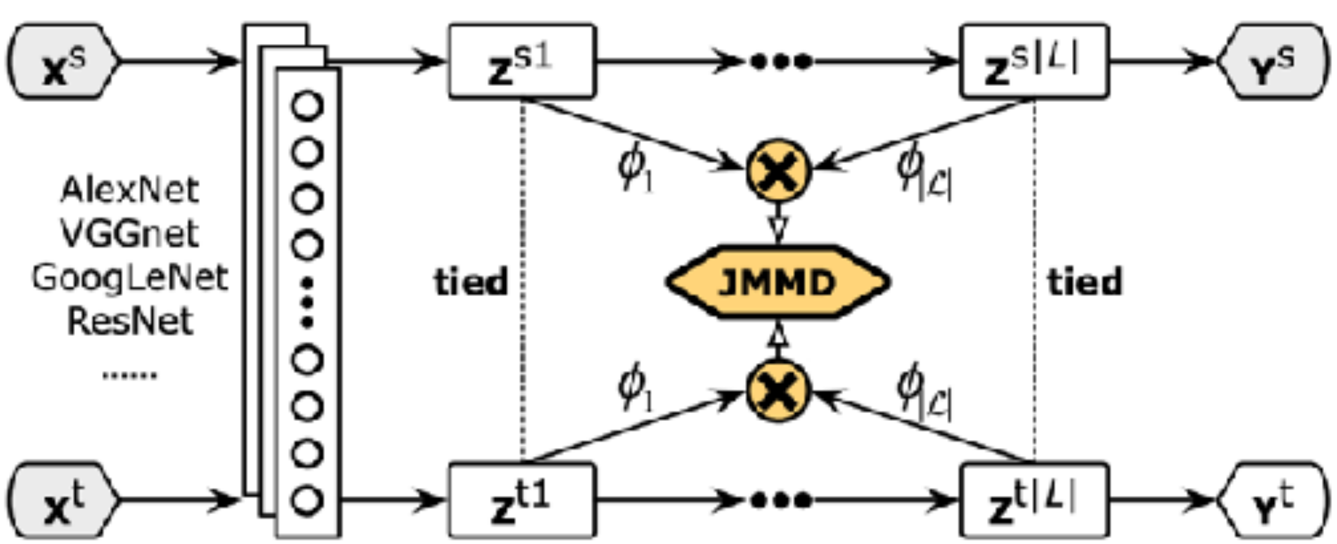
Maximum Mean Discrepancy (MMD) Based

$$\mathcal{D}_{\mathcal{H}}(P, Q) \triangleq \sup_{f \sim \mathcal{H}} \left(\mathbb{E}_{\boldsymbol{X}^s}[f(\boldsymbol{X}^s)] - \mathbb{E}_{\boldsymbol{X}^t}[f(\boldsymbol{X}^t)] \right)_{\mathcal{H}}$$



DAN

$$\begin{aligned}\widehat{D}_{\mathcal{L}}(P, Q) = &\ \frac{1}{n_s^2} \sum_{i=1}^{n_s} \sum_{j=1}^{n_s} \boxed{\sum_{\ell \in \mathcal{L}}} k^\ell \left(\mathbf{z}_i^{s\ell}, \mathbf{z}_j^{s\ell} \right) \\ &+ \frac{1}{n_t^2} \sum_{i=1}^{n_t} \sum_{j=1}^{n_t} \boxed{\sum_{\ell \in \mathcal{L}}} k^\ell \left(\mathbf{z}_i^{t\ell}, \mathbf{z}_j^{t\ell} \right) \\ &- \frac{2}{n_s n_t} \sum_{i=1}^{n_s} \sum_{j=1}^{n_t} \boxed{\sum_{\ell \in \mathcal{L}}} k^\ell \left(\mathbf{z}_i^{s\ell}, \mathbf{z}_j^{t\ell} \right)\end{aligned}$$



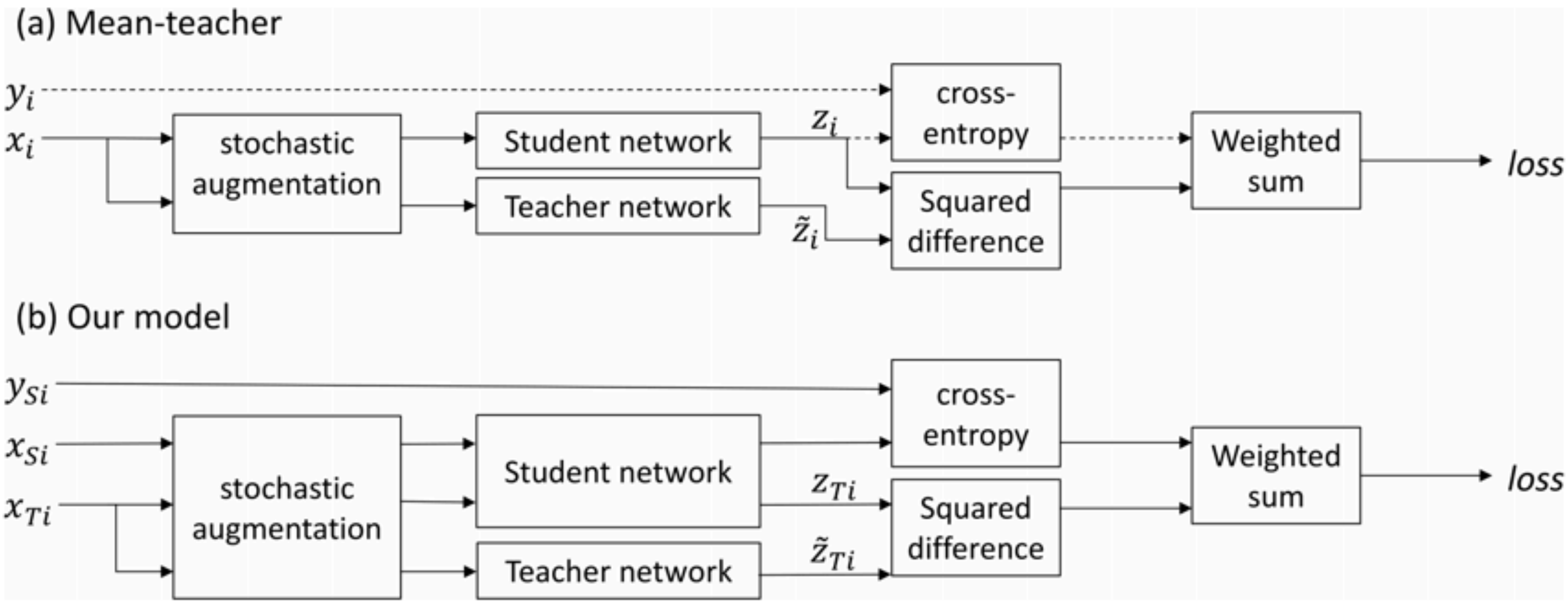
JAN

$$\begin{aligned}\widehat{D}_{\mathcal{L}}(P, Q) = &\ \frac{1}{n_s^2} \sum_{i=1}^{n_s} \sum_{j=1}^{n_s} \boxed{\prod_{\ell \in \mathcal{L}}} k^\ell \left(\mathbf{z}_i^{s\ell}, \mathbf{z}_j^{s\ell} \right) \\ &+ \frac{1}{n_t^2} \sum_{i=1}^{n_t} \sum_{j=1}^{n_t} \boxed{\prod_{\ell \in \mathcal{L}}} k^\ell \left(\mathbf{z}_i^{t\ell}, \mathbf{z}_j^{t\ell} \right) \\ &- \frac{2}{n_s n_t} \sum_{i=1}^{n_s} \sum_{j=1}^{n_t} \boxed{\prod_{\ell \in \mathcal{L}}} k^\ell \left(\mathbf{z}_i^{s\ell}, \mathbf{z}_j^{t\ell} \right)\end{aligned}$$

[1] Long, Mingsheng, et al. "Learning transferable features with deep adaptation networks." ICML, 2015.
[2] Long, Mingsheng, et al. "Deep transfer learning with joint adaptation networks." ICML, 2017.

Consistency Regularization

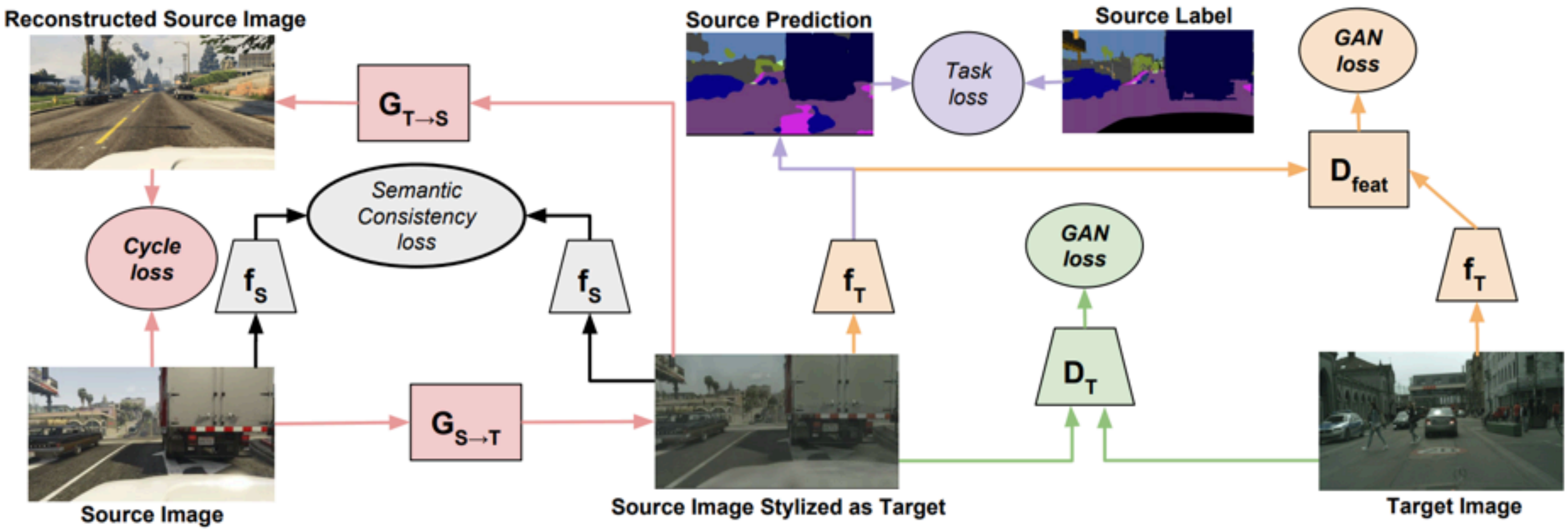
Self-ensembling



[1] French, Geoffrey, Michal Mackiewicz, and Mark Fisher. "Self-ensembling for visual domain adaptation." ICLR, 2017.

Consistency Regularization

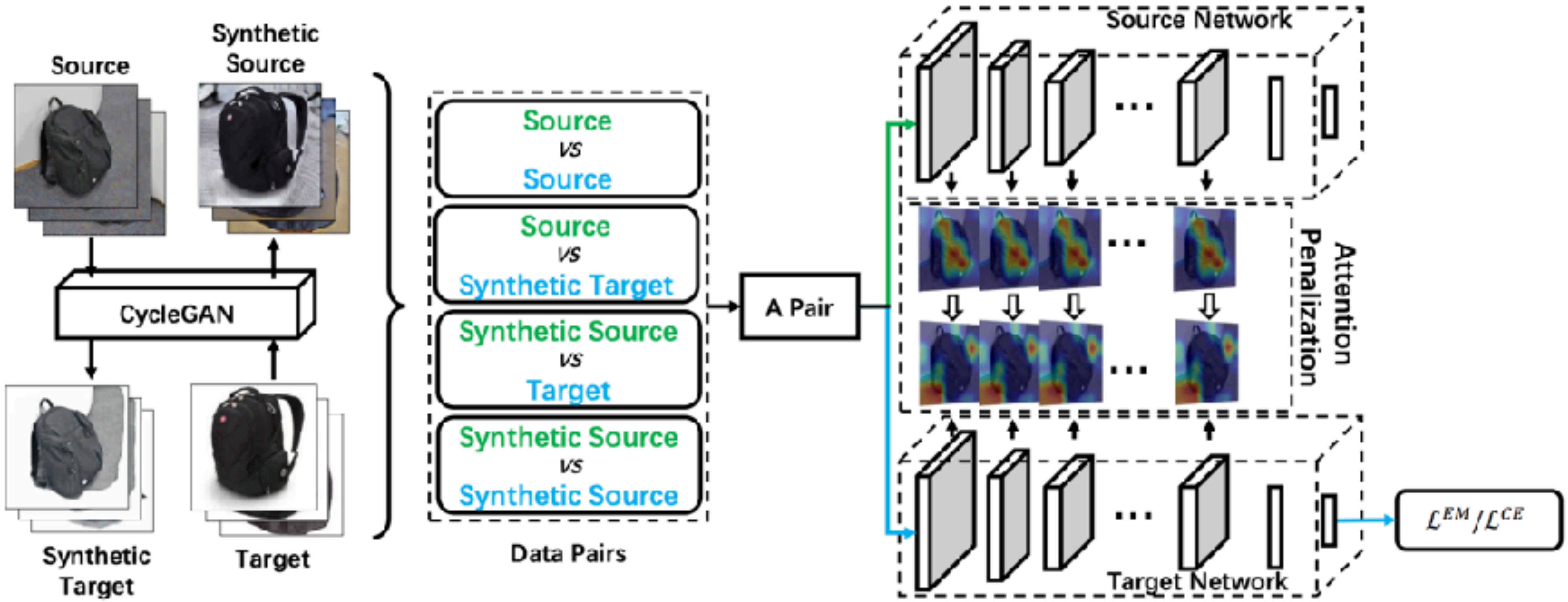
Style transfer + adversarial training + semantic consistency



[1] Hoffman, Judy, et al. "Cycada: Cycle-consistent adversarial domain adaptation." ICML, 2018.

Consistency Regularization

Attention Alignment



[1] Kang, Guoliang, et al. "Deep adversarial attention alignment for unsupervised domain adaptation: the benefit of target expectation maximization." ECCV. 2018..

Discriminative Domain-Invariant Feature Learning

Through domain adaptation, we expect the learned features satisfy:

- Domain-Invariant: indistinguishable from features
- **Discriminative: good inter-class separability and high intra-class compactness**

Contrastive Adaptation Network for the Image Classification

- Class-aware alignment vs. Class-agnostic alignment (previous)

[1] Kang, Guoliang, et al. "Contrastive adaptation network for unsupervised domain adaptation." CVPR. 2019.

[2] Kang, Guoliang, et al. "Contrastive adaptation network for single-and multi-source domain adaptation." IEEE TPAMI (2020).

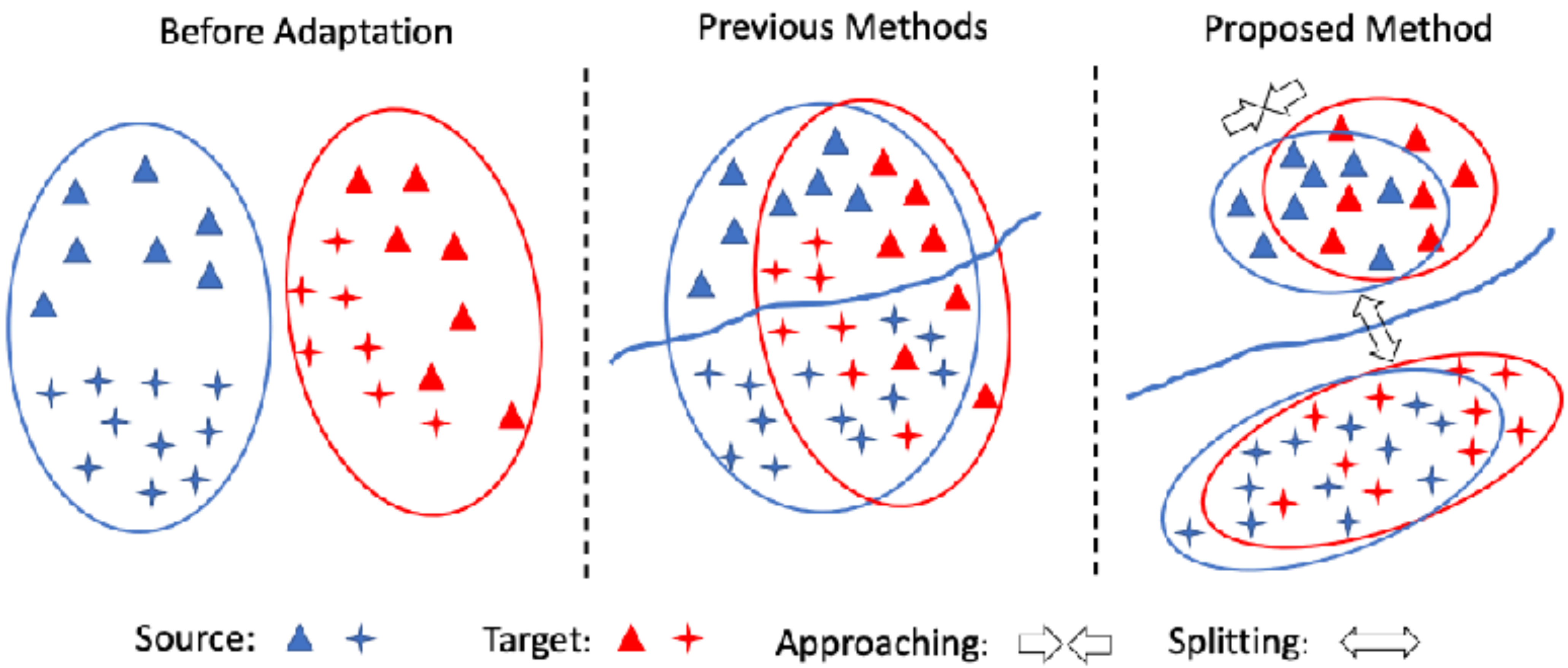
Pixel-Level Cycle Association for Domain Adaptive Semantic Segmentation

- Align semantic-consistent pixel pairs vs. Align globally (previous)

[3] Kang, Guoliang, et al. "Pixel-Level Cycle Association: A New Perspective for Domain Adaptive Semantic Segmentation." NeurIPS (2020).

- Introduction
- **Contrastive Adaptation Network**
- Pixel-Level Cycle Association
- Summary

Motivation



Class-aware alignment

Contrastive Domain Discrepancy

MMD measuring conditional distribution discrepancy

$$\mathcal{D}_{\mathcal{H}}(P,Q) \triangleq \sup_{f \sim \mathcal{H}} \left(\mathbb{E}_{\boldsymbol{X}^s}[f(\phi(\boldsymbol{X}^s) \quad)] - \mathbb{E}_{\boldsymbol{X}^t}[f(\phi(\boldsymbol{X}^t) \quad)]\right)_{\mathcal{H}}$$

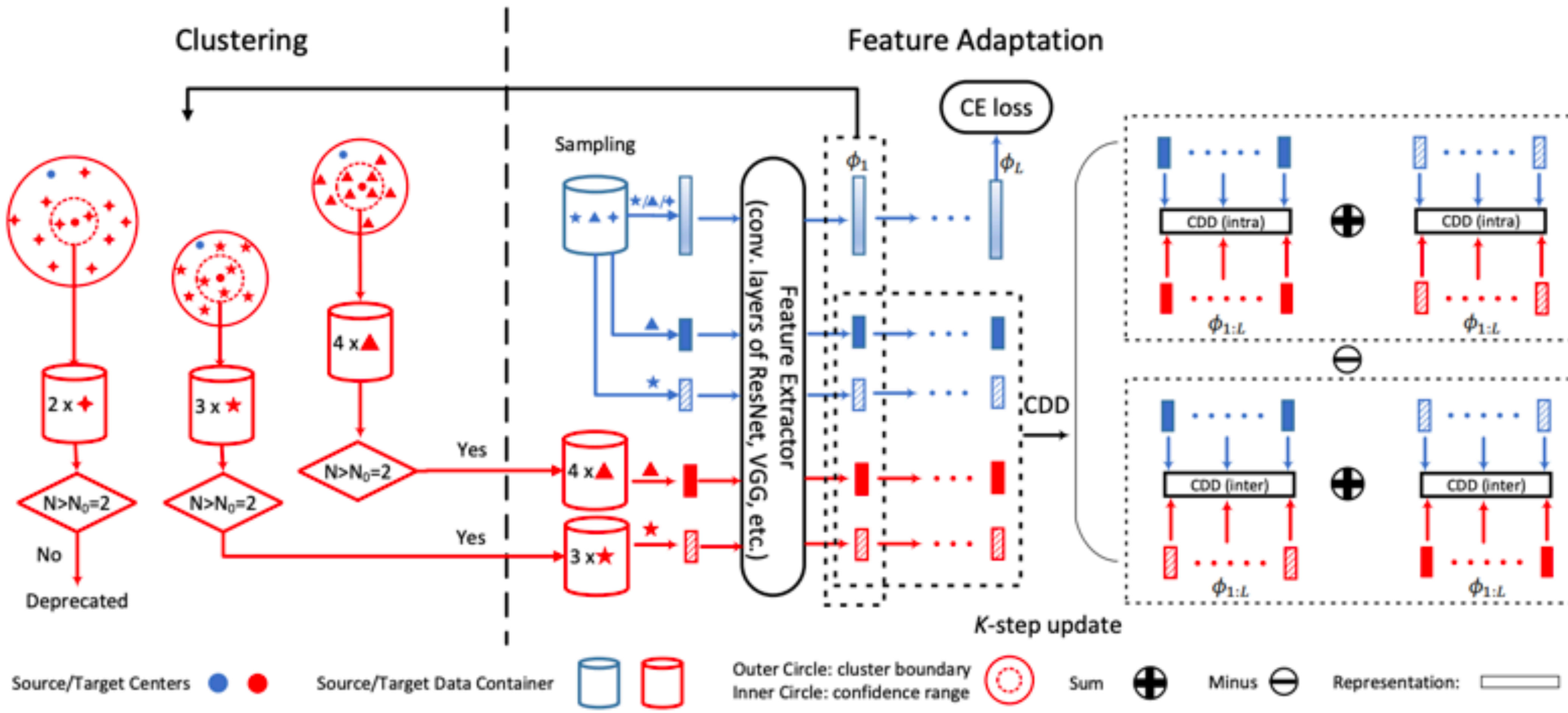
Contrastive Domain Discrepancy (CDD)

$$\hat{\mathcal{D}}^{cdd} = \underbrace{\frac{1}{M}\sum_{c=1}^{M} \hat{\mathcal{D}}^{cc}(\boxed{\hat{y}^t_{1:n_t}}, \boxed{\phi})}_{intra} - \underbrace{\frac{1}{M(M-1)}\sum_{c=1}^{M}\sum_{\substack{c'=1 \\ c' \neq c}}^{M} \hat{\mathcal{D}}^{cc'}(\boxed{\hat{y}^t_{1:n_t}}, \boxed{\phi})}_{inter}$$

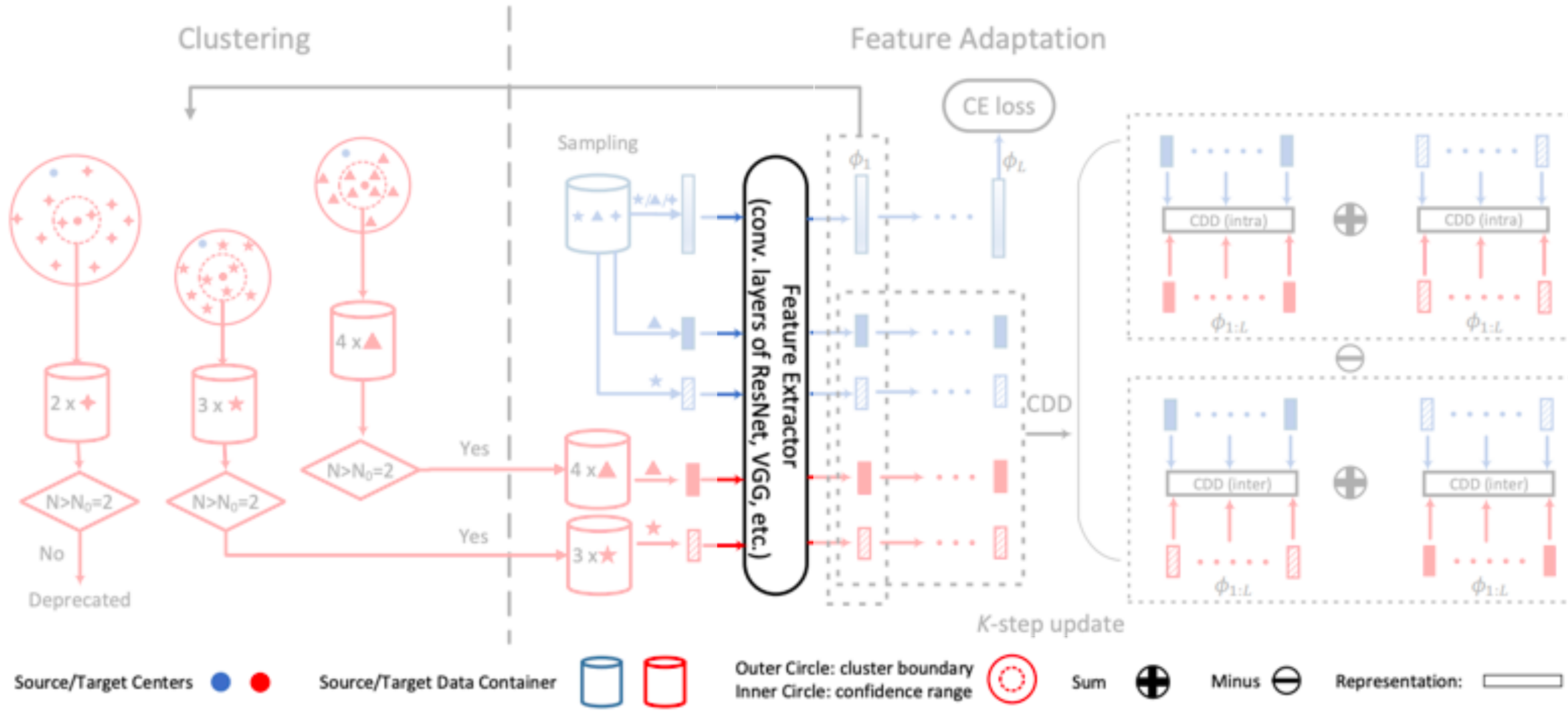
Intra: The MMD distance between cross-domain distributions conditioned on the same class.

Inter: The MMD distance between cross-domain distributions conditioned on different classes.

Contrastive Adaptation Network

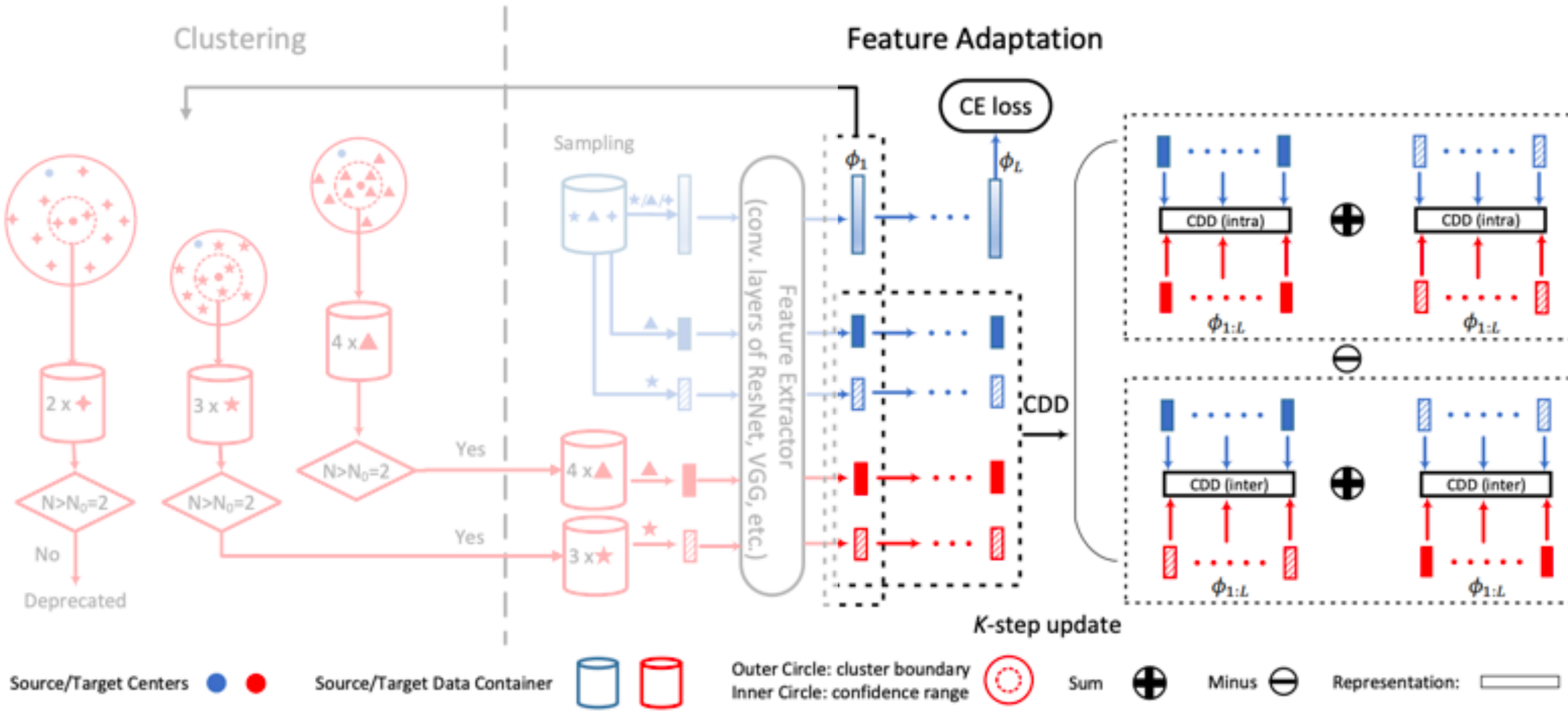


Contrastive Adaptation Network



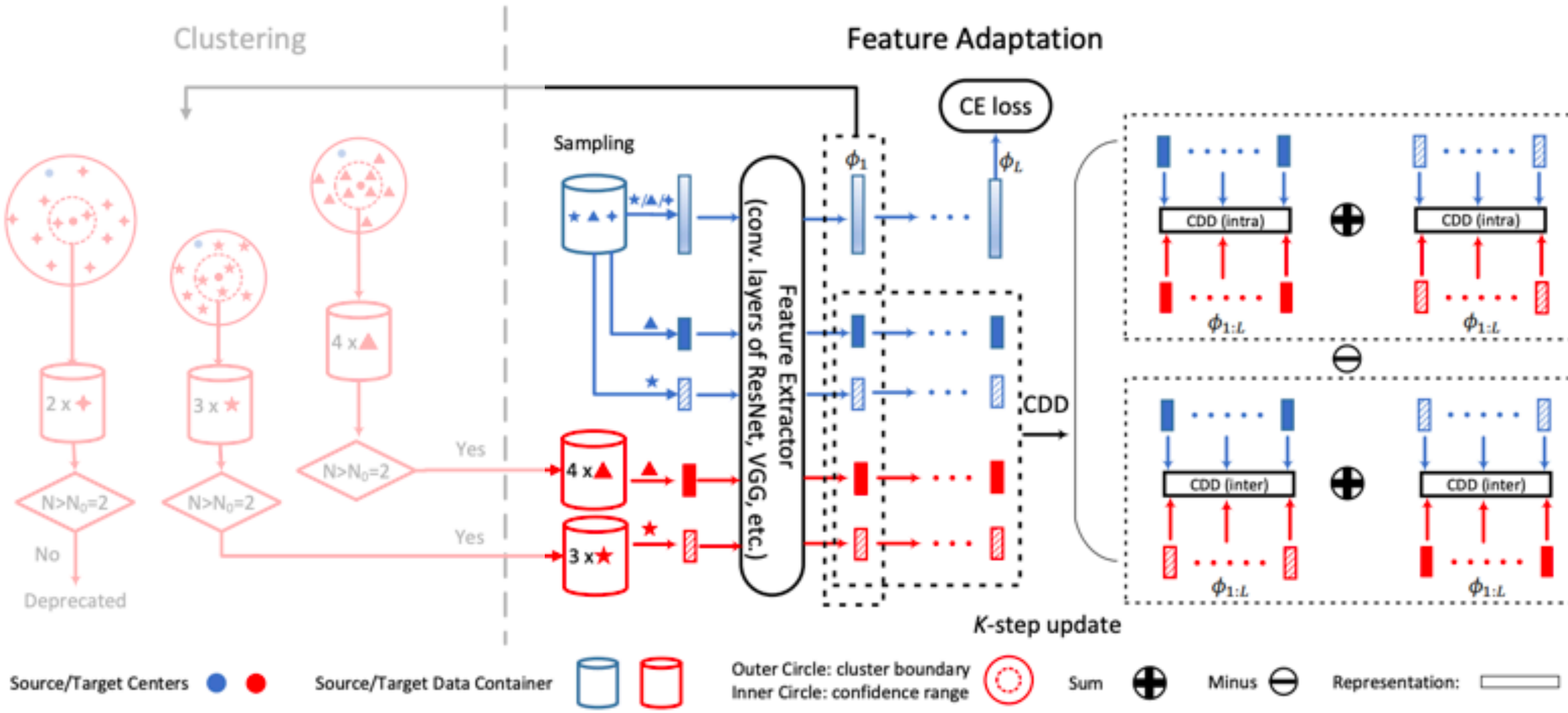
- **ImageNet pertained weights to initialize the backbone**

Contrastive Adaptation Network



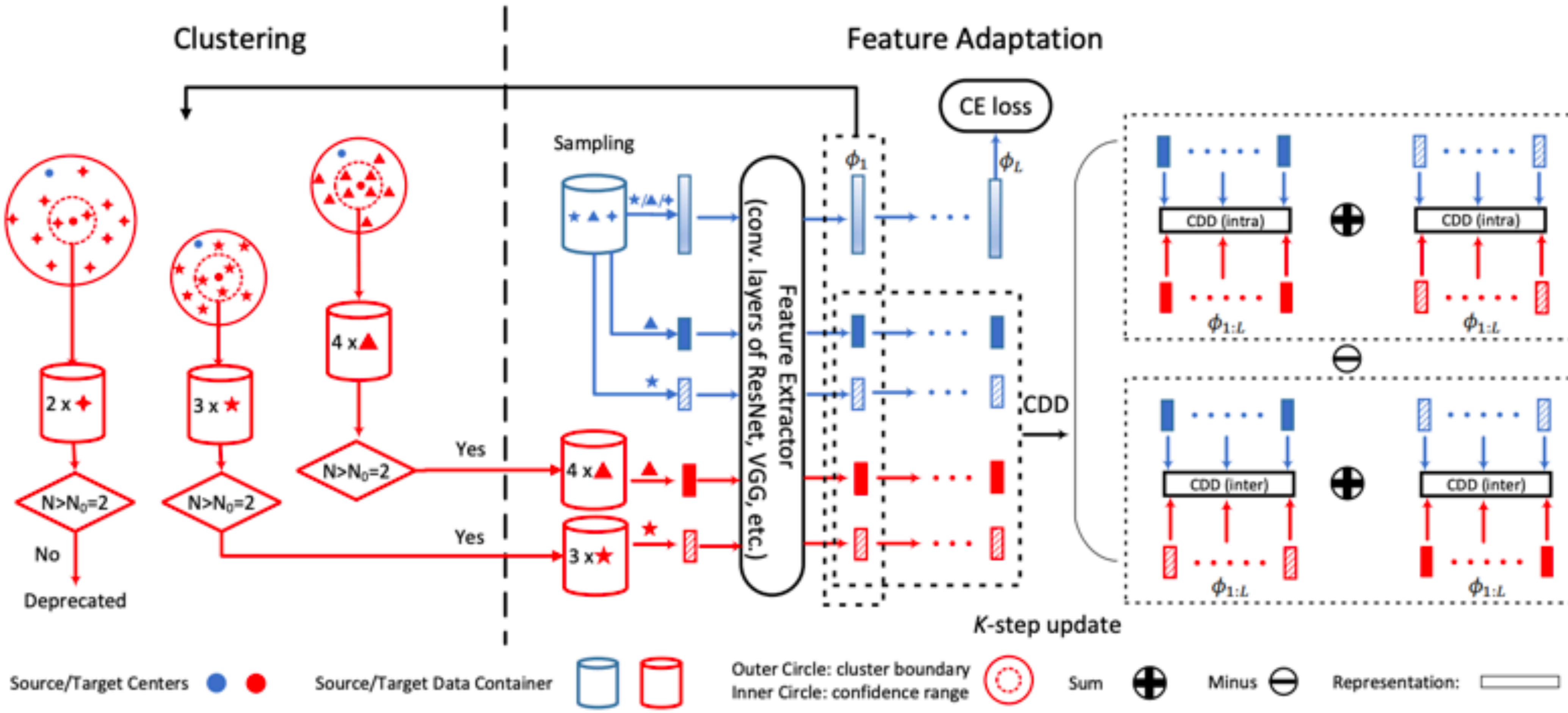
- **ImageNet pertained weights to initialize the backbone**
- **Align multiple Fully-Connected layers (including the final outputs)**

Contrastive Adaptation Network



- **ImageNet pertained weights to initialize the backbone**
- **Align multiple Fully-Connected layers (including the final outputs)**
- **Overall Objective** $\min_{\theta} \ell = \ell^{ce} + \beta \hat{\mathcal{D}}^{cdd}_{\mathcal{L}}$ where $\hat{\mathcal{D}}^{cdd}_{\mathcal{L}} = \sum_{l=1}^{L} \hat{\mathcal{D}}^{cdd}_{l}$

Contrastive Adaptation Network



- **ImageNet pertained weights to initialize the backbone**
- **Align multiple Fully-Connected layers (including the final outputs)**
- **Overall Objective** $\min_{\theta} \ell = \ell^{ce} + \beta \hat{\mathcal{D}}^{cdd}_{\mathcal{L}}$ where $\hat{\mathcal{D}}^{cdd}_{\mathcal{L}} = \sum_{l=1}^{L} \hat{\mathcal{D}}^{cdd}_{l}$

Generate Target Label Hypotheses

Motivation/Assumption

- The data from different categories is less likely to concentrate
- The peaks of target feature distribution are good representatives for the underlying categories.

Initialize with Source Centers

$$O^{tc} \leftarrow O^{sc} = \sum_{i=1}^{N_s} \mathbf{1}_{y_i^s=c} \frac{\phi_1(\boldsymbol{x}_i^s)}{\|\phi_1(\boldsymbol{x}_i^s)\|}$$

Iterative Refinement via Spherical K-means Clustering

$$\boxed{\text{Attach Target Labels}} \rightleftarrows \boxed{\text{Update Target Centers}}$$

$$\hat{y}_i^t \leftarrow \mathrm{argmin}_c\, dist(\phi_1(\boldsymbol{x}_i^t), O^{tc})$$

$$O^{tc} \leftarrow \sum_{i=1}^{N_t} \mathbf{1}_{\hat{y}_i^t=c} \frac{\phi_1(\boldsymbol{x}_i^t)}{\|\phi_1(\boldsymbol{x}_i^t)\|}$$

Filtering

The ambiguous target data (i.e. far from the cluster centers) and ambiguous classes (i.e. containing few target samples around the cluster centers) are zeroed out in estimating the CDD.

Alternative Optimization

- **Algorithm**

The loop of AO is repeated multiple times in our experiments.
Asynchronously update of the target labels and the network parameters.

Algorithm 1: Optimization of CAN at loop T_e.

Input:
source data: $\mathcal{S} = \{(\boldsymbol{x}_1^s, y_1^s), \cdots, (\boldsymbol{x}_{N_s}^s, y_{N_s}^s)\}$,
target data: $\mathcal{T} = \{\boldsymbol{x}_1^t, \cdots, \boldsymbol{x}_{N_t}^t\}$

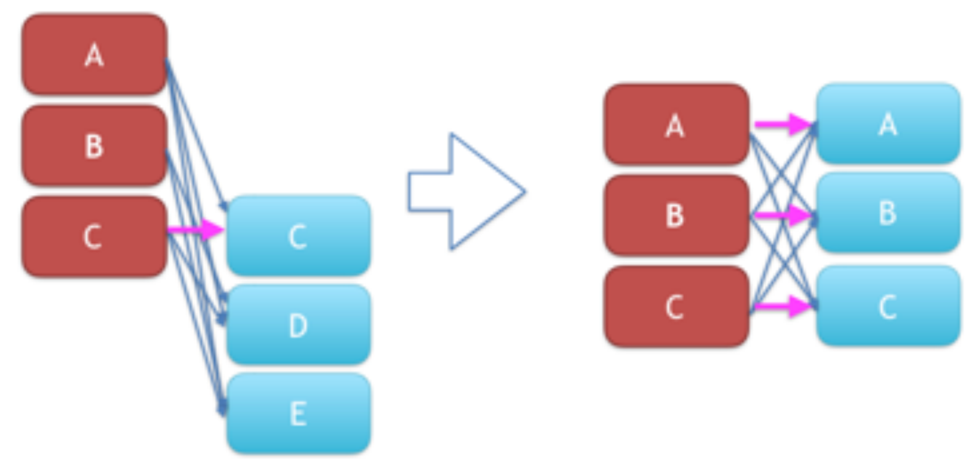
Procedure:

Clustering
- Forward $\mathcal{S}$ and compute the M cluster centers O^{sc} ;
- Initialize O^{tc}: $O^{tc} \leftarrow O^{sc}$;
- Cluster target samples $\mathcal{T}$ using spherical K-means;
- Filter the ambiguous target samples and classes;

for $(k \leftarrow 1; k \le K; k \leftarrow k+1)$ **do**

Feature Adaptation
- $\boxed{\text{Class-aware sampling}}$ based on $\mathcal{C}'_{T_e}$, $\tilde{\mathcal{T}}$, and $\mathcal{S}$;
- Compute $\hat{\mathcal{D}}_{\mathcal{L}}^{cdd}$ using Eq. (6);
- Sample from $\mathcal{S}$ and compute ℓ^{ce} using Eq. (7);
- Back-propagate with the objective ℓ (Eq.(8));
- Update network parameters θ.

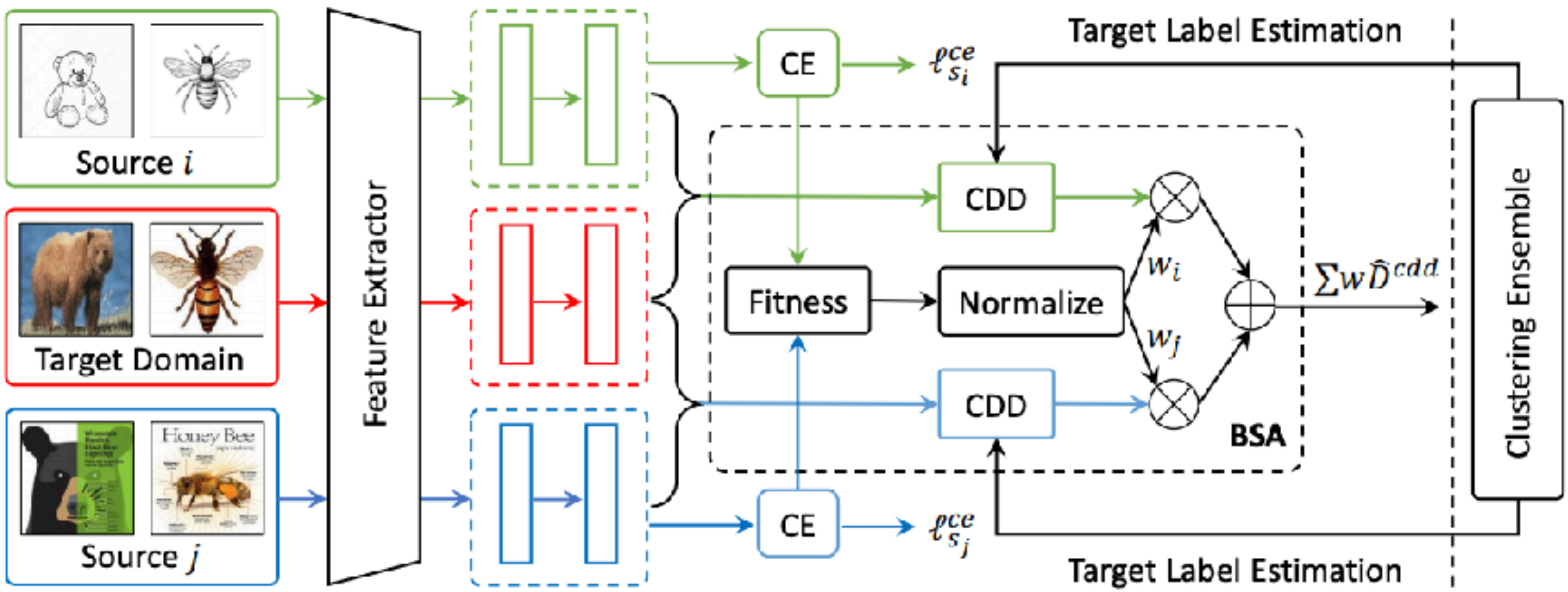
end



Random Sampling   Class-Aware Sampling

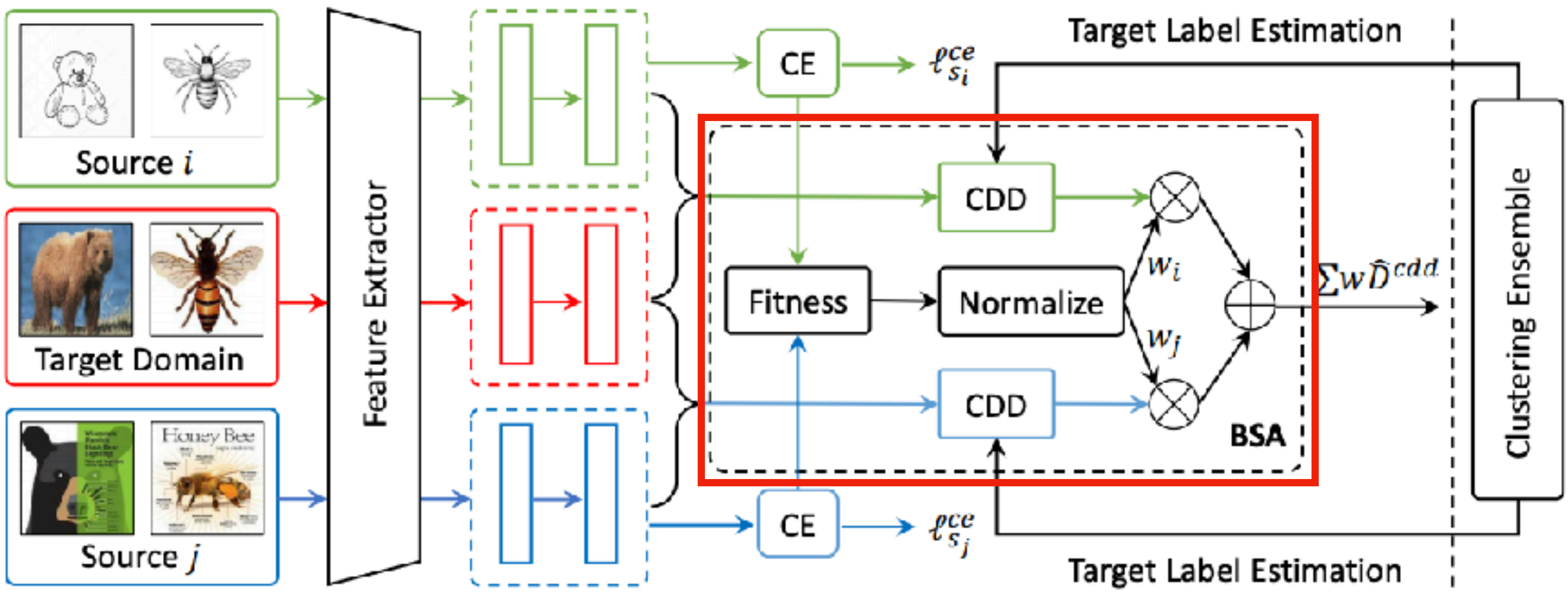
Extension to Multi-Source Setting

Framework



Extension to Multi-Source Setting

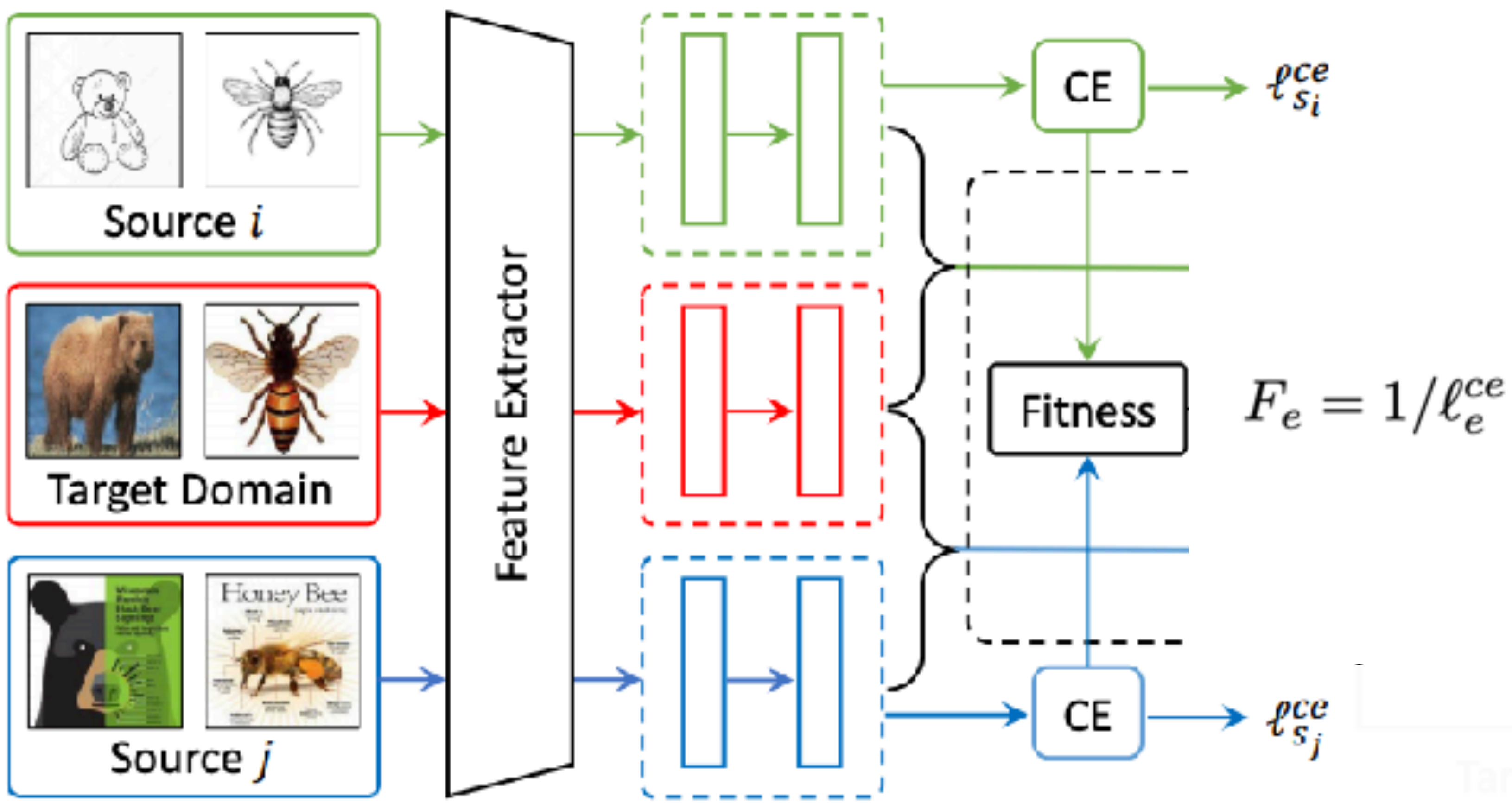
Boundary Sensitive Alignment



$$\min_{\theta} \ell = \sum_{e=1}^{E} \ell^{ce}_{e} + \beta w_e \hat{\mathcal{D}}^{cdd}_{\mathcal{L},e} \quad \textbf{where} \quad F_e = 1/\ell^{ce}_{e} \quad \textbf{and} \quad w_e = \frac{F_e}{\sum_i F_e}$$

Extension to Multi-Source Setting

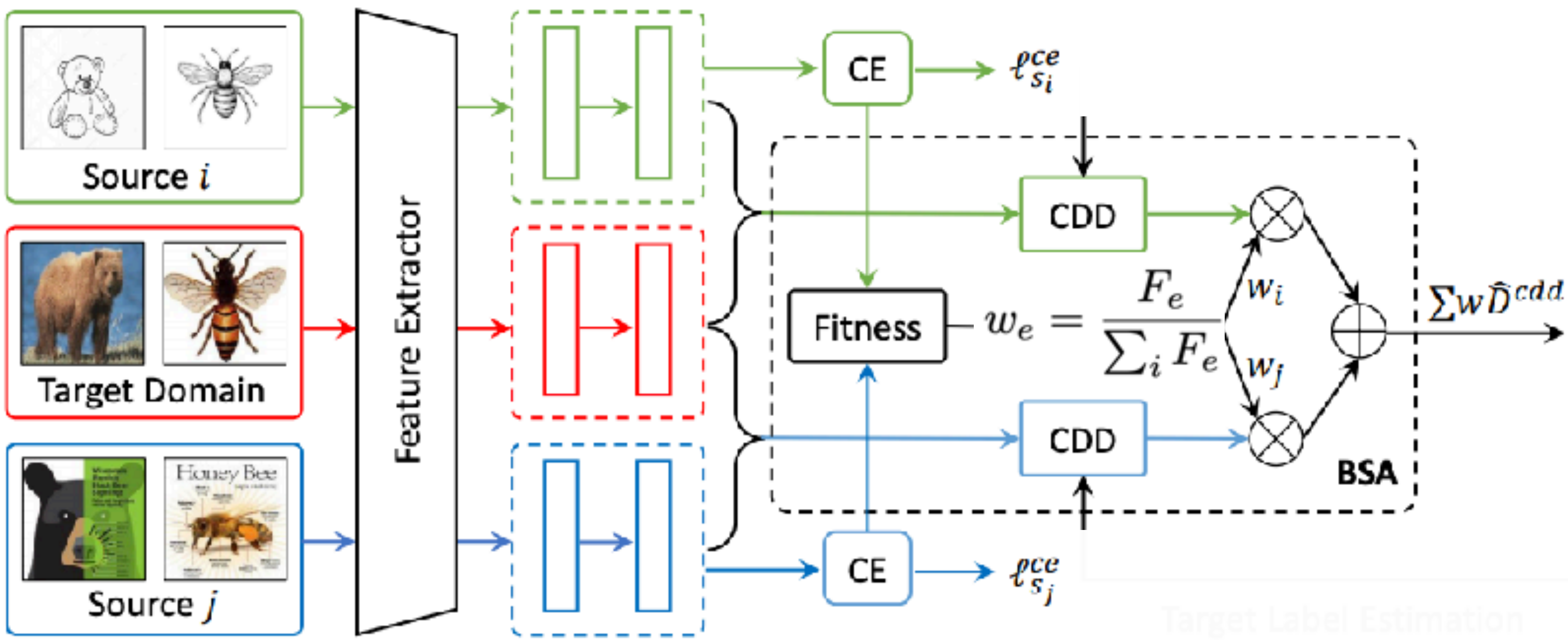
Boundary Sensitive Alignment



$$\min_{\theta} \ell = \sum_{e=1}^{E} \ell_e^{ce} + \beta w_e \hat{\mathcal{D}}_{\mathcal{L},e}^{cdd}$$

Extension to Multi-Source Setting

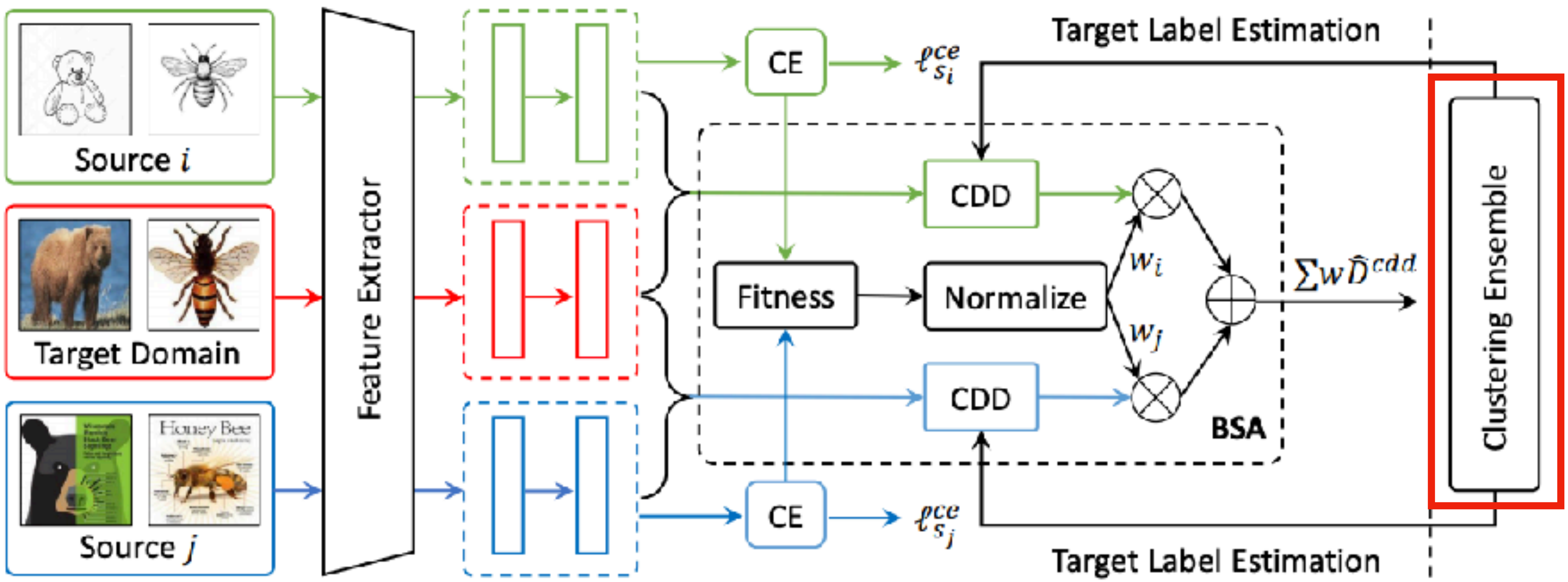
Boundary Sensitive Alignment



$$\min_{\theta} \ell = \sum_{e=1}^{E} \ell_e^{ce} + \beta w_e \hat{\mathcal{D}}_{\mathcal{L},e}^{cdd}$$

Extension to Multi-Source Setting

Clustering Ensemble



Experiment Results

Datasets

Single-Source

Office-31 VisDA-2017

Multi-Source

DomainNet

Experiment Results

Single-Source

Office-31

Method	A → W	D → W	W → D	A → D	D → A	W → A	Average
Source-finetune	68.4 ± 0.2	96.7 ± 0.1	99.3 ± 0.1	68.9 ± 0.2	62.5 ± 0.3	60.7 ± 0.3	76.1
RevGrad [18], [46]	82.0 ± 0.4	96.9 ± 0.2	99.1 ± 0.1	79.7 ± 0.4	68.2 ± 0.4	67.4 ± 0.5	82.2
DAN [13]	80.5 ± 0.4	97.1 ± 0.2	99.6 ± 0.1	78.6 ± 0.2	63.6 ± 0.3	62.8 ± 0.2	80.4
JAN [14]	85.4 ± 0.3	97.4 ± 0.2	99.8 ± 0.2	84.7 ± 0.3	68.6 ± 0.3	70.0 ± 0.4	84.3
MADA [28]	90.0 ± 0.2	97.4 ± 0.1	99.6 ± 0.1	87.8 ± 0.2	70.3 ± 0.3	66.4 ± 0.3	85.2
CDAN [31]	94.1 ± 0.1	98.6 ± 0.1	**100.0 ± 0.0**	92.9 ± 0.2	71.0 ± 0.3	69.3 ± 0.3	87.7
GSDA [33]	**95.7**	**99.1**	**100.0**	94.8	73.5	74.9	89.7
Ours (intra only)	93.2 ± 0.2	98.4 ± 0.2	99.8 ± 0.2	92.9 ± 0.2	76.5 ± 0.3	76.0 ± 0.3	89.5
Ours (CAN)	94.5 ± 0.3	**99.1 ± 0.2**	99.8 ± 0.2	**95.0 ± 0.3**	**78.0 ± 0.3**	**77.0 ± 0.3**	**90.6**

VisDA-2017

Method	airplane	bicycle	bus	car	horse	knife	motorcycle	person	plant	skateboard	train	truck	Average
Source-finetune	72.3	6.1	63.4	91.7	52.7	7.9	80.1	5.6	90.1	18.5	78.1	25.9	49.4
RevGrad [18], [46]	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
DAN [13]	68.1	15.4	76.5	87.0	71.1	48.9	82.3	51.5	88.7	33.2	88.9	42.2	62.8
JAN [14]	75.7	18.7	82.3	86.3	70.2	56.9	80.5	53.8	92.5	32.2	84.5	54.5	65.7
MCD [27]	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
ADR [26]	87.8	79.5	83.7	65.3	92.3	61.8	88.9	73.2	87.8	60.0	85.5	32.3	74.8
SE [47]	95.9	**87.4**	85.2	58.6	96.2	95.7	90.6	80.0	94.8	90.8	88.4	47.9	84.3
DTA [32]	93.7	82.2	**85.6**	83.8	93.0	81.0	90.7	82.0	95.1	78.1	86.4	32.1	81.5
Ours (intra only)	96.5	72.1	80.9	70.8	94.6	**98.0**	**91.7**	**84.2**	90.3	89.8	**89.4**	47.9	83.9
Ours (CAN)	**97.0**	87.2	82.5	74.3	**97.8**	96.2	90.8	80.7	**96.6**	**96.3**	87.5	**59.9**	**87.2**

Experiment Results

Multi-Source

DomainNet

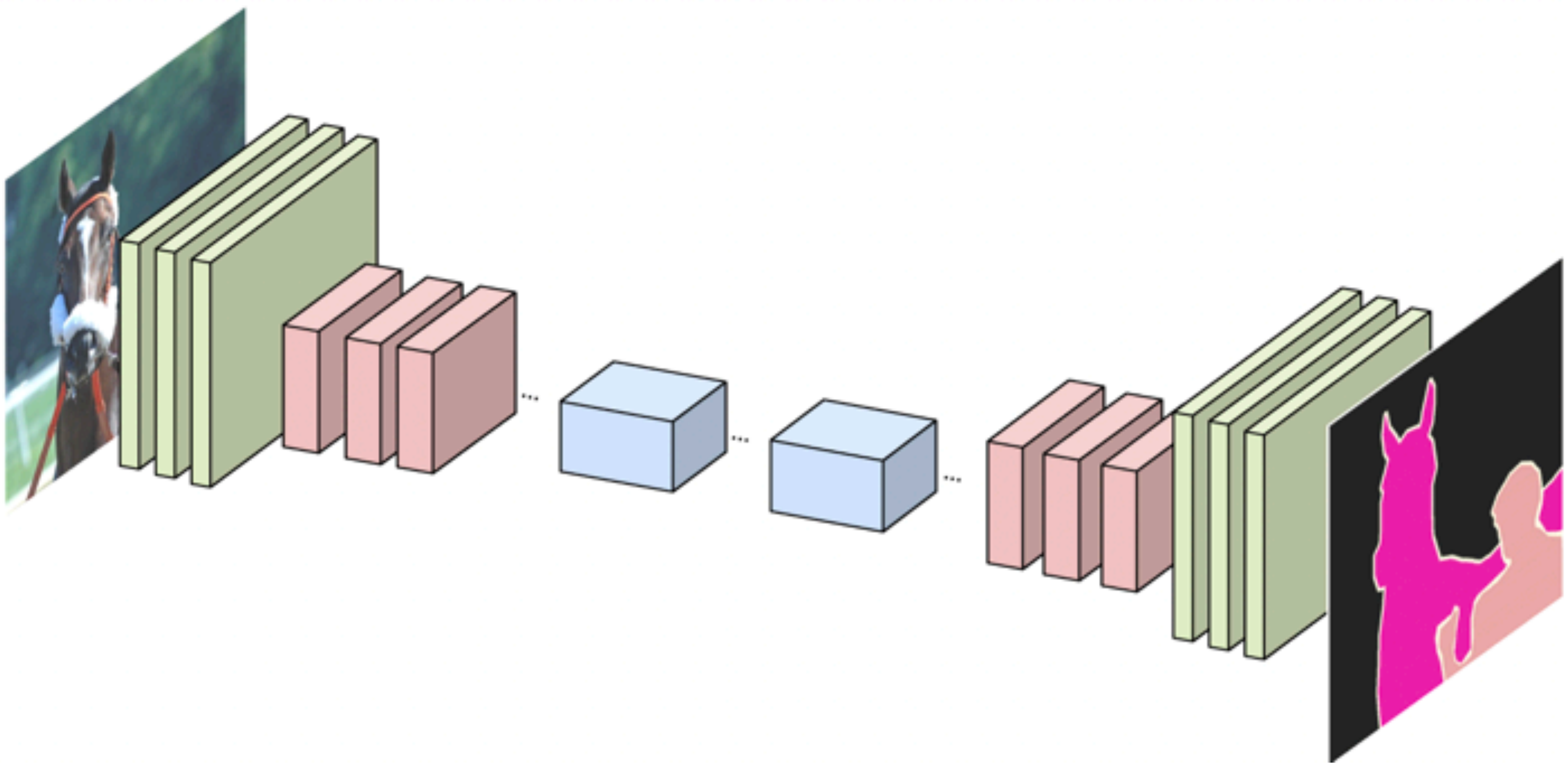
Domain	Method	Target						
		Clipart	Infograph	Painting	Quickdraw	Real	Sketch	Average
Single-Best	Source-finetune	39.6	8.2	33.9	11.8	41.6	23.1	26.4
	RevGrad [18], [46]	37.9	11.4	33.9	13.7	41.5	28.6	27.8
	DAN [13]	39.1	11.4	33.3	16.2	42.1	29.7	28.6
	JAN [14]	35.3	9.1	32.5	14.3	43.1	25.7	26.7
	MCD [27]	42.6	19.6	42.6	3.8	50.5	33.8	32.2
	SE [47]	31.7	12.9	19.9	7.7	33.4	26.3	22.0
	Ours (CAN)	**63.8**	**24.0**	**55.7**	**27.1**	**67.7**	**51.9**	**48.4**
Multi-Source	DCTN [41]	48.6	23.5	48.8	7.2	53.5	47.3	38.2
	M^3SDA [16]	58.6	26.0	52.3	6.3	62.7	49.5	42.6
	Ours (CAN)	67.4	25.3	56.2	26.3	72.5	56.2	50.7
	Ours (MSCAN w/o. BSA)	68.5	27.3	57.4	28.1	72.5	58.1	51.9
	Ours (MSCAN w. BSA)	**69.3**	**28.0**	**58.6**	**30.3**	**73.3**	**59.5**	**53.2**
Oracle	ResNet-101	69.3	34.5	66.3	66.8	80.1	60.7	63.0

Failure Case Analysis

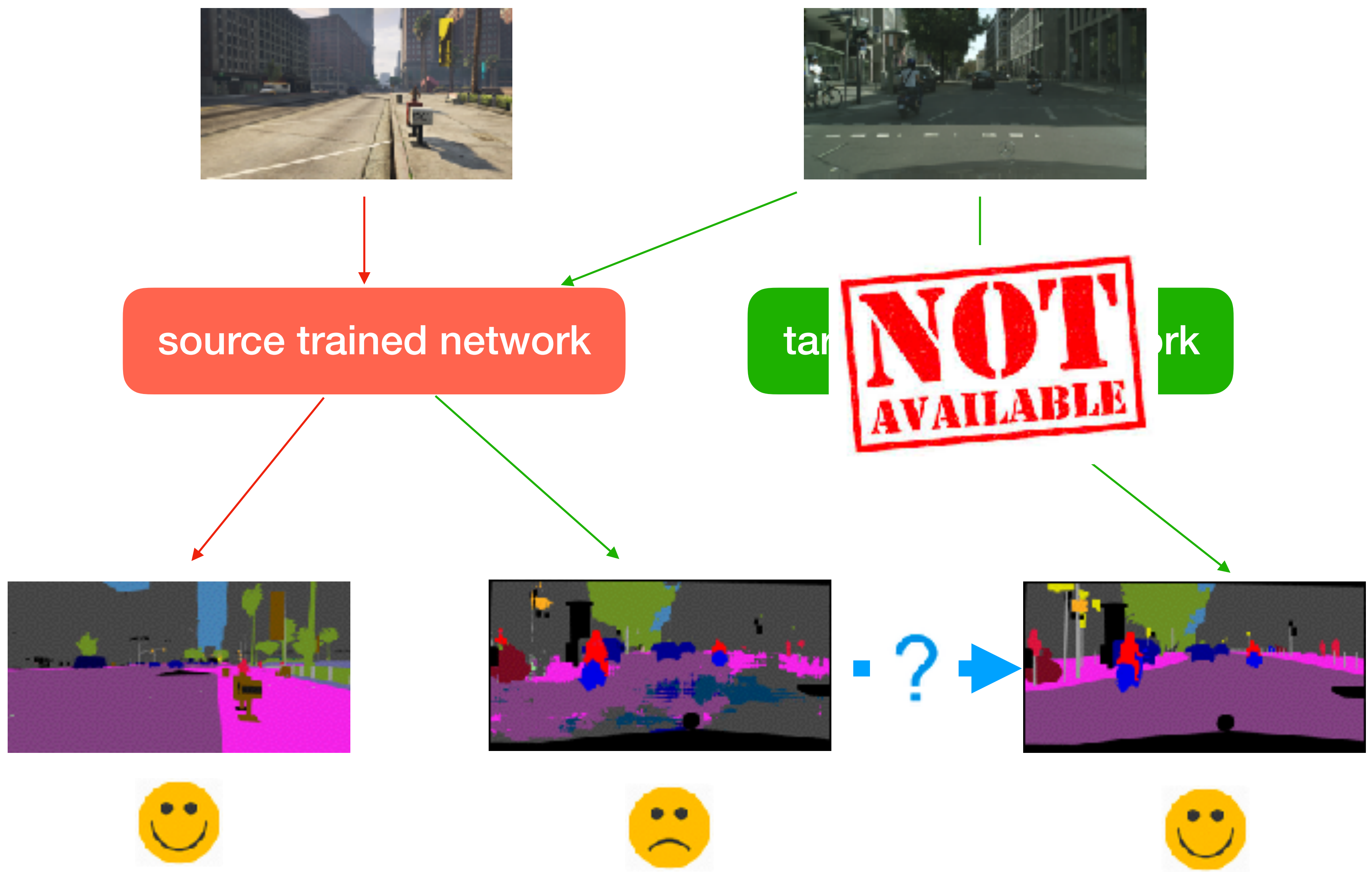
- Introduction
- Contrastive Adaptation Network
- **Pixel-Level Cycle Association**
- Summary

Domain Adaptive Semantic Segmentation

Semantic Segmentation

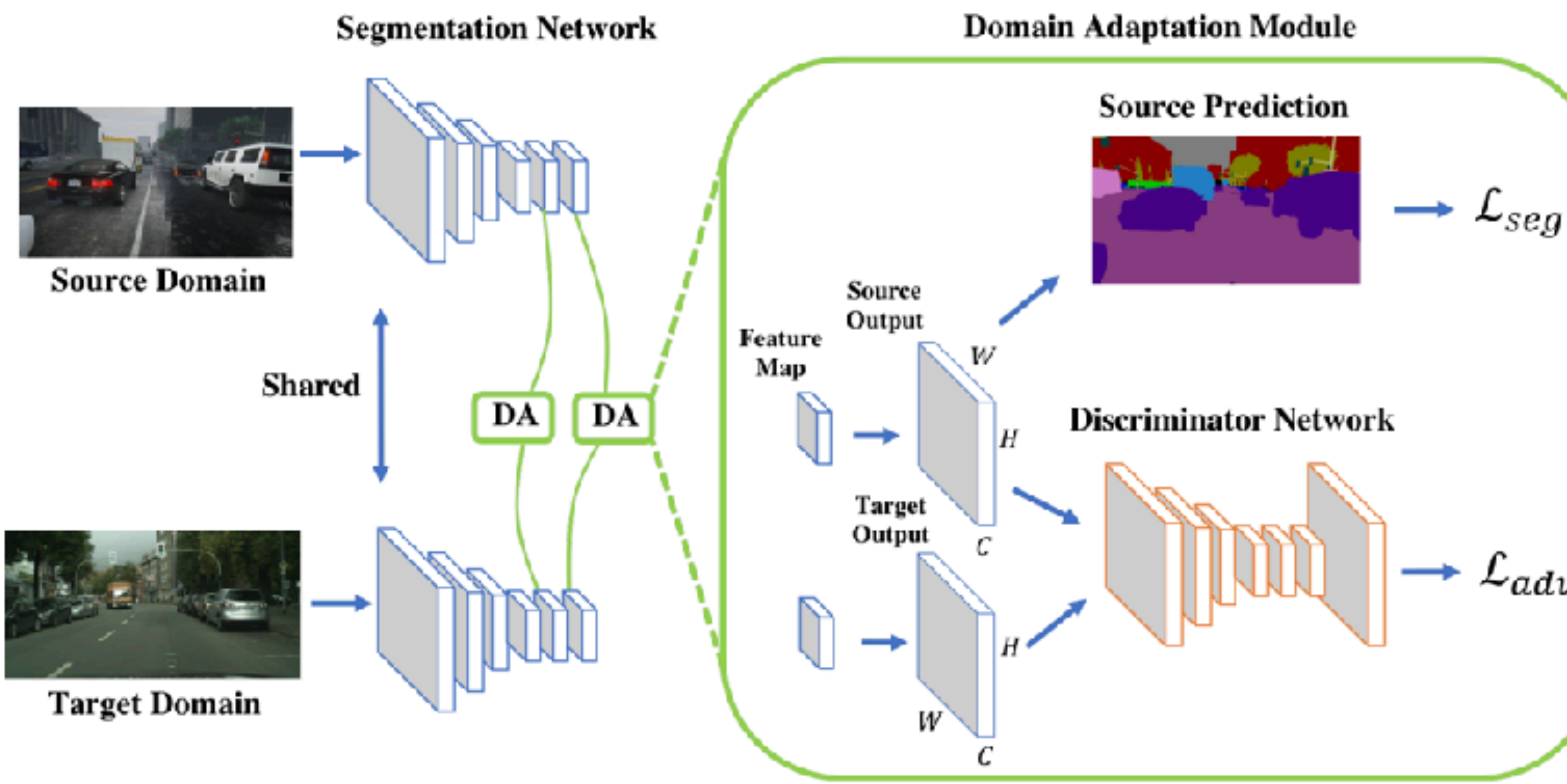


Domain Adaptive Semantic Segmentation



Previous Methods

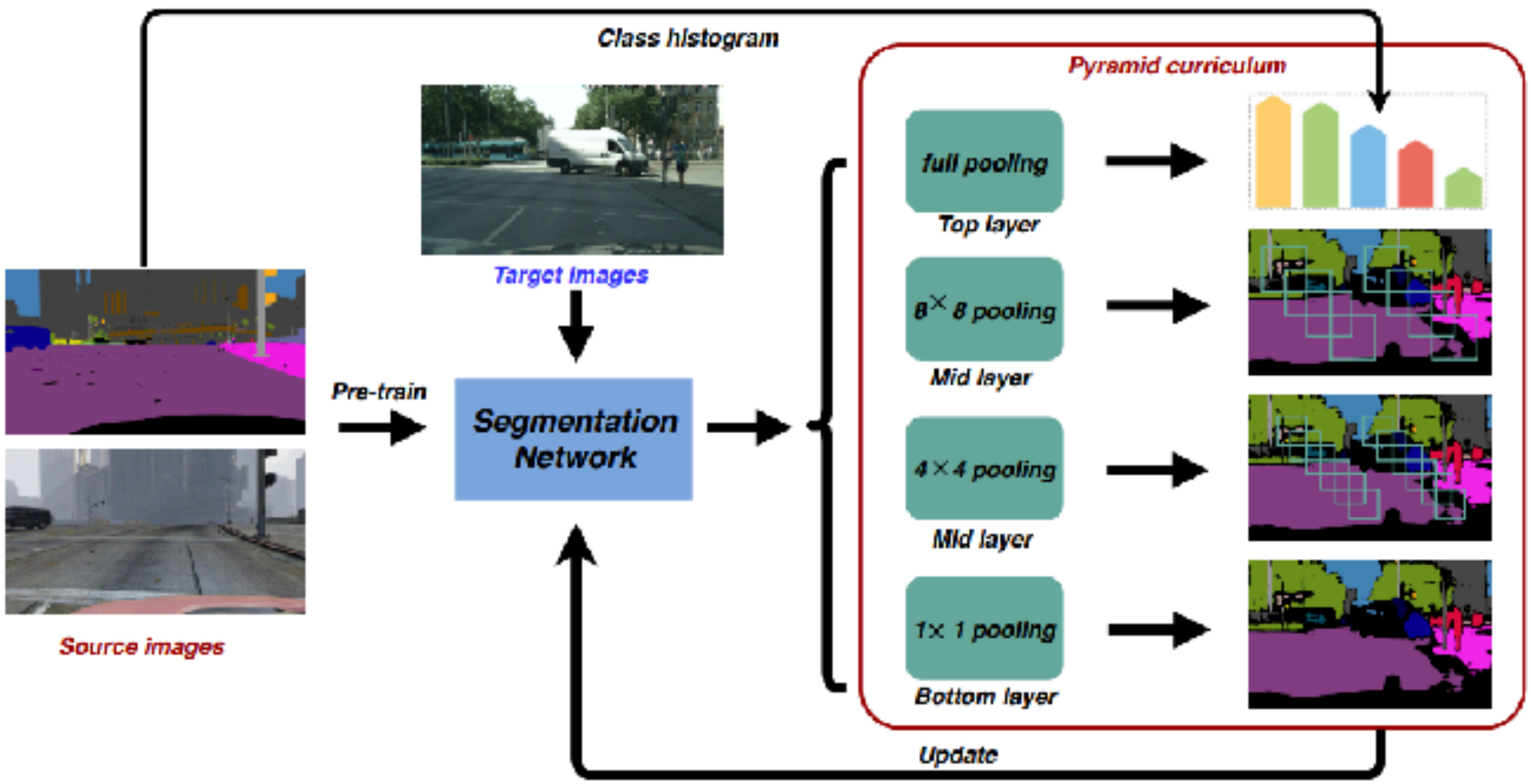
Adversarial training based method



[1] Tsai, Yi-Hsuan, et al. "Learning to adapt structured output space for semantic segmentation." CVPR. 2018.

Previous Methods

Self-training based method



[1] Lian, Qing, et al. "Constructing self-motivated pyramid curriculums for cross-domain semantic segmentation: A non-adversarial approach." ICCV. 2019.

Motivation

Drawbacks of previous methods

- adversarial training based methods:

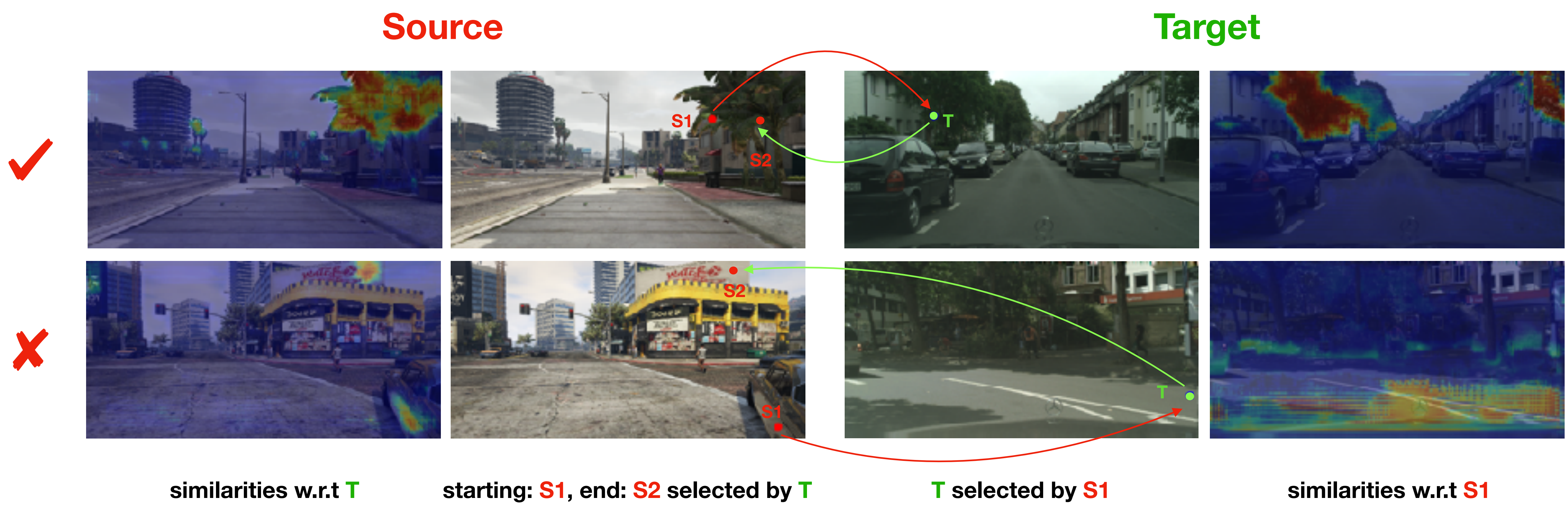
  1) Align globally; 2) Not discriminative enough.

- self-training based methods:

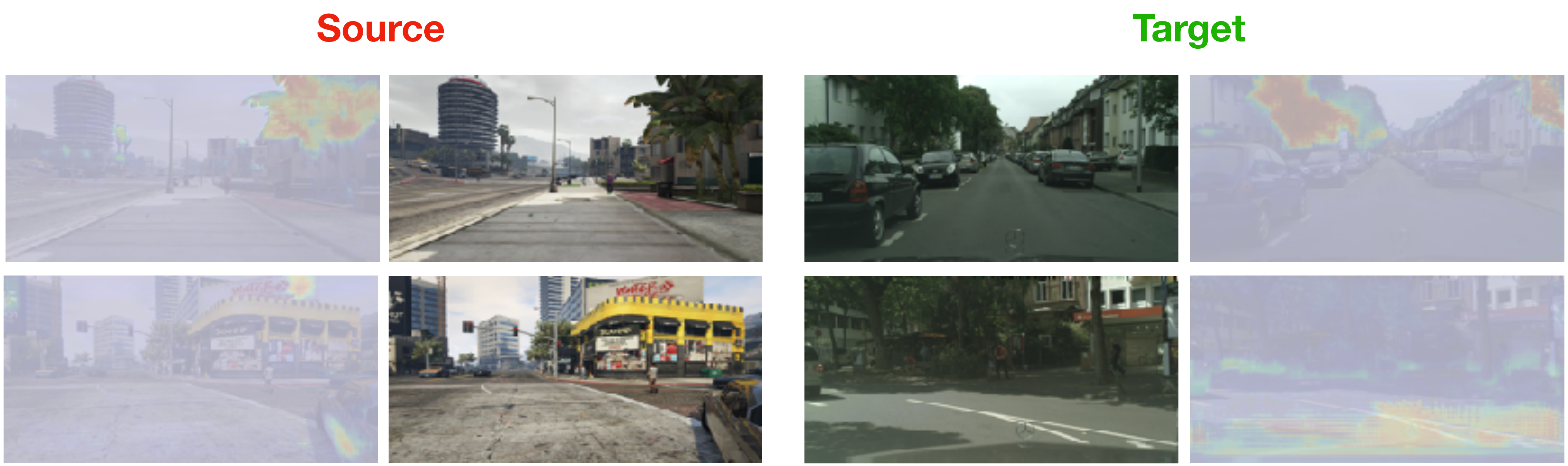
  1) Need good initialization; 2) Sensitive to the noise; 3) Not stable enough.

Build associations between target and source pixels, and diminish pair-wise discrepancy

Pixel-Level Cycle Association



Pixel-Level Cycle Association



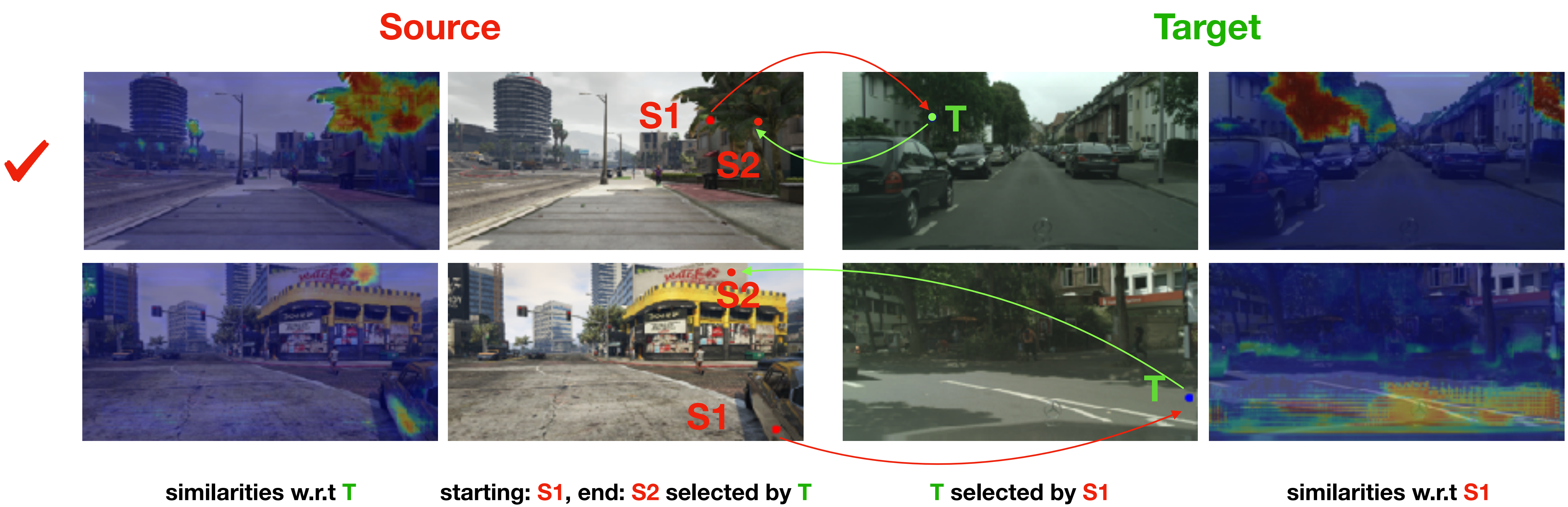
Pixel-Level Cycle Association

Pixel-Level Cycle Association

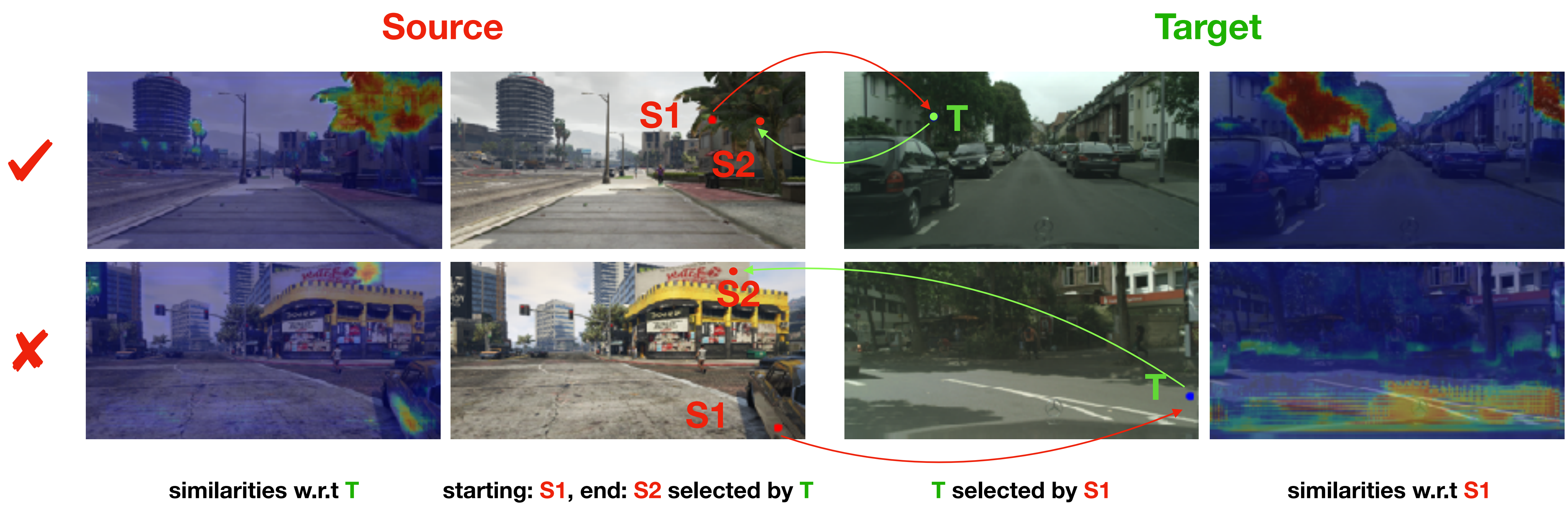
Pixel-Level Cycle Association

Pixel-Level Cycle Association

Pixel-Level Cycle Association



Pixel-Level Cycle Association



Pixel-Level Cycle Association

Similarity between features

$$D(F_i^s, F_j^t) = \langle \frac{F_i^s}{\|F_i^s\|}, \frac{F_j^t}{\|F_j^t\|} \rangle$$

Association loss

$$\cancel{\mathcal{L}^{fass} =} \mathcal{L}^{fass} = -\frac{1}{|\hat{I}^s|} \sum_{i \in \hat{I}^s} \log\{D(F_i^s, F_{j^*}^t) D(F_{j^*}^t, F_{i^*}^s)\} \frac{{}^{t}_{j^*}, F_{i^*}^s)\}}{(F_{j^*}^t, F_{i'}^s)\}} \}$$

Contrast Normalization

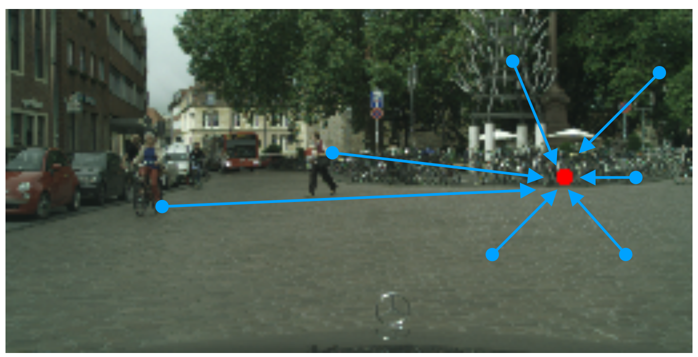
$$D(F_i^s, F_{j'}^t) \leftarrow \frac{D(F_i^s, F_{j'}^t) - \mu_{s \to t}}{\boxed{\sigma_{s \to t}}}, D(F_{j^*}^t, F_{i'}^s) \leftarrow \frac{D(F_{j^*}^t, F_{i'}^s) - \mu_{t \to s}}{\boxed{\sigma_{t \to s}}}$$

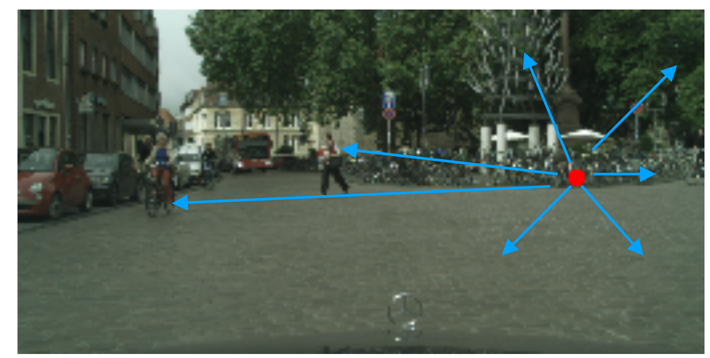
$$\frac{\partial D}{\partial F} \propto \frac{1}{\sigma}$$

Gradient Diffusion via Spatial Aggregation

Spatial Aggregation

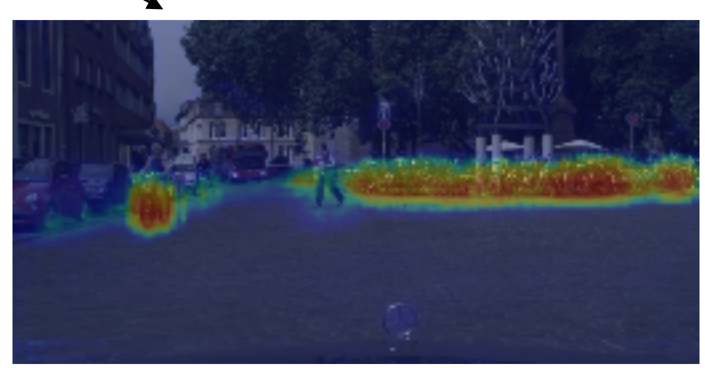
Gradient Diffusion





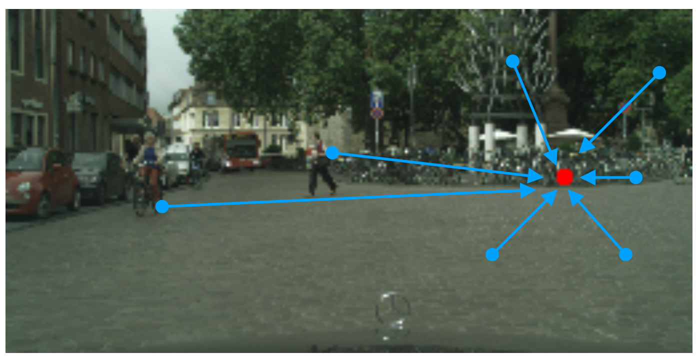
$$\hat{F}_j^t = (1-\alpha)F_j^t + \alpha \sum_{j'} \boxed{w_{j'}} F_{j'}^t$$

$$\frac{\partial \hat{F}_j^t}{\partial \hat{F}_{j' \neq j}^t} = \alpha \times \boxed{w_{j'}}$$



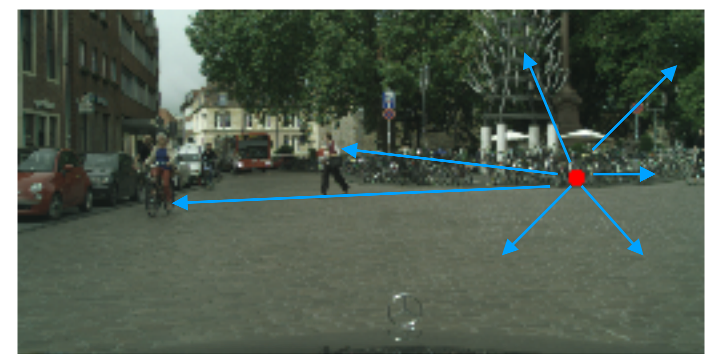
Gradient Diffusion via Spatial Aggregation

Spatial Aggregation

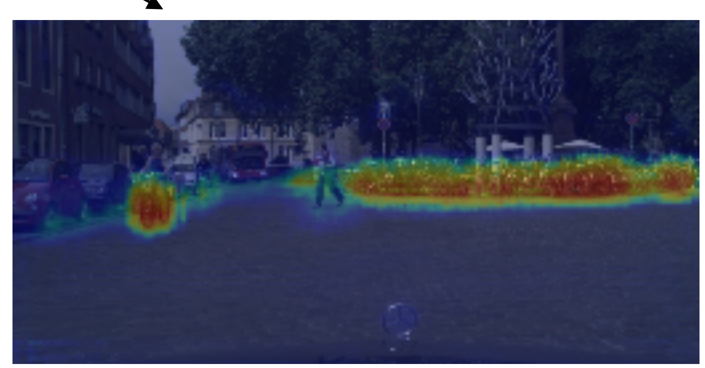


$$\hat{F}_j^t = (1-\alpha)F_j^t + \alpha \sum_{j'} \boxed{w_{j'}} F_{j'}^t$$

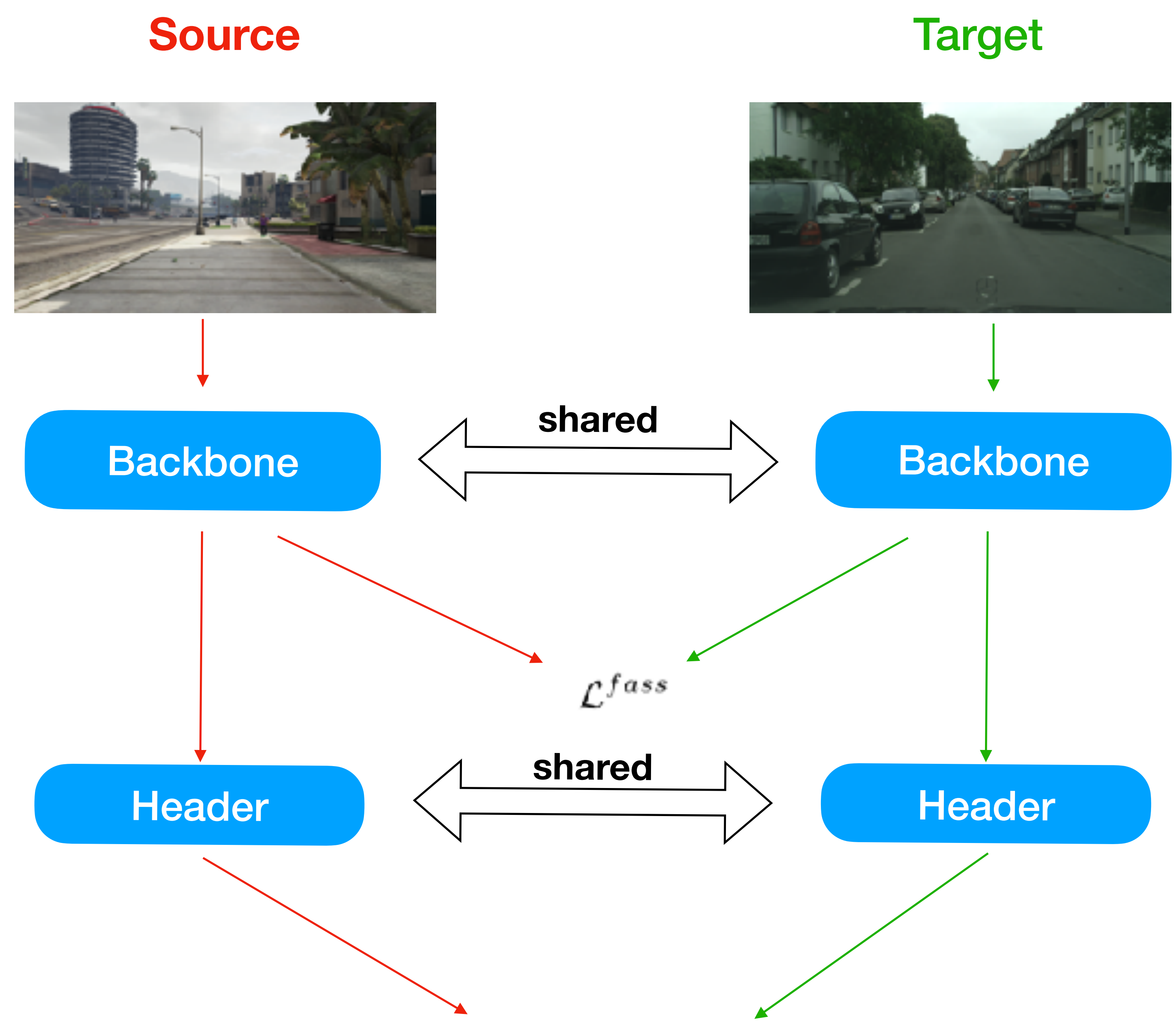
Gradient Diffusion



$$\frac{\partial \hat{F}_j^t}{\partial \hat{F}_{j'\neq j}^t} = \alpha \times \boxed{w_{j'}}$$



Multi-Layer Association



$$D^{kl}(P_i^s, P_j^t) = -\sum_c P_i^s(c) \log \frac{P_i^s(c)}{P_j^t(c)}, \quad D^{kl}(P_j^t, P_i^s) = -\sum_c P_j^t(c) \log \frac{P_j^t(c)}{P_i^s(c)}$$

Objective

$$\mathcal{L}^{full} = \underbrace{\underbrace{\mathcal{L}^{ce} + \beta_1 \mathcal{L}^{lov}}_{\text{source-only}} + \beta_2 \mathcal{L}^{asso} + \beta_3 \mathcal{L}^{lsr}}_{\text{source+target}}$$

Cross-domain association loss

$$\mathcal{L}^{asso} = \mathcal{L}^{fass} + \mathcal{L}^{cass}$$

Adaptive LSR regularizer

$$\mathcal{L}^{lsr} = -\frac{1}{M}\{\frac{1}{|I^s|}\sum_{i \in I^s} \gamma_i \sum_c \log P_i^s(c) + \frac{1}{|I^t|}\sum_{j \in I^t} \gamma_j \sum_c \log P_j^t(c)\}$$

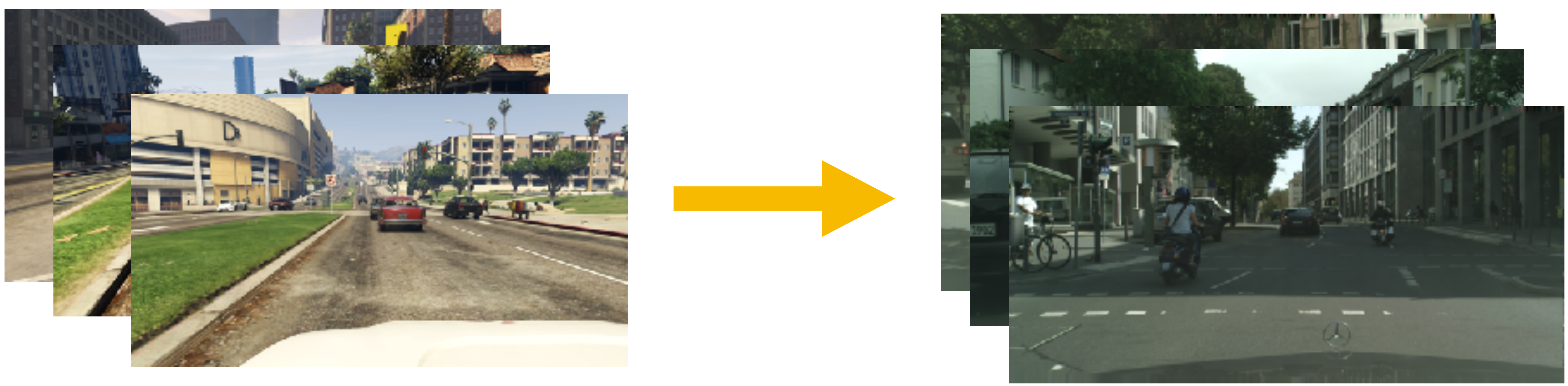
where $\gamma_i = \frac{-\frac{1}{M}\sum_c \log P_i^s(c)}{\lambda} - 1 \quad \gamma_j = \frac{-\frac{1}{M}\sum_c \log P_j^t(c)}{\lambda} - 1$

[1] Maxim Berman, et al. The Lovász-Softmax Loss: A Tractable Surrogate for the Optimization of the Intersection-Over-Union Measure in Neural Networks, CVPR 2018

Experiment Results

Datasets

Source | **Target**



Synthetic Images (SYNTHIA/GTAV)

Real-World Images (Cityscapes)

Experiment Results

Ablation Study

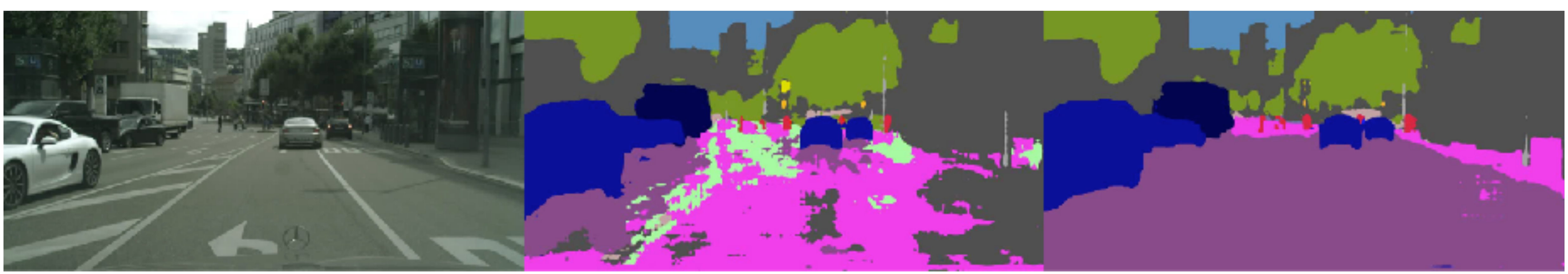
Source Dataset	source-only	source + target			
	$\mathcal{L}^{ce}$				
GTAV	31.5				
SYNTHIA	35.4				

Comparison with previous SOTA

SYNTHIA → Cityscapes																			
	Method	road	side.	build.	wall*	fence*	pole*	light	sign	veg.	sky	person	rider	car	bus	motor	bike	mIoU	mIoU*
Source Only	–	51.2	21.8	67.8	8.2	0.1	26.2	17.7	17.3	69.2	67.1	52.7	22.8	62.3	31.6	21.0	46.1	36.4	42.2
AdaptSeg[41]	AT	84.3	42.7	77.5	–	–	–	4.7	7.0	77.9	82.5	54.3	21.0	72.3	32.2	18.9	32.3	–	46.7
CLAN[31]	AT	81.3	37.0	80.1	–	–	–	16.1	13.7	78.2	81.5	53.4	21.2	73.0	32.9	22.6	30.7	–	47.8
SSF-DAN[15]	AT	84.6	41.7	80.8	–	–	–	11.5	14.7	80.8	**85.3**	57.5	21.6	82.0	36.0	19.3	34.5	–	50.0
ADVENT[44]	AT	85.6	42.2	79.7	8.7	0.4	25.9	5.4	8.1	80.4	84.1	57.9	23.8	73.3	36.4	14.2	33.0	41.2	48.0
DISE [7]	AT	**91.7**	**53.5**	77.1	2.5	0.2	27.1	6.2	7.6	78.4	81.2	55.8	19.2	82.3	30.3	17.1	34.3	41.5	48.7
PatchAlign [42]	AT	82.4	38.0	78.6	8.7	0.6	26.0	3.9	11.1	75.5	84.6	53.5	21.6	71.4	32.6	19.3	31.7	40.0	46.5
MaxSquare[9]	ST	82.9	40.7	80.3	10.2	0.8	25.8	12.8	18.2	82.5	82.2	53.1	18.0	79.0	31.4	10.4	35.6	41.4	48.2
CRST [54]	ST	67.7	32.2	73.9	10.7	**1.6**	**37.4**	22.2	**31.2**	80.8	80.5	60.8	**29.1**	82.8	25.0	19.4	45.3	43.8	50.1
ours	–	82.6	29.0	**81.0**	**11.2**	0.2	33.6	**24.9**	18.3	**82.8**	82.3	**62.1**	26.5	**85.6**	**48.9**	**26.8**	**52.2**	**46.8**	**54.0**

Experiment Results

GTAV → Cityscapes																					
	Method	road	side.	build.	wall	fence	pole	light	sign	veg.	terrain	sky	person	rider	car	truck	bus	train	motor	bike	mIoU
Source Only	–	34.8	14.9	53.4	15.7	21.5	29.7	35.5	18.4	81.9	13.1	70.4	62.0	**34.4**	62.7	21.6	10.7	0.7	**34.9**	35.7	34.3
AdaptSeg[41]	AT	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
ADVENT[44]	AT	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
CLAN[31]	AT	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
DISE[7]	AT	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
SSF-DAN [15]	AT	90.3	38.9	81.7	24.8	22.9	30.5	**37.0**	21.2	84.8	38.8	76.9	58.8	30.7	**85.7**	30.6	38.1	5.9	28.3	36.9	45.4
PatchAlign [42]	AT	**92.3**	51.9	82.1	29.2	25.1	24.5	33.8	**33.0**	82.4	32.8	82.2	58.6	27.2	84.3	33.4	46.3	2.2	29.5	32.3	46.5
MaxSquare[9]	ST	89.4	43.0	82.1	30.5	21.3	30.3	34.7	24.0	85.3	**39.4**	78.2	63.0	22.9	84.6	36.4	43.0	5.5	34.7	33.5	46.4
CRST[54]	ST	91.0	**55.4**	80.0	33.7	21.4	**37.3**	32.9	24.5	85.0	34.1	80.8	57.7	24.6	84.1	27.8	30.1	**26.9**	26.0	**42.3**	47.1
ours	–	84.0	30.4	82.4	**35.3**	24.8	32.2	36.8	24.5	**85.5**	37.2	78.6	**66.9**	32.8	85.5	**40.4**	**48.0**	8.8	29.8	41.8	**47.7**



Cityscapes **Source-only** **Our PLCA**

- Introduction
- Contrastive Adaptation Network
- Pixel-Level Cycle Association
- **Summary**

Summary

- Without considering the discriminative ability of features, the adapted features would be sub-optimal for the downstream task.

- Class-aware alignment helps avoid the misalignment and improve the generalization ability of features.

- In the semantic segmentation task, taking the pixel-wise discrepancy into consideration is beneficial.

- In future, how to automatically optimize the discrepancy/alignment metric is worth investigating.

Thank you for listening !