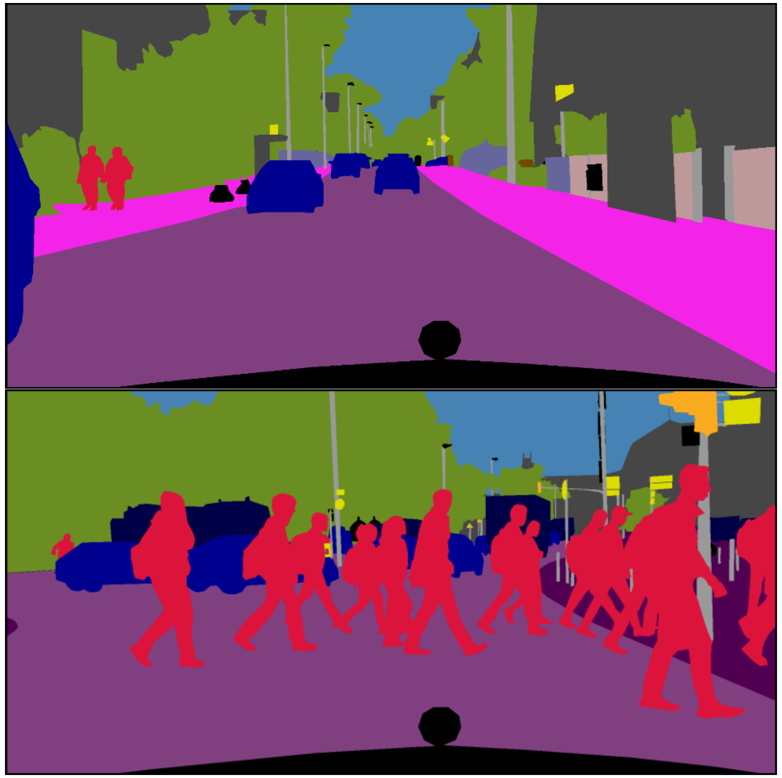
Prototype-based

Domain Adaptation

01 Background

Annotations are particularly costly as every pixel has to be labeled

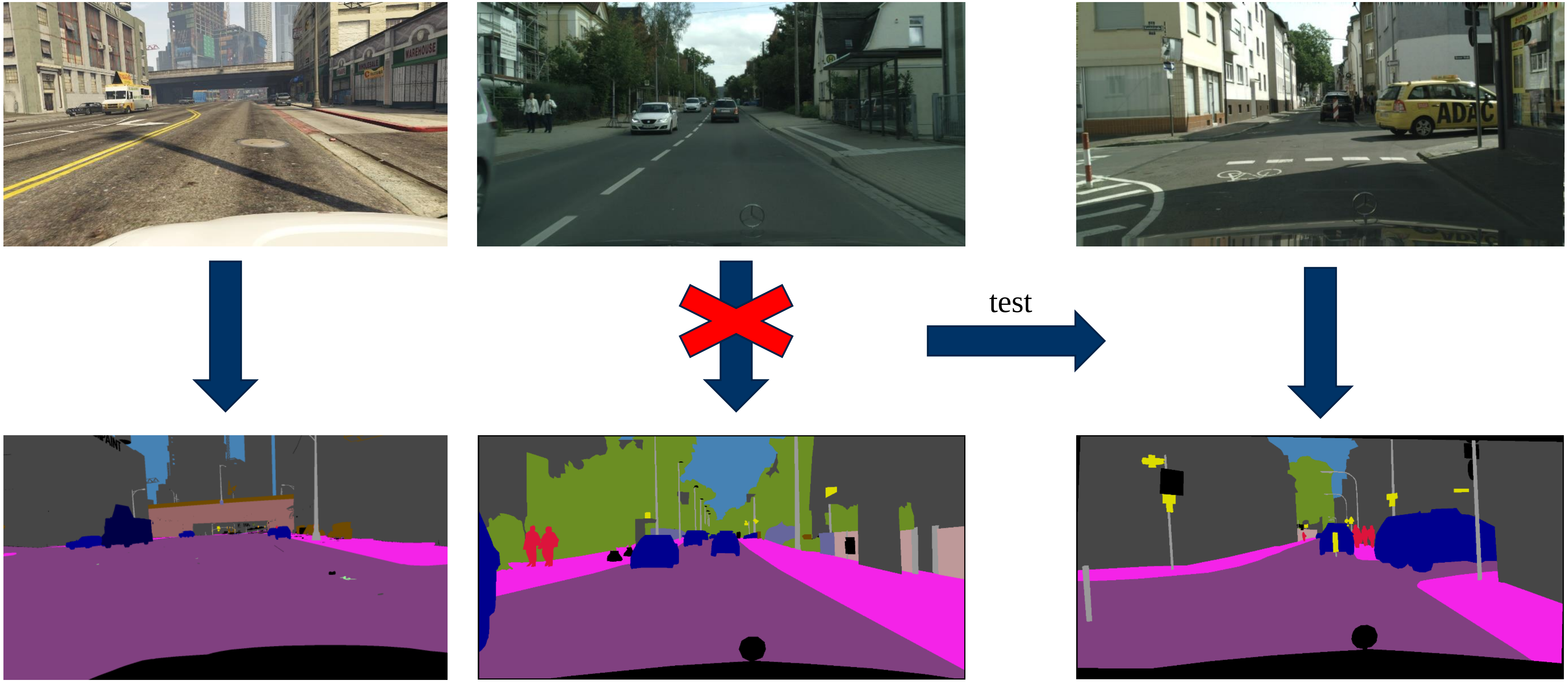


01 What is UDA?

UDA: Unsupervised Domain Adaptation

Image:

Label:



Domain Adaptation

Bi-directional Contrastive Learning for Domain Adaptive Semantic Segmentation

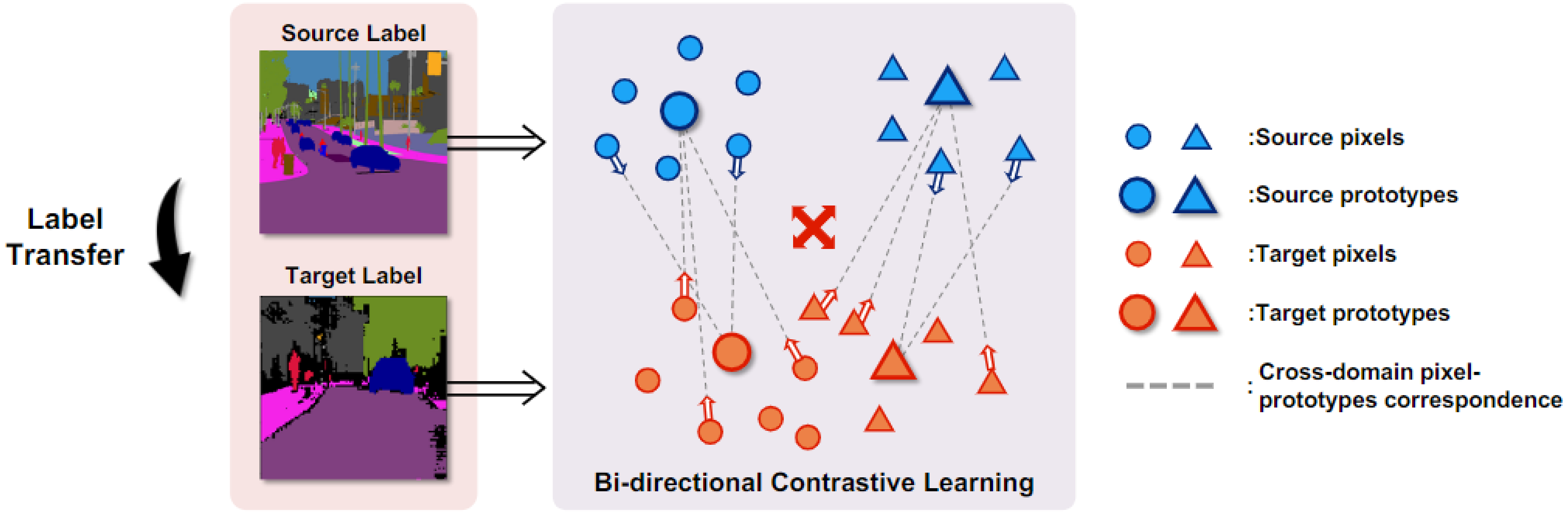
Geon Lee, Chanho Eom, Wonkyung Lee, Hyekang Park, and Bumsub Ham*
https://cvlab.yonsei.ac.kr/projects/DASS

Yonsei University

01 Bi-directional Contrastive Learning for Domain Adaptive Semantic Segmentation

- **ECCV 2022**
- **Motivation**

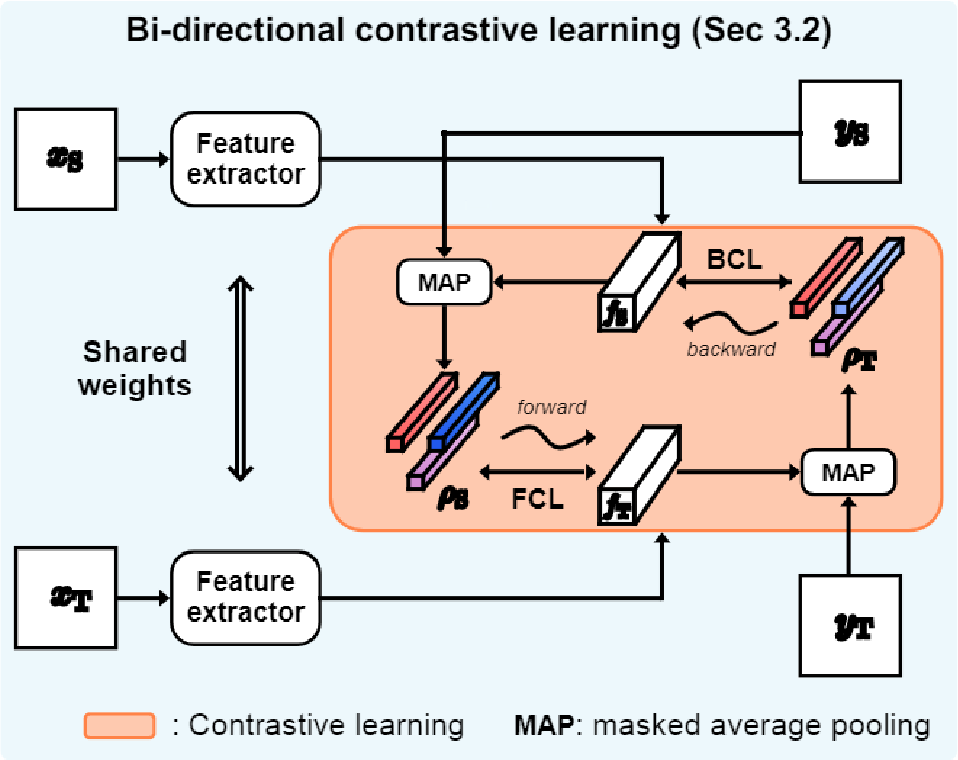
However, they typically focus on reducing the domain discrepancy globally, and fail to keep pixel-level semantics



01 Bi-directional Contrastive Learning for Domain Adaptive Semantic Segmentation

- **Contribution**
  - We introduce a novel contrastive learning framework using bi-directional pixel-prototype correspondences to learn domain-invariant and discriminative feature representations for UDASS
  - We propose a nonparametric approach to generating dynamic pseudo labels. We also present a calibration method to reduce domain biases for pixel-prototype correspondences between target and source domains
  - We set a new state of the art on standard benchmarks for UDASS, and demonstrate the effectiveness of our contrast learning framework

01 Bi-directional Contrastive Learning for Domain Adaptive Semantic Segmentation

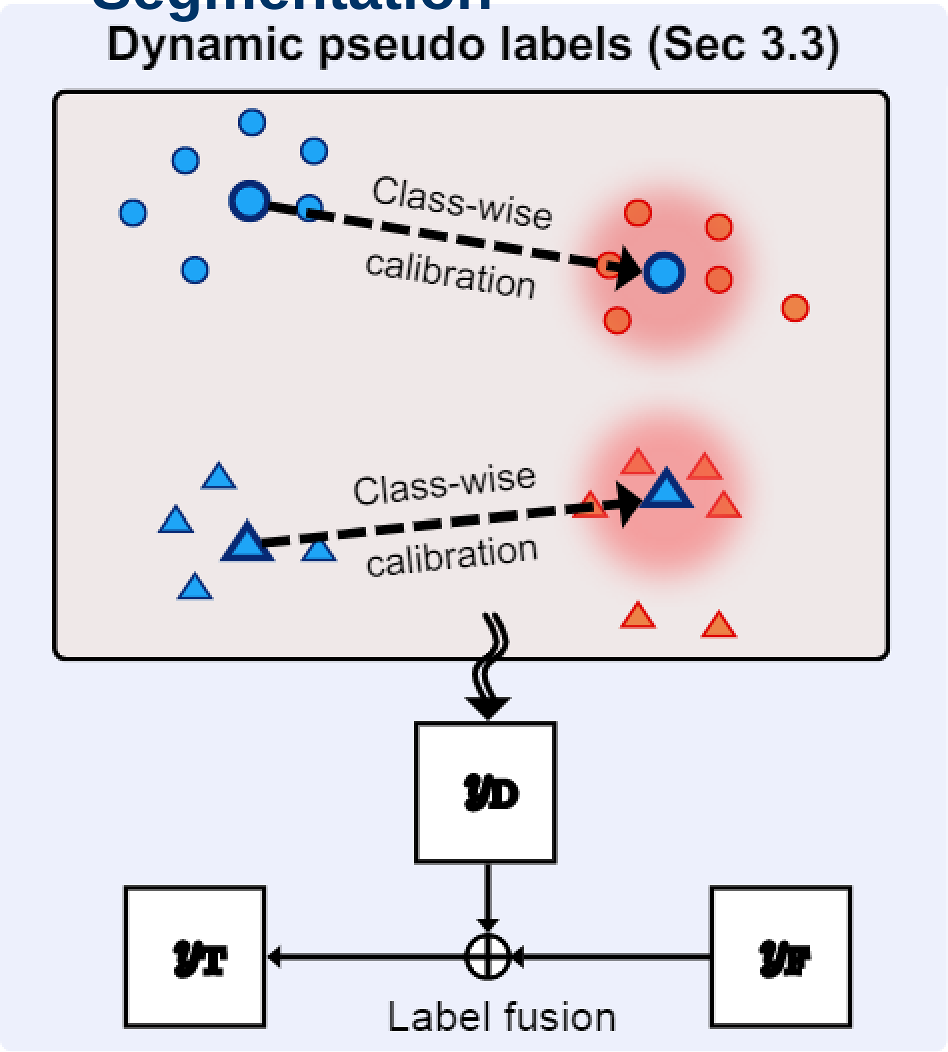


$$\rho_S(c) = \frac{\sum_p f_S(p) y_S(p,c)}{\sum_p y_S(p,c)}, \rho_T(c) = \frac{\sum_p f_T(p) y_T(p,c)}{\sum_p y_T(p,c)},$$

$$\mathcal{L}_{FC} = -\sum_c \sum_p y_T(p,c) \log \frac{\exp\left(s(f_T(p), \rho_S(c))/\tau\right)}{\sum_c \exp\left(s(f_T(p), \rho_S(c))/\tau\right)},$$

$$\mathcal{L}_{BC} = -\sum_c \sum_p y_S(p,c) \log \frac{\exp\left(s(f_S(p), \rho_T(c))/\tau\right)}{\sum_c \exp\left(s(f_S(p), \rho_T(c))/\tau\right)}.$$

01 Bi-directional Contrastive Learning for Domain Adaptive Semantic Segmentation



$$\mu_S(c) \leftarrow \lambda \mu_S(c) + (1-\lambda)\rho_S(c)$$
$$\mu_T(c) \leftarrow \lambda \mu_T(c) + (1-\lambda)\rho_T(c)$$

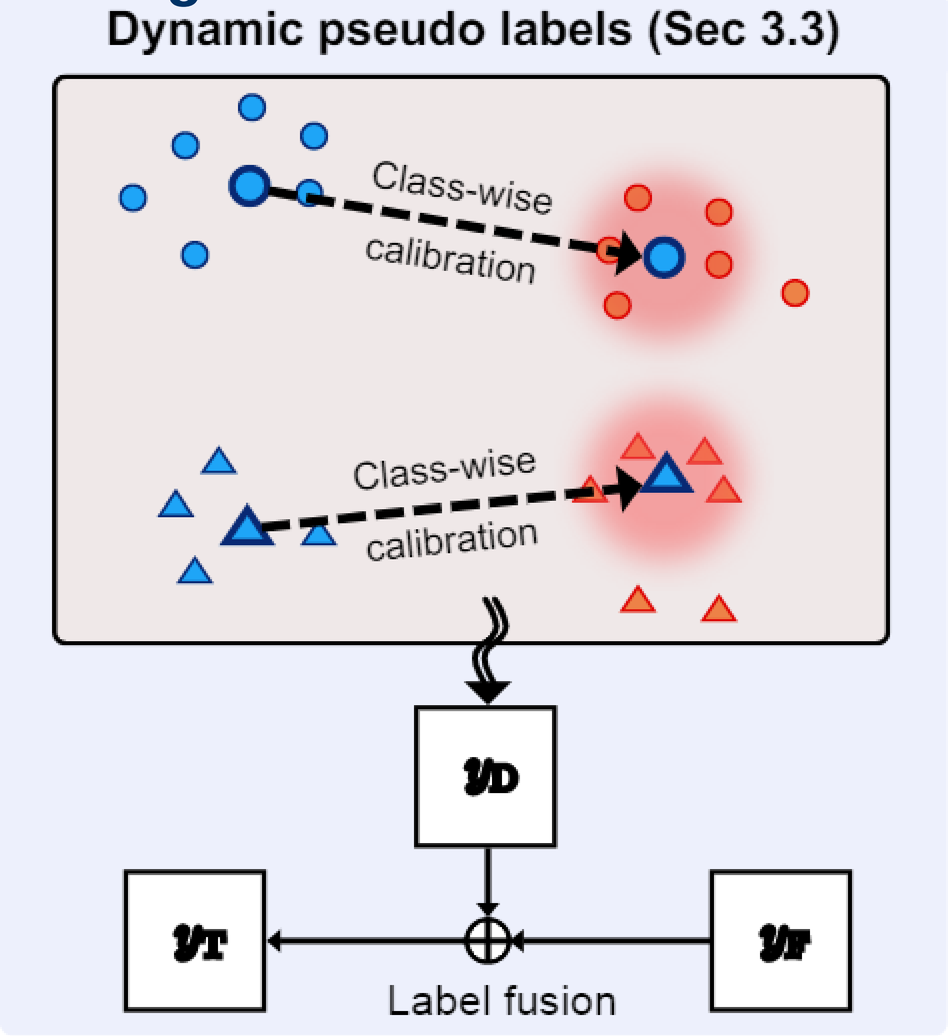
$$\xi(c) = \mu_T(c) - \mu_S(c)$$

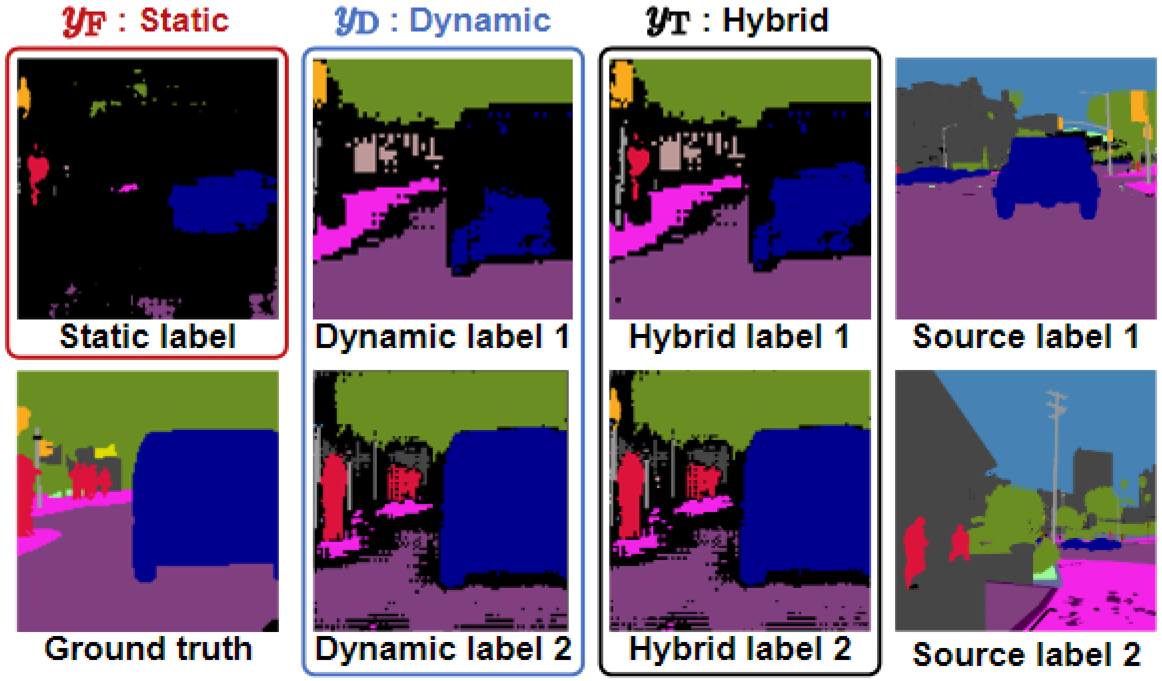
$$\rho_{S\to T}(c) = \rho_S(c) + \xi(c).$$

$$y_D(p,c) = \begin{cases} 1, & \text{if } s(f_T(p)), \rho_{S\to T}(c)) > \mathcal{T} \text{and } c = c' \\ 0, & \text{otherwise} \end{cases}$$

$$y_T(p,c) = \begin{cases} y_D(p,c), & \text{if } y_D(p,c) = 1 \\ y_F(p,c), & \text{if } y_D(p,c') = 0 \text{ for } c' \in \mathcal{C}, \text{ and } y_F(p,c) = 1 \\ 0, & \text{otherwise} \end{cases}$$

01 Bi-directional Contrastive Learning for Domain Adaptive Semantic Segmentation





01 Bi-directional Contrastive Learning for Domain Adaptive Semantic Segmentation

- **Experimental Results**
- **GTA5:**

GTA5 → Cityscapes																						
Split	Methods	Type	road	side.	build.	wall	fence	pole	light	sign	veg.	terrian	sky	person	rider	car	truck	bus	train	motor	bike	mIoU
Validation	Source-only	-	45.4	16.5	66.4	14.4	21.6	25.1	36.3	17.2	80.1	16.3	69.1	61.4	24.9	68.6	28.4	4.7	4.4	40.8	27.5	35.2
	AdaptSeg [43]	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
	CBST [58]	ST	91.8	53.5	80.5	32.7	21.0	34.0	28.9	20.4	83.9	34.2	80.9	53.1	24.0	82.7	30.3	35.9	16.0	25.9	42.8	45.9
	CRST [59]	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
	PLCA [20]	-	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
	CAG_UDA [53]	ST	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
	FDA [51]	ST	92.5	53.5	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
	TPLD [37]	ST	94.2	60.5	82.8	36.6	16.6	39.3	29.0	25.5	85.6	44.9	84.4	60.6	27.4	84.1	37.0	47.0	31.2	36.1	46.4	51.2
	CorDA [47]	ST	**94.7**	**63.1**	87.6	30.7	40.6	40.2	47.8	51.6	87.6	**47.0**	89.7	66.7	35.9	90.2	**48.9**	57.5	0.0	39.8	56.0	56.6
	ProDA [52]	ST	87.1	55.1	78.1	**45.6**	**43.8**	**44.6**	52.5	**53.4**	89.1	44.7	82.1	70.1	**39.1**	88.4	43.8	**59.1**	1.0	**48.7**	54.4	56.5
	Ours	ST	93.5	60.2	**88.1**	31.1	37.0	41.9	**54.7**	37.8	**89.9**	45.5	**89.9**	**72.7**	38.2	**90.7**	34.3	53.2	4.4	47.2	**58.5**	**57.1**
Test	AdaptSeg [43]	AT	88.5	40.4	81.0	26.3	20.6	25.6	36.0	12.9	84.8	45.5	87.2	63.7	35.8	76.4	27.7	28.0	2.9	33.0	26.1	44.3
	CBST [58]	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
	CRST [59]	ST	93.5	57.6	84.6	**39.3**	24.1	25.2	35.0	17.3	85.0	40.6	86.5	58.7	28.7	85.8	**49.0**	56.4	5.4	31.9	43.2	49.9
	FDA-MBT [51]	ST	93.4	55.8	83.6	25.4	23.1	33.2	39.0	36.9	84.0	47.2	88.8	66.3	40.6	87.4	26.9	49.6	12.8	35.2	42.8	51.2
	CorDA [47]	ST	**94.2**	**62.9**	88.1	30.2	41.2	40.1	49.1	49.9	89.1	49.1	90.1	69.1	28.9	86.2	46.2	59.5	1.2	35.2	52.3	57.5
	ProDA [52]	ST	88.1	57.1	81.2	46.1	**45.2**	**41.5**	**55.1**	**56.2**	86.1	45.1	78.1	73.2	40.1	88.8	48.7	**60.1**	1.1	50.3	53.1	57.6
	Ours	ST	93.8	59.7	**90.1**	38.0	33.4	39.9	45.3	30.5	**92.2**	**58.2**	**94.8**	**81.9**	**47.9**	**93.2**	40.1	53.1	13.1	**51.2**	**58.2**	**58.5**

01 Bi-directional Contrastive Learning for Domain Adaptive Semantic Segmentation

- **Experimental Results**
- **SYNTHIA :**

SYNTHIA → Cityscapes																			
Methods	Type	road	side.	build.	wall*	fence*	pole*	light	sign	veg.	sky	person	rider	car	bus	motor	bike	mIoU	mIoU*
Source-only	AT	53.4	23.4	73.0	5.5	0.0	25.7	6.6	7.0	77.9	55.3	52.9	21.0	60.9	6.6	21.8	33.7	32.5	37.6
AdaptSeg [43]	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
CBST [58]	ST	68.0	29.9	76.3	10.8	1.4	33.9	22.8	29.5	77.6	78.3	60.6	28.3	81.6	23.5	18.8	39.8	38.9	42.6
CRST [59]	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
CAG_UDA [53]	ST	84.7	40.8	81.7	7.8	0.0	35.1	13.3	22.7	84.5	77.6	64.2	27.8	80.9	19.7	22.7	48.3	44.5	51.5
FDA [51]	ST	79.3	35.0	73.2	-	-	-	19.9	24.0	61.7	82.6	61.4	31.1	83.9	40.8	38.4	51.1	-	52.5
PLCA [20]	-	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
TPLD [37]	ST	80.9	44.3	82.2	19.9	0.3	40.6	20.5	30.1	77.2	80.9	60.6	25.5	84.8	41.1	24.7	43.7	47.3	53.5
CorDA [47]	ST	**93.3**	**61.6**	85.3	19.6	5.1	37.8	36.6	**42.8**	84.9	90.4	69.7	41.8	85.6	38.4	32.6	53.9	55.0	62.8
ProDA [52]	ST	87.3	45.1	84.2	**36.5**	0.0	**43.3**	**54.7**	36.0	88.3	83.1	71.5	24.4	88.4	**50.1**	40.1	45.6	55.1	61.3
Ours	ST	83.8	42.2	**85.3**	16.4	**5.7**	43.1	48.3	30.2	**89.3**	**92.1**	**68.2**	**43.1**	**89.7**	47.2	**42.2**	**54.2**	**55.6**	**62.9**

01 Bi-directional Contrastive Learning for Domain Adaptive Semantic Segmentation

- **Ablation study**

$\mathcal{L}_{\text{base}}$	$\mathcal{L}_{FC}$	$\mathcal{L}_{BC}$	$+y_{\text{D}}$ (w/o cal.)	$+y_{\text{D}}$ (w/ cal.)	Source dataset	
					GTA5	SYNTHIA
✓					49.5	45.1
✓	✓				51.2	48.8
✓	✓	✓			53.5	51.3
✓	✓	✓	✓		55.3	53.5
✓	✓	✓		✓	57.1	55.6

(a) w/o cal. (b) w/ cal. (c) GT labels.

Fig. 6: Visualization of dynamic pseudo labels. (a-b) Pseudo labels obtained without and with calibrating prototypes of a source domain; (c) Target labels.

Table 4: Quantitative results for various pseudo labels of a target domain. We report the densities of static, dynamic, and hybrid pseudo labels and corresponding label accuracies.

Pseudo labels	Density(%)	Accuracy(%)
Static [58]	20.1	98.5
Dyn. (w/o cal.)	22.2	98.6
Dyn. (w/ cal.)	34.3	98.6
Hybrid	42.3	98.8