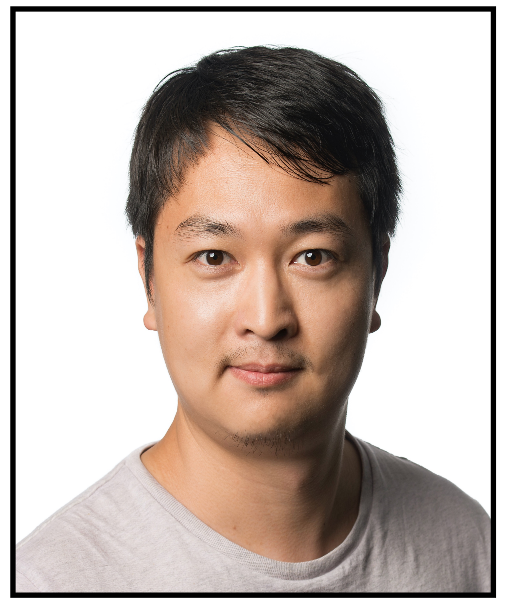
max planck institut informatik
JUNE 18-22, 2023 CVPR VANCOUVER, CANADA
SIC Saarland Informatics Campus

Weakly-Supervised Domain Adaptive Semantic Segmentation with Prototypical Contrastive Learning


Anurag Das


Yongqin Xian*

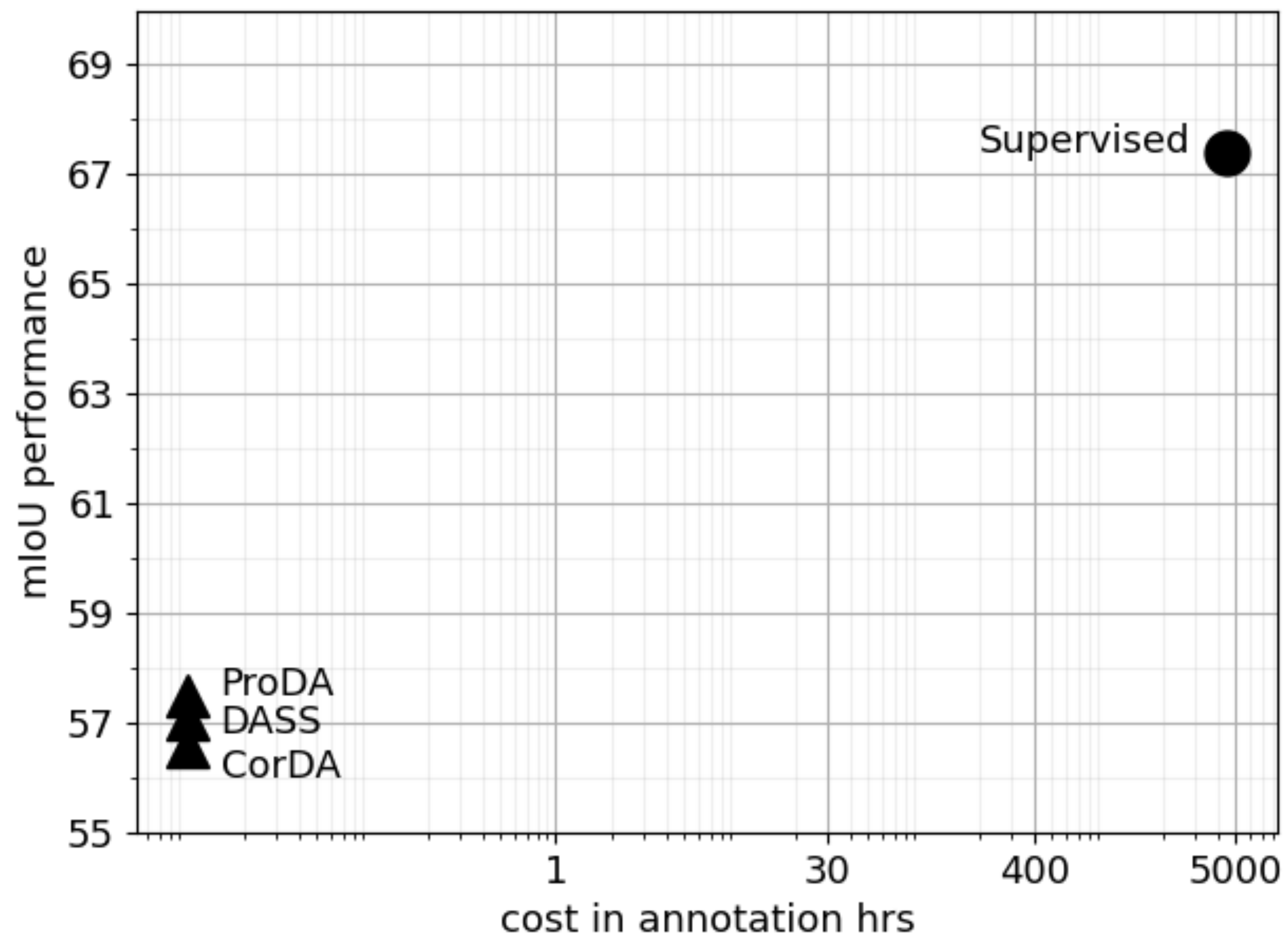

Dengxin Dai


Bernt Schiele

Max Planck Institute for Informatics, Saarland Informatics Campus
*ETH Zürich

Motivation

- **Huge performance gap** between Unsupervised Domain Adaptive Semantic Segmentation (UDASS) and Supervised Learning



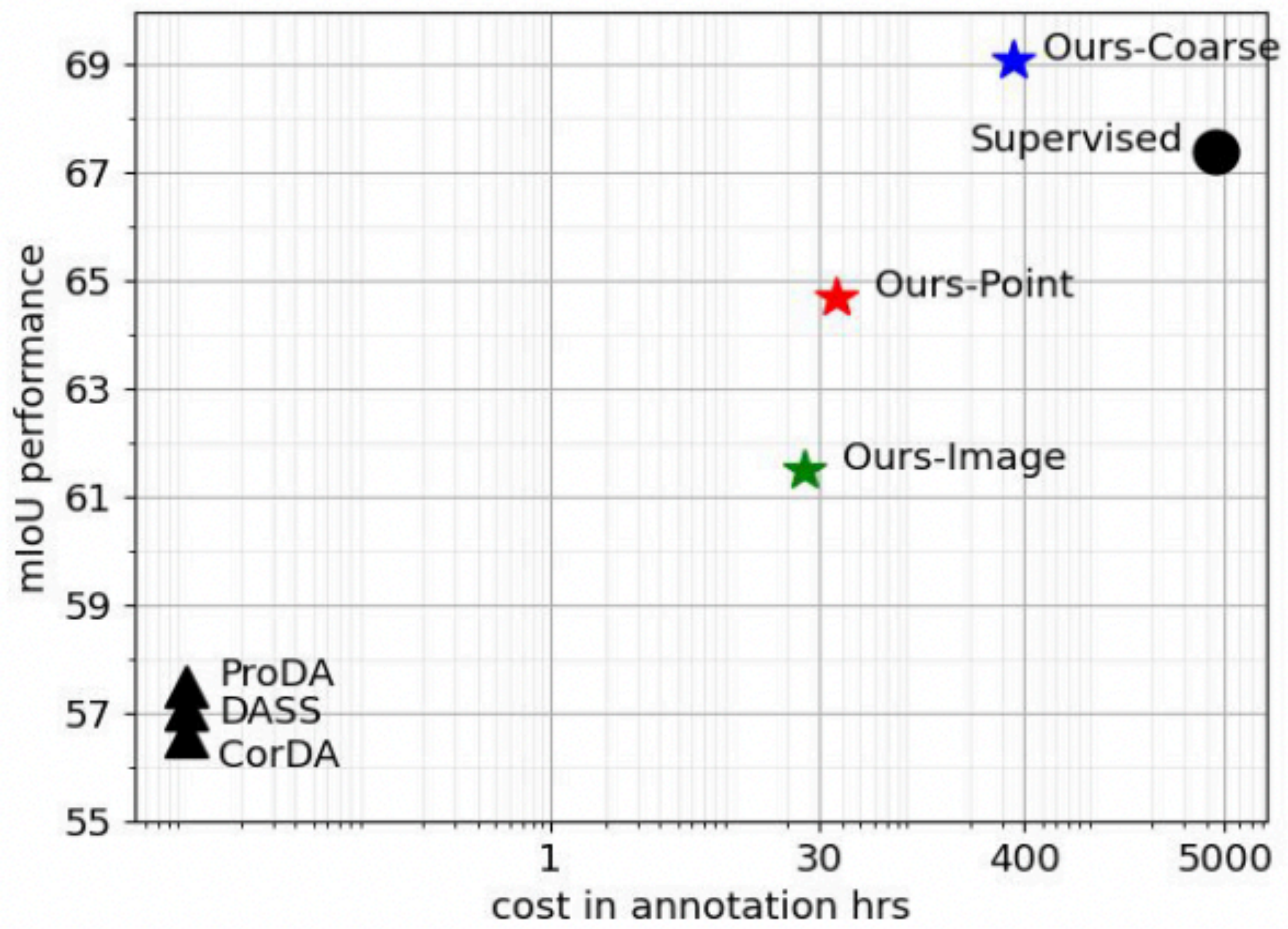
Motivation

- **Huge performance gap** of Unsupervised Domain Adaptive Semantic Segmentation (UDASS) and Supervised Learning

- Idea : **Use additional cheap weak labels from real domain**

Road, Sidewalk, building, traffic sign, car, rider, sky, vegetation

Cost/image: 30 sec 45 sec 7 min 90 min



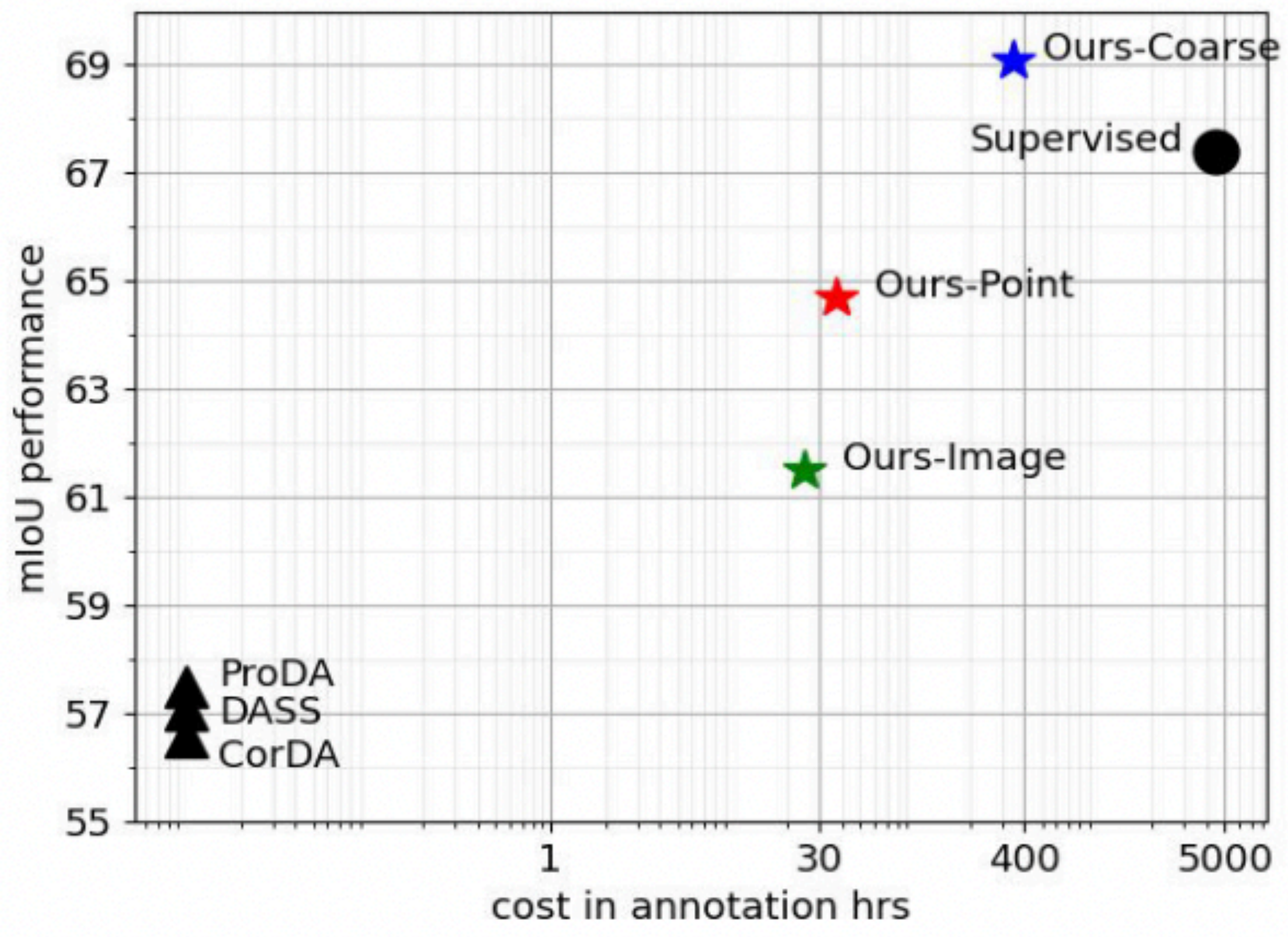
Motivation

- **Huge performance gap** of Unsupervised Domain Adaptive Semantic Segmentation (UDASS) and Supervised Learning

- Idea : **Use additional cheap weak labels from real domain**

Road, Sidewalk, building, traffic sign, car, rider, sky, vegetation

Cost/image:	30 sec	45 sec	7 min	90 min



- Task : **Weakly Supervised Domain Adaptive Semantic Segmentation (WDASS)**

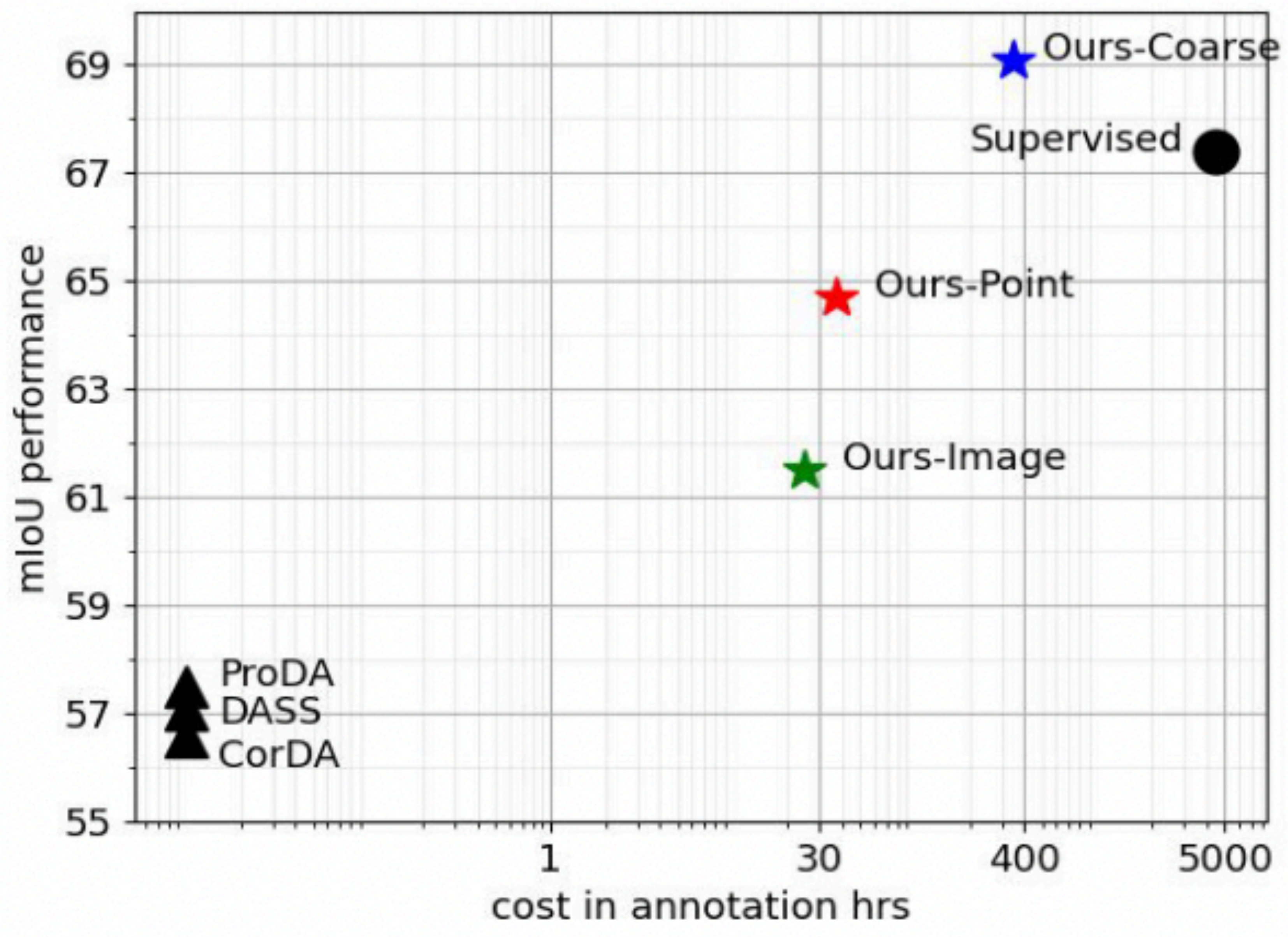
Motivation

- **Huge performance gap** of Unsupervised Domain Adaptive Semantic Segmentation (UDASS) and Supervised Learning

- Idea : **Use additional cheap weak labels from real domain**

Road, Sidewalk, building, traffic sign, car, rider, sky, vegetation

Cost/image: 30 sec 45 sec 7 min 90 min



- Task : **Weakly Supervised Domain Adaptive Semantic Segmentation (WDASS)**

- Prior Works :
  - Ignore source-target feature alignment using weak labels
  - Lack of common framework for different weak labels

Contributions

- **Present common framework for WDASS task using image, point and coarse labels**
  - Two components :
    - Construct better prototypes using weak labels
    - Contrastive alignment of features using prototypes

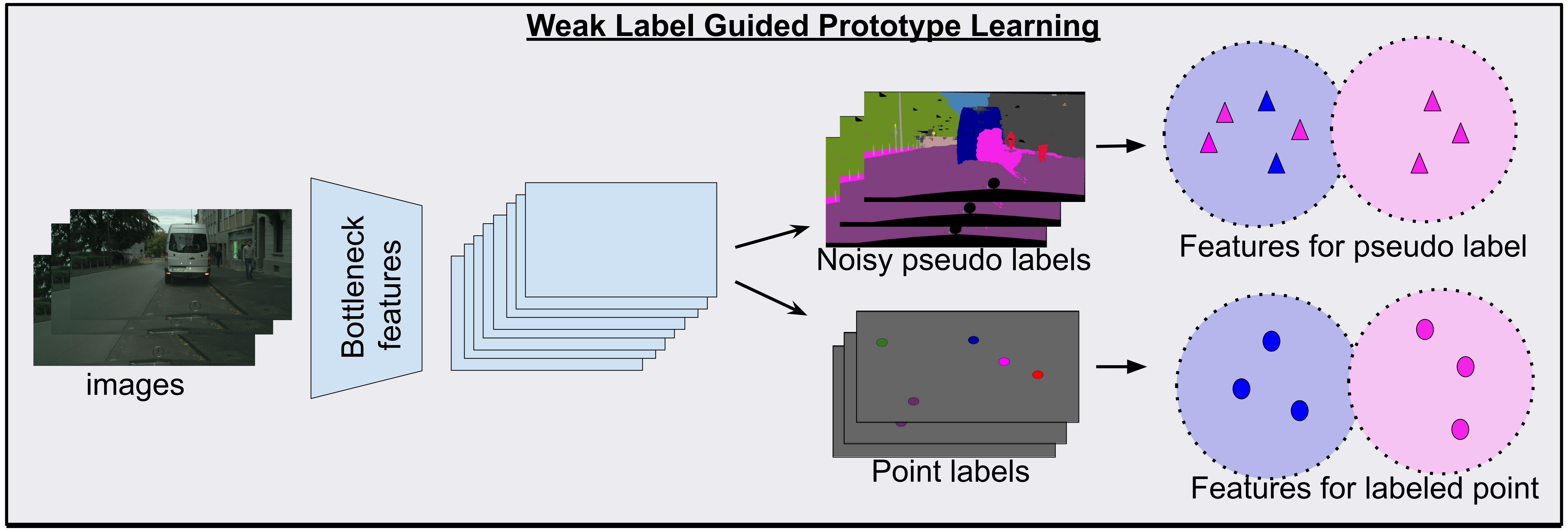
- **Bridge the gap between UDASS and supervised learning**
  - Notably with coarse annotation our framework outperforms supervised learning

- **Show tradeoff between annotation cost vs performance for different weak labels**
  - Point label achieves better performance for low annotation budget

- **Achieves new state-of-the-art on WDASS for different weak labels**

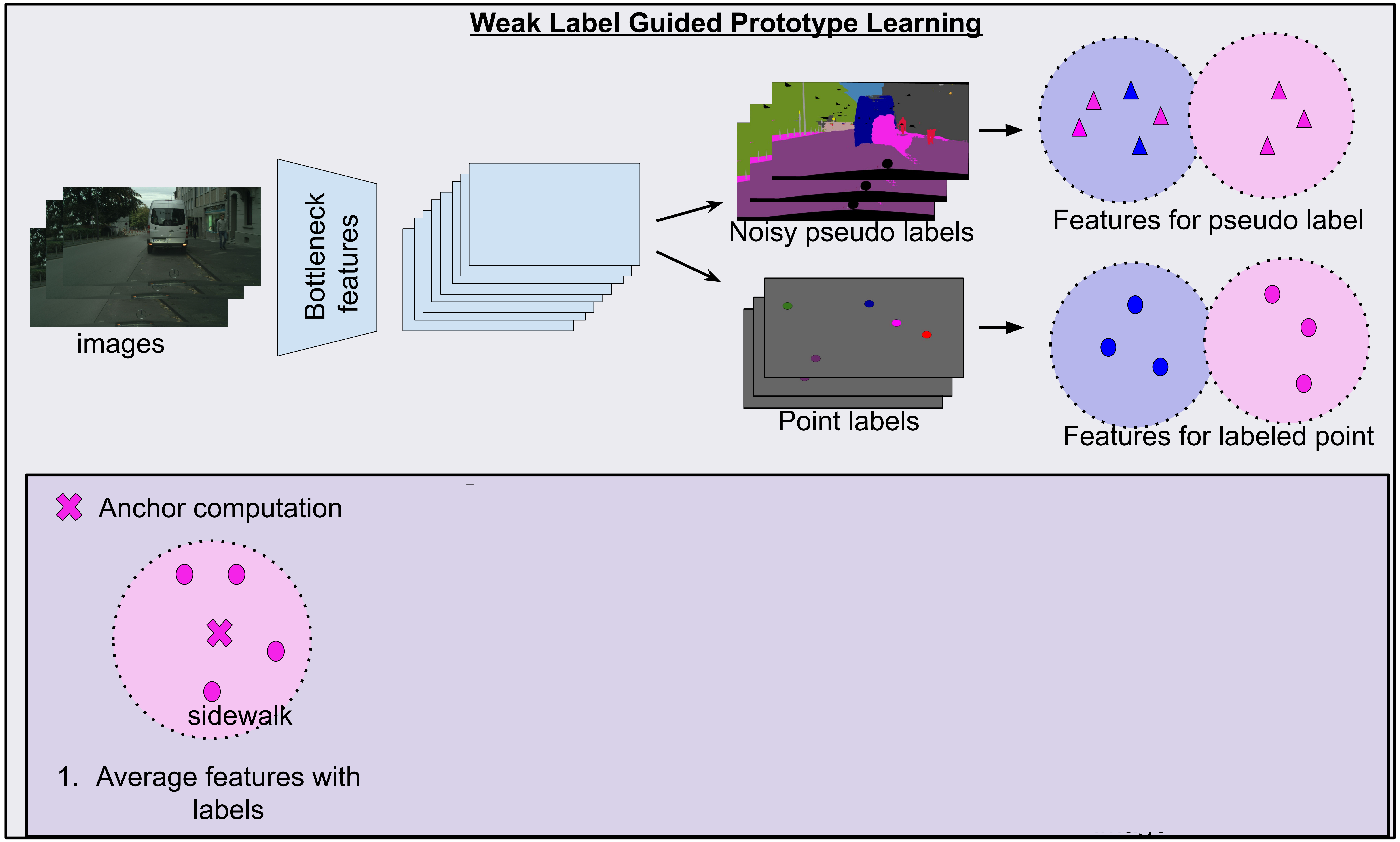
Framework

-



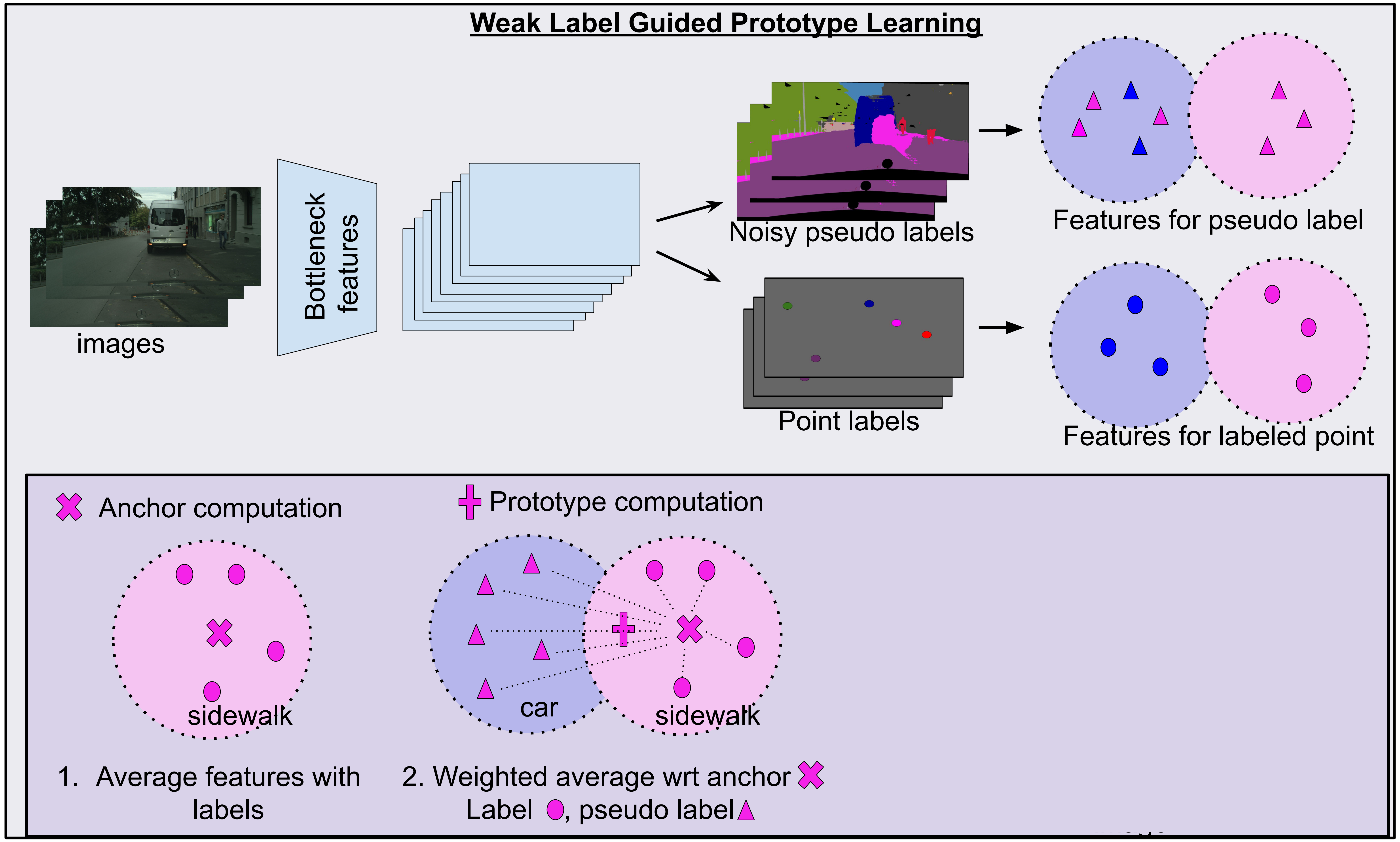
Weak Label Guided Prototype Learning
images
Bottleneck features
Noisy pseudo labels
Features for pseudo label
Point labels
Features for labeled point

Framework



Weak Label Guided Prototype Learning
images
Bottleneck features
Noisy pseudo labels
Features for pseudo label
Point labels
Features for labeled point
Anchor computation
sidewalk
1. Average features with labels

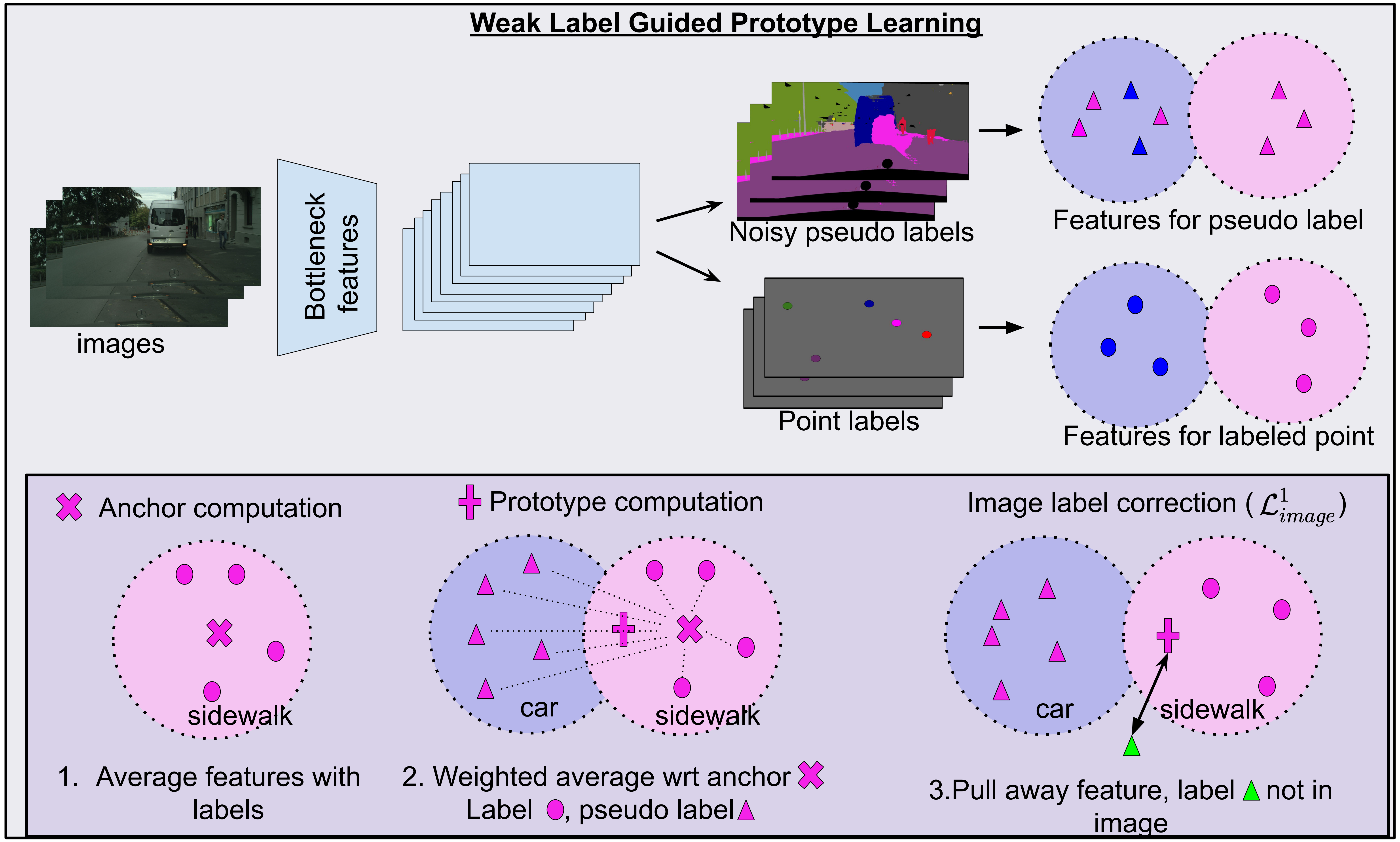
Framework



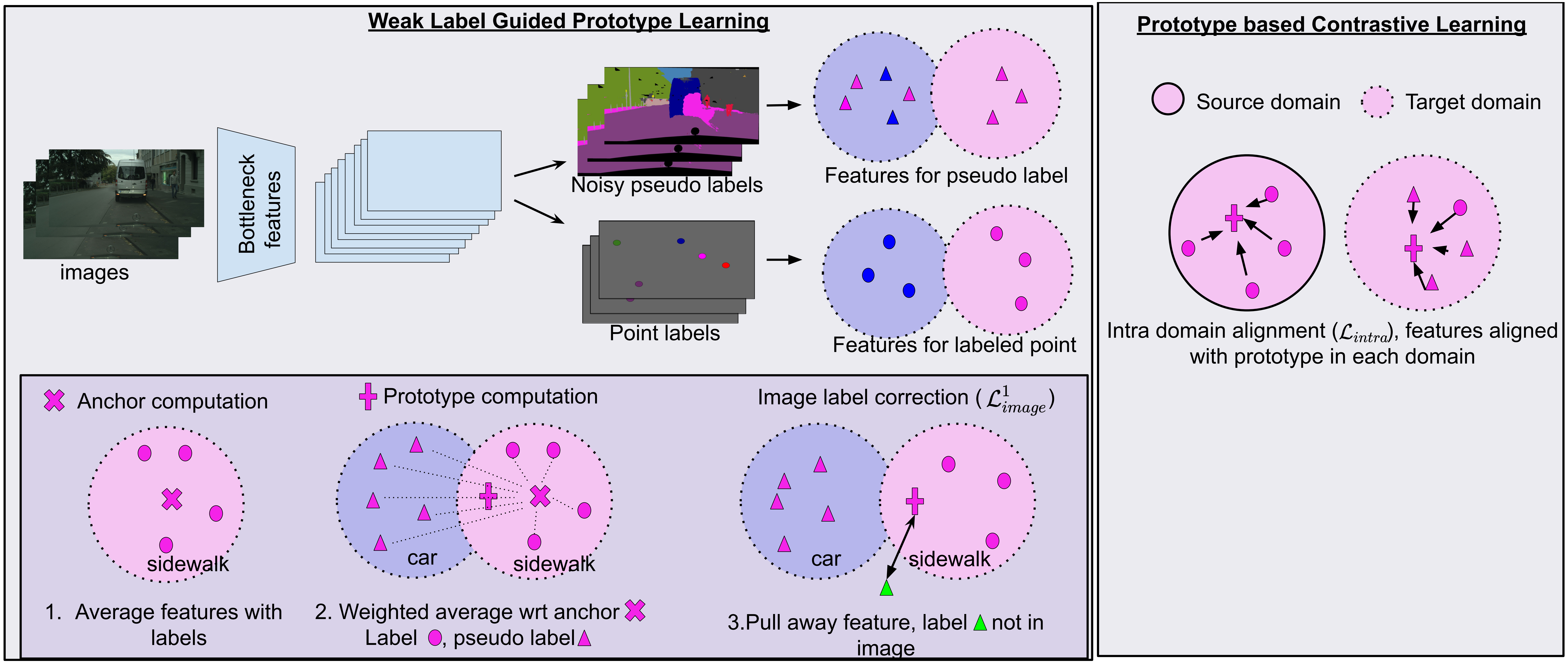
Weak Label Guided Prototype Learning
images
Bottleneck features
Noisy pseudo labels
Point labels
Features for pseudo label
Features for labeled point
Anchor computation
Prototype computation
sidewalk
car
sidewalk
1. Average features with labels
2. Weighted average wrt anchor
Label , pseudo label

Framework

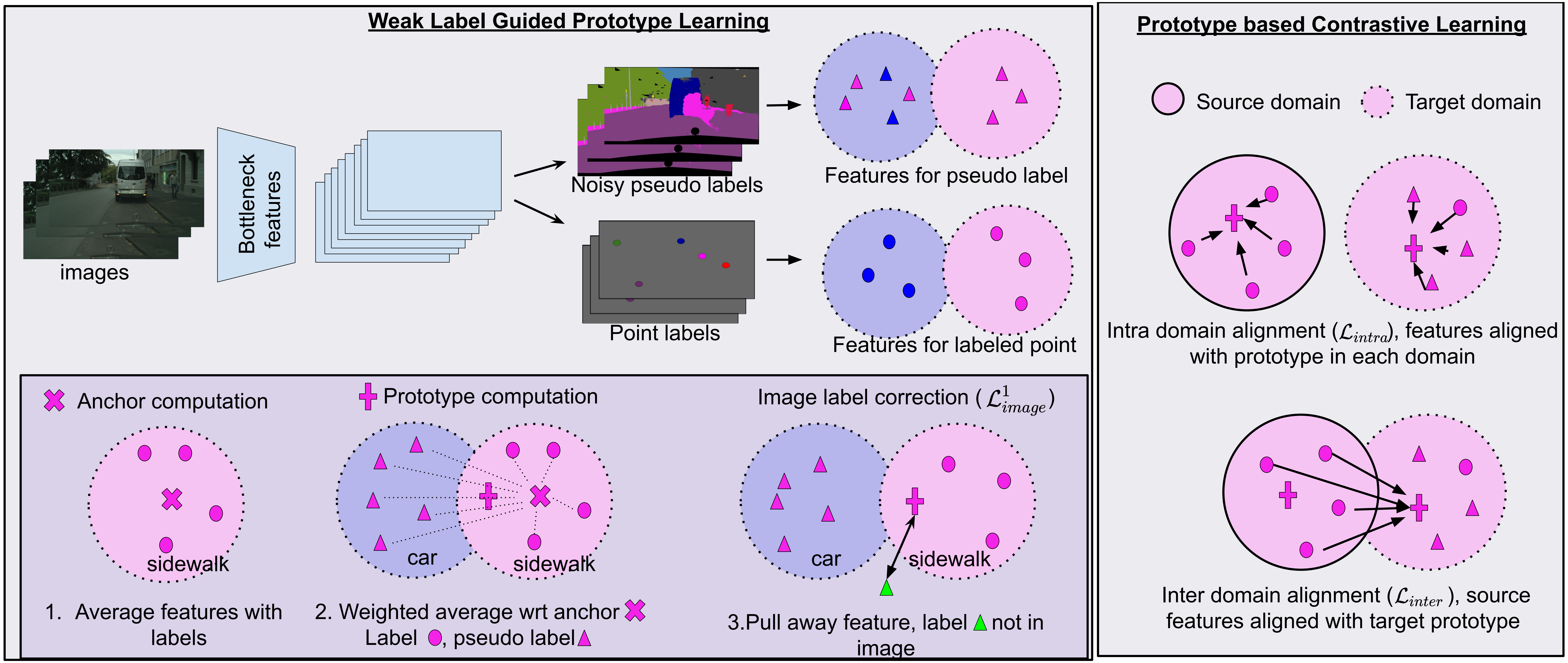
-


Weak Label Guided Prototype Learning
images
Bottleneck features
Noisy pseudo labels
Features for pseudo label
Point labels
Features for labeled point
Anchor computation
Prototype computation
Image label correction ($\mathcal{L}^1_{image}$)
sidewalk
car
sidewalk
car
sidewalk
1. Average features with labels
2. Weighted average wrt anchor
Label , pseudo label
3.Pull away feature, label not in image

Framework


Weak Label Guided Prototype Learning
images
Bottleneck features
Noisy pseudo labels
Features for pseudo label
Point labels
Features for labeled point
Anchor computation
Prototype computation
Image label correction ($\mathcal{L}^1_{image}$)
sidewalk
car
sidewalk
car
sidewalk
1. Average features with labels
2. Weighted average wrt anchor
Label , pseudo label
3.Pull away feature, label not in image
Prototype based Contrastive Learning
Source domain
Target domain
Intra domain alignment ($\mathcal{L}_{intra}$), features aligned with prototype in each domain

Framework



Weak Label Guided Prototype Learning
images
Bottleneck features
Noisy pseudo labels
Features for pseudo label
Point labels
Features for labeled point
Anchor computation
Prototype computation
Image label correction ($\mathcal{L}^1_{image}$)
sidewalk
car
sidewalk
car
sidewalk
1. Average features with labels
2. Weighted average wrt anchor
Label , pseudo label
3.Pull away feature, label not in image
Prototype based Contrastive Learning
Source domain
Target domain
Intra domain alignment ($\mathcal{L}_{intra}$), features aligned with prototype in each domain
Inter domain alignment ($\mathcal{L}_{inter}$), source features aligned with target prototype

Experimental setting

- **Standard UDASS setting**
  - GTA5 to Cityscapes
  - Synthia to Cityscapes
- **Additional weak labels from Cityscapes dataset**
  - image, point and coarse label
- **Metric for evaluation**
  - Mean Intersection of Union (mIoU) score
- **Segmentation network**
  - DeepLabv2 with ImageNet pre-training

Results (Comparison with SoTA)

	Method	GTA5 → Cityscapes mIoU	gap	Synthia → Cityscapes mIoU†	mIoU*	gap
	Source	36.6	+30.8	34.9	40.3	+33.6
UDA	CorDA[1]	56.6	+10.8	55.0	62.8	+11.1
	ProDA[2]	57.5	+9.9	55.5	62.0	+11.9
	DASS [3]	57.1	+10.3	55.6	62.9	+11.0
image	baseline	51.4	+16.0	36.1	39.1	+34.8
	WeakSegDA [4]	53.0	+14.4	50.6	58.5	+15.4
	Ours	**61.5**	**+5.9**	**61.3**	**63.9**	**+10.0**
point	baseline	54.9	+12.9	48.5	53.3	+20.6
	WeakSegDA [4]	56.4	+11.0	57.2	63.7	+10.2
	Ours	**64.7**	**+2.7**	**62.8**	**68.7**	**+5.2**
coarse	baseline	60.8	+6.6	54.6	59.1	+14.8
	Coarse-to-fine [5]	66.7	+0.7	61.6	67.2	+6.7
	Ours	**69.1**	**-1.7**	**66.0**	**71.0**	**+2.9**
	Supervised	67.4	0.0	68.8	73.9	0.0

- Our framework outperforms prior works and baseline by significant difference

[1] Domain adaptive semantic segmentation with self-supervised depth estimation, ICCV 2021
[2] Prototypical pseudo label denoising and target structure learning for domain adaptive semantic segmentation. CVPR 2021
[3] Bi-directional contrastive learning for domain adaptive semantic segmentation, ECCV 2022
[4] Domain adaptive semantic segmentation using weak labels, ECCV 2020
[5] Urban scene semantic segmentation with lowcost coarse annotation, WACV 2023

Results (Comparison with SoTA)

		GTA5 → Cityscapes		Synthia → Cityscapes		
	Method	mIoU	gap	mIoU†	mIoU*	gap
	Source	36.6	+30.8	34.9	40.3	+33.6
UDA	CorDA [1]	56.6	+10.8	55.0	62.8	+11.1
UDA	ProDA [2]	57.5	+9.9	55.5	62.0	+11.9
UDA	DASS [3]	57.1	+10.3	55.6	62.9	+11.0
image	baseline	51.4	+16.0	36.1	39.1	+34.8
image	WeakSegDA [4]	53.0	+14.4	50.6	58.5	+15.4
image	Ours	**61.5**	**+5.9**	**61.3**	**63.9**	**+10.0**
point	baseline	54.9	+12.9	48.5	53.3	+20.6
point	WeakSegDA [4]	56.4	+11.0	57.2	63.7	+10.2
point	Ours	**64.7**	**+2.7**	**62.8**	**68.7**	**+5.2**
coarse	baseline	60.8	+6.6	54.6	59.1	+14.8
coarse	Coarse-to-fine [5]	66.7	+0.7	61.6	67.2	+6.7
coarse	Ours	**69.1**	**-1.7**	**66.0**	**71.0**	**+2.9**
	Supervised	67.4	0.0	68.8	73.9	0.0

- Our framework outperforms prior works and baseline by significant difference
- Our framework bridges gap between UDA and supervised learning.

[1] Domain adaptive semantic segmentation with self-supervised depth estimation, ICCV 2021
[2] Prototypical pseudo label denoising and target structure learning for domain adaptive semantic segmentation. CVPR 2021
[3] Bi-directional contrastive learning for domain adaptive semantic segmentation, ECCV 2022
[4] Domain adaptive semantic segmentation using weak labels, ECCV 2020
[5] Urban scene semantic segmentation with lowcost coarse annotation, WACV 2023

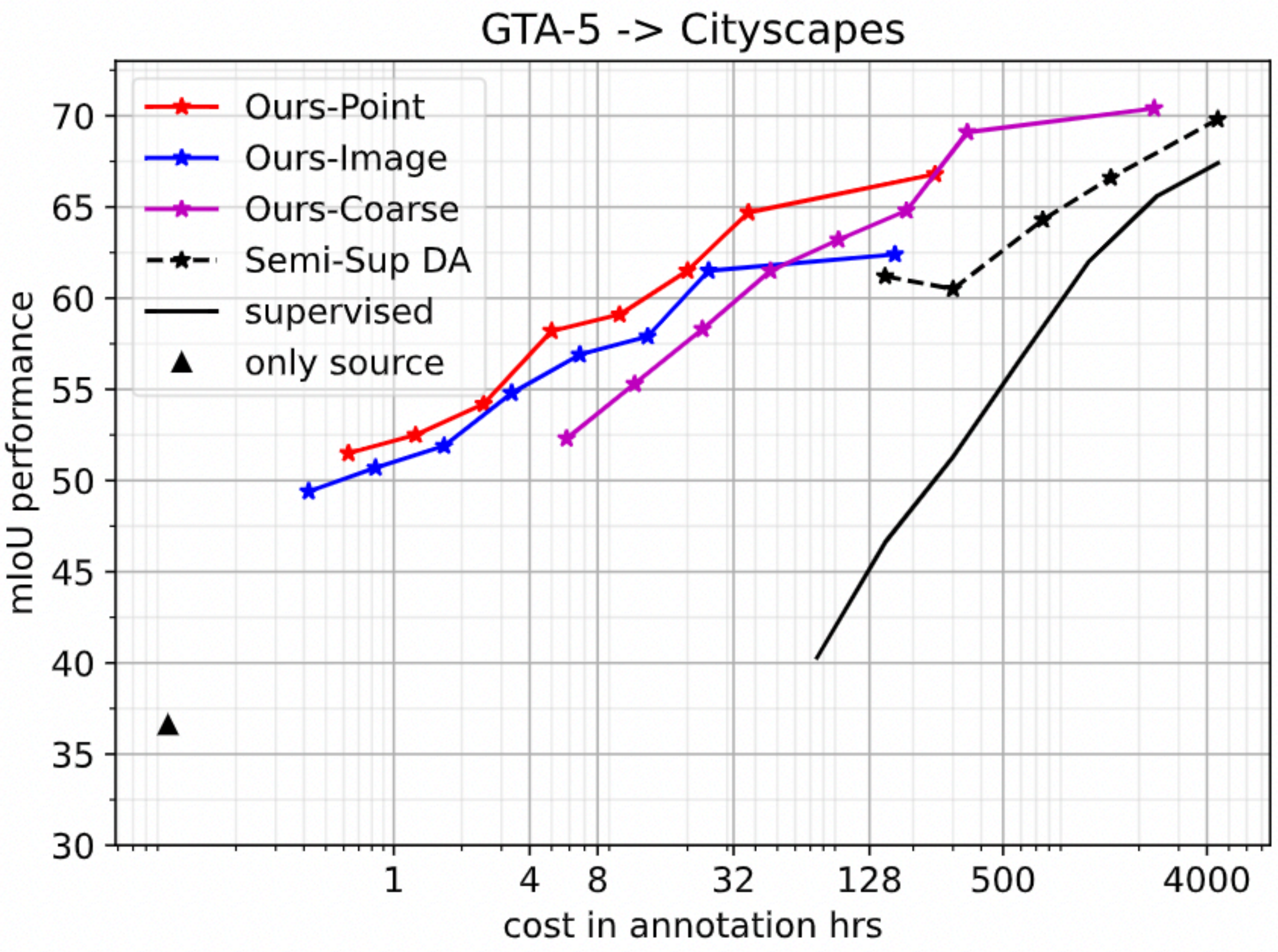
Results (Comparison with SoTA)

		GTA5 → Cityscapes		Synthia → Cityscapes		
	Method	mIoU	gap	mIoU†	mIoU*	gap
	Source	36.6	+30.8	34.9	40.3	+33.6
UDA	CorDA [1]	56.6	+10.8	55.0	62.8	+11.1
UDA	ProDA [2]	57.5	+9.9	55.5	62.0	+11.9
UDA	DASS [3]	57.1	+10.3	55.6	62.9	+11.0
image	baseline	51.4	+16.0	36.1	39.1	+34.8
image	WeakSegDA [4]	53.0	+14.4	50.6	58.5	+15.4
image	Ours	**61.5**	**+5.9**	**61.3**	**63.9**	**+10.0**
point	baseline	54.9	+12.9	48.5	53.3	+20.6
point	WeakSegDA [4]	56.4	+11.0	57.2	63.7	+10.2
point	Ours	**64.7**	**+2.7**	**62.8**	**68.7**	**+5.2**
coarse	baseline	60.8	+6.6	54.6	59.1	+14.8
coarse	Coarse-to-fine [5]	66.7	+0.7	61.6	67.2	+6.7
coarse	Ours	**69.1**	**-1.7**	**66.0**	**71.0**	**+2.9**
	Supervised	67.4	0.0	68.8	73.9	0.0

- Our framework outperforms prior works and baseline by significant difference
- Our framework bridges gap between UDA and supervised learning.
- For GTA5 to Cityscapes setting, coarse labels outperforms supervised learning.

[1] Domain adaptive semantic segmentation with self-supervised depth estimation, ICCV 2021
[2] Prototypical pseudo label denoising and target structure learning for domain adaptive semantic segmentation. CVPR 2021
[3] Bi-directional contrastive learning for domain adaptive semantic segmentation, ECCV 2022
[4] Domain adaptive semantic segmentation using weak labels, ECCV 2020
[5] Urban scene semantic segmentation with lowcost coarse annotation, WACV 2023

Results (Cost vs performance)



	Cost / image
Image label	30 sec
Point label	45 sec
Coarse label	7 min
Fine label	90 min

- Our framework using weak labels performs better than supervised learning and semi-supervised domain adaptation
- Within different weak labels, point labels gives best tradeoff for cost vs performance labels outperforms supervised learning.

max planck institut
informatik

- **Poster Session:** WED-PM
- **Poster Id:** 293
- **Date and Time:** June 21, 4:30 pm - 6:00 pm