Classes Matter: A Fine-grained Adversarial Approach to Cross-domain Semantic Segmentation

Haoran Wang*, Tong Shen*, Wei Zhang, Ling-Yu Duan, and Tao Mei

ETH Zurich, JD AI Research, Peking University

ECCV 2020

Presenter: Minho Park

Domain Adaptation

- Synthetic-to-Real Translation on GTAV-to-Cityscapes Labels
    - https://paperswithcode.com/sota/synthetic-to-real-translation-on-gtav-to

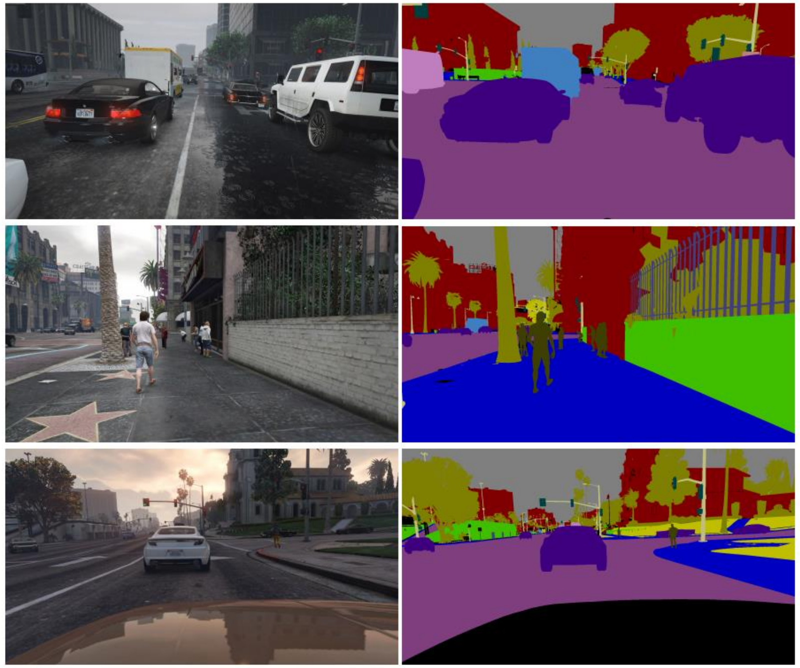
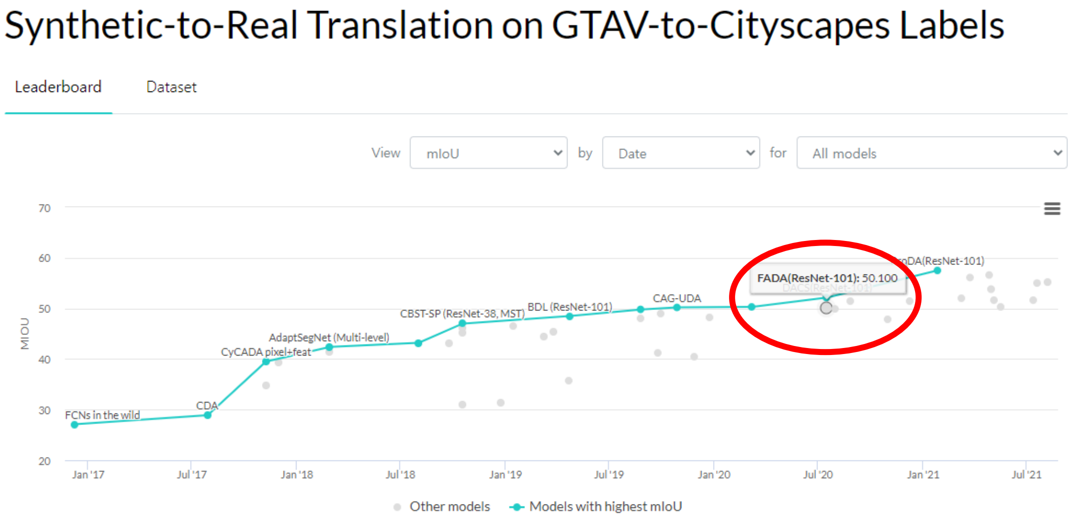


Fig. 1. Images and ground-truth semantic label maps produced by the presented approach. Left: images extracted from the game Grand Theft Auto V. Right: semantic label maps. The color coding is defined in Fig. 4.



S. Richter, V. Vineet et al., Playing for Data: Ground Truth from Computer Games, ECCV 2016, https://download.visinf.tu-darmstadt.de/data/from_games/

Domain Adaptation

- Synthetic-to-Real Translation on GTAV-to-Cityscapes Labels
  - https://paperswithcode.com/sota/synthetic-to-real-translation-on-gtav-to
- E.g. GTA5 → Cityscapes:
- Use all source domain (GTA5) dataset + Unlabeled target domain (Cityscapes) dataset

S. Richter, V. Vineet et al., Playing for Data: Ground Truth from Computer Games, ECCV 2016, https://download.visinf.tu-darmstadt.de/data/from_games/

Other Domain Related Tasks

- Domain Adaptation, Domain Generalization, and Single Domain Generalization.

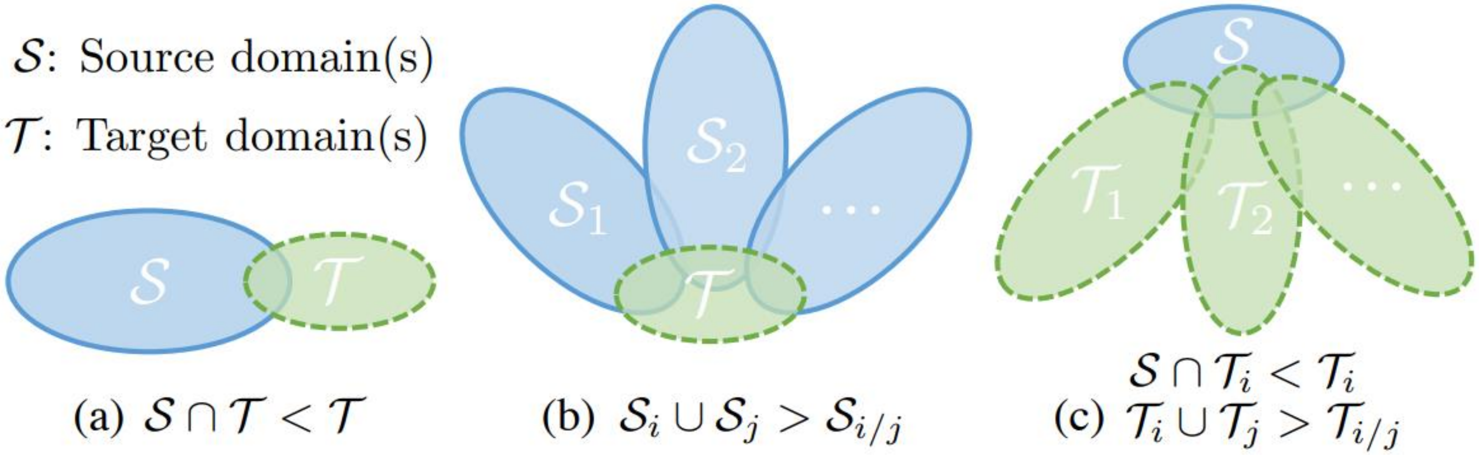


Figure 1. The domain discrepancy: (a) domain adaptation, (b) domain generalization, and (c) single domain generalization.

F. Qiao et al., Learning to Learn Single Domain Generalization, CVPR 2020, https://arxiv.org/abs/2003.13216

Adversarial Framework

- A domain discriminator is trained to distinguish the target from the source.
- At the same time the feature network tries to fool the discriminator **by generating domain-invariant features.**

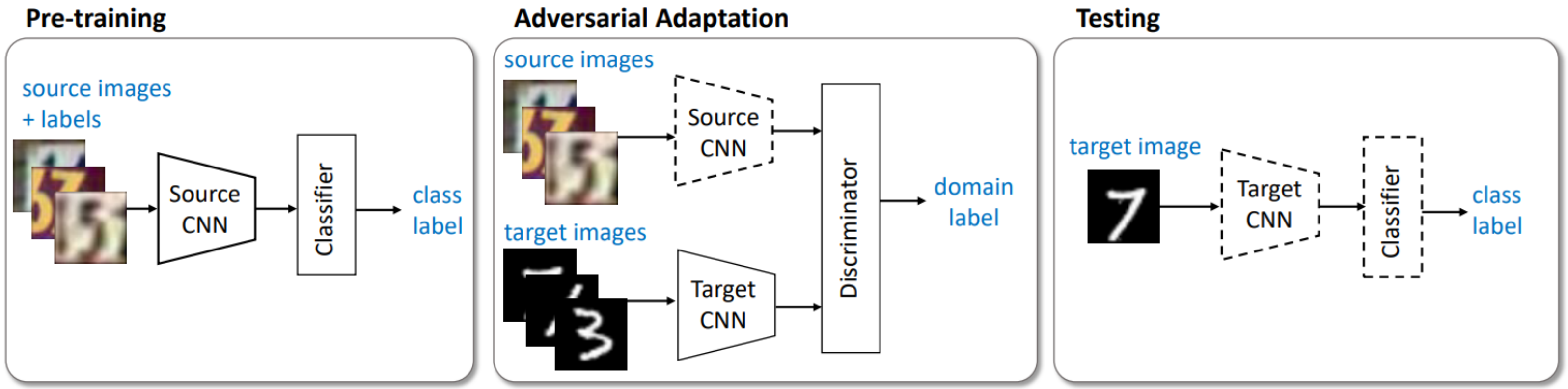


Figure 3: An overview of our proposed Adversarial Discriminative Domain Adaptation (ADDA) approach. We first pre-train a source encoder CNN using labeled source image examples. Next, we perform adversarial adaptation by learning a target encoder CNN such that a discriminator that sees encoded source and target examples cannot reliably predict their domain label. During testing, target images are mapped with the target encoder to the shared feature space and classified by the source classifier. Dashed lines indicate fixed network parameters.

E. Tzeng et al., Adversarial Discriminative Domain Adaptation, CVPR 2017, https://arxiv.org/abs/1702.05464

Motivation

- The class conditional distributions should also be aligned, meaning that class-level alignment also plays an important role.

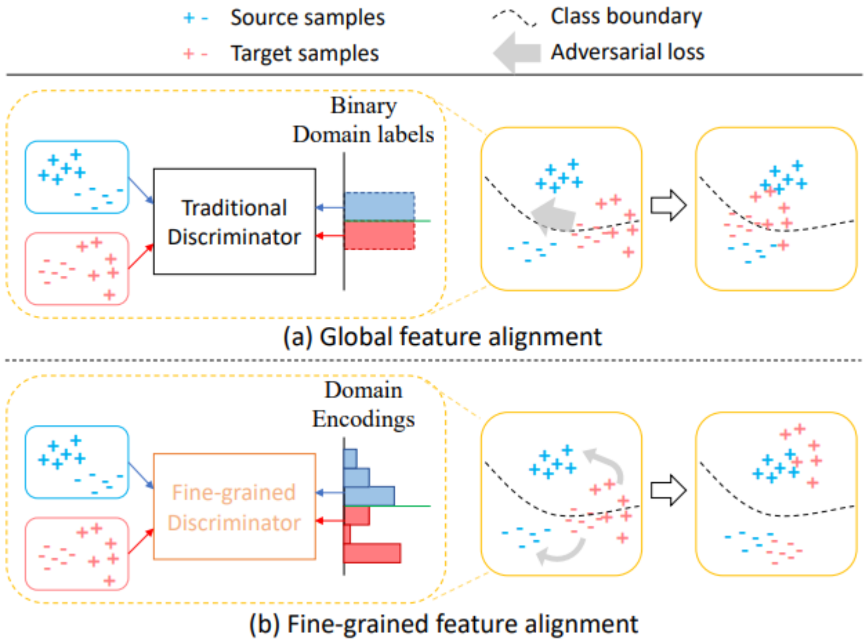


Fig. 1: Illustration of traditional and our fine-grained adversarial learning.

Contribution

- We propose a **fine-grained adversarial learning framework** for cross-domain semantic segmentation that **explicitly incorporates class-level information.**
- The fine-grained learning framework enables **class-level feature alignment**, which is further verified by analysis **using Class Center Distance.**
- **SOTA** on popular domain adaptive segmentation tasks including GTA5 → Cityscapes, SYNTHIA → Cityscapes and Cityscapes → Cross-City.

Revisit Traditional Feature Alignment

- Traditional feature-level adversarial training

$$\min_{D} \mathcal{L}_D = -\sum_{i=1}^{n_S} (1-d) \log P(d=0|f_i) - \sum_{j=1}^{n_T} d \log P\left(d=1\middle|f_j\right)$$

$$\min_{F,G} \mathcal{L}_{seg} + \lambda_{adv} \mathcal{L}_{adv}$$

- d: Domain index
- n_S, n_T: Number of samples from each domain
- f: Feature vector from feature extractor

Method Overview

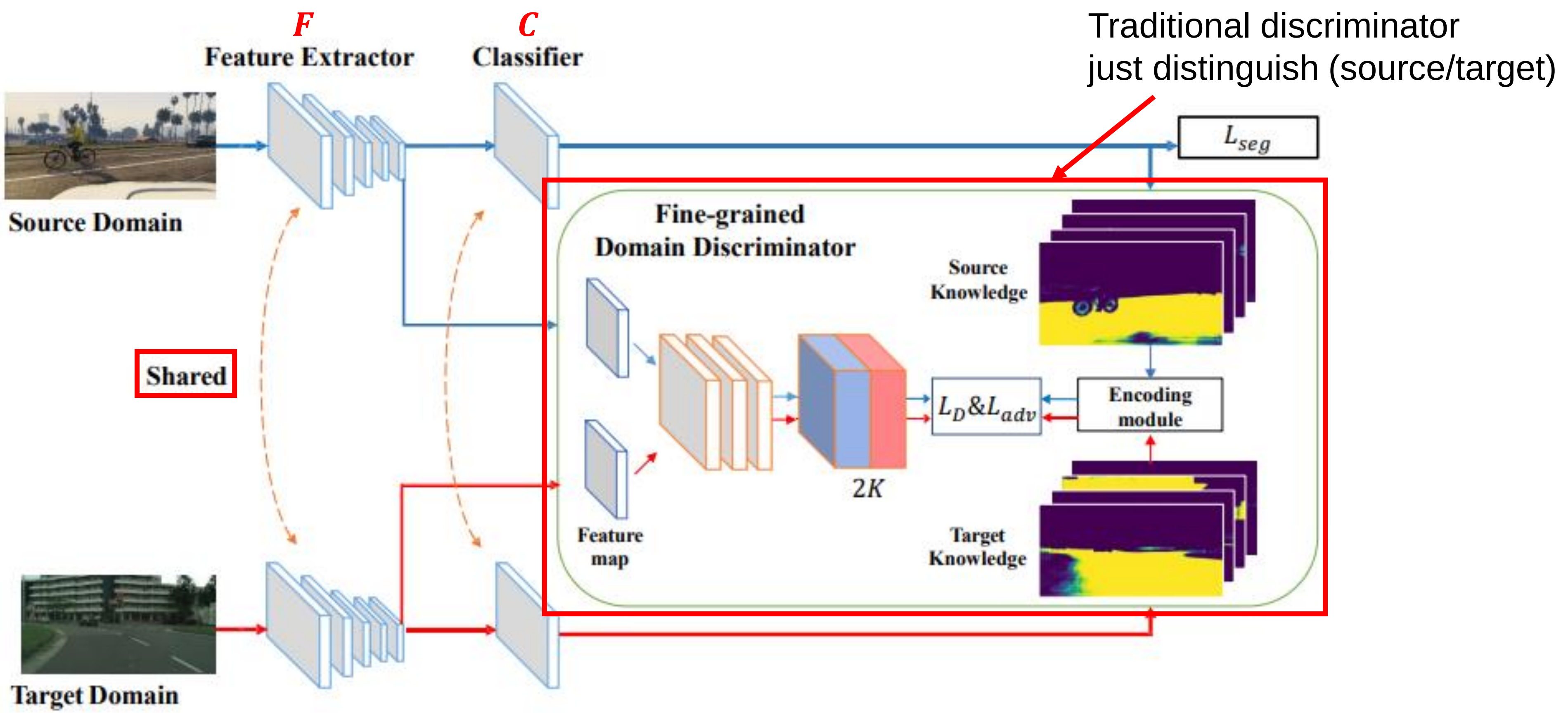


Fig. 2: Overview of the proposed fine-grained adversarial framework.

Fine-grained Domain Discriminator

- Discriminator not only tries to distinguish domains, but also learns to model class structures.

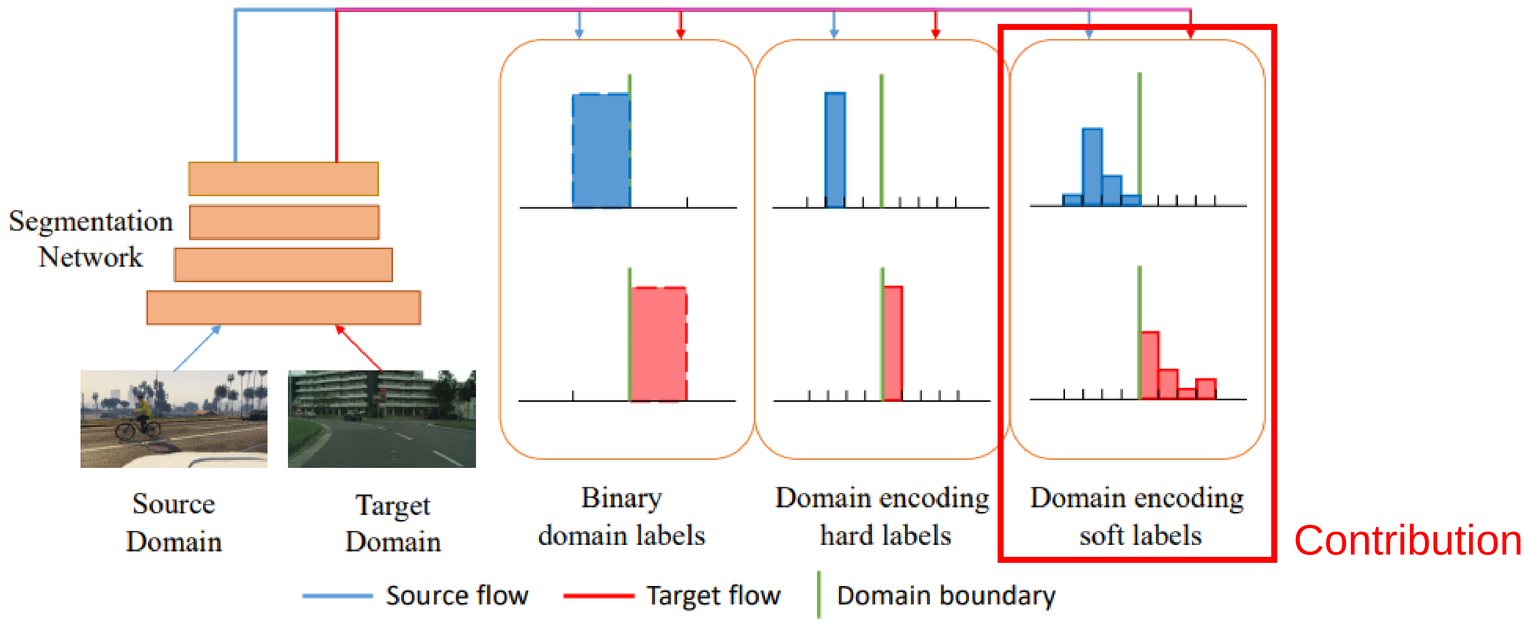


Fig. 3: Illustration of different strategies to generate domain encodings.

Extracting Class Knowledge

- New loss functions

$$\mathcal{L}_D = -\sum_{i=1}^{n_S}\sum_{k=1}^{K} a_{ik}^{(s)} \log P(d=0, c=k|f_i) - \sum_{j=1}^{n_T}\sum_{k=1}^{K} a_{jk}^{(t)} \log P\left(d=1, c=k\middle|f_j\right)$$

$$\mathcal{L}_{adv} = -\sum_{j=1}^{n_T}\sum_{k=1}^{K} a_{jk}^{(t)} \log P\left(d=0, c=k\middle|f_j\right)$$

1. One-hot hard labels: $a_k = \begin{cases} 0, & \text{if } k = \arg\max_k p_k \\ 1, & \text{otherwise} \end{cases}$

2. Soft labels: $a_k = \frac{\exp\left(\frac{z_k}{T}\right)}{\sum_{j=1}^{K} \exp\left(\frac{z_j}{T}\right)}$

p_k: kth softmax probability
z_k: kth entry of logits
T: Temperature

Practically, clipping by given thresholding to achieve more stable performance.

Experiments

Not SOTA now…
Today: **ProDA !**

Table 3: Experimental results for GTA5 → Cityscapes.

GTA5 → Cityscapes																					
Backbone	Method	Road	SW	Build	Wall	Fence	Pole	TL	TS	Veg.	Terrain	Sky	PR	Rider	Car	Truck	Bus	Train	Motor	Bike	mIoU
VGG-16	FCNs in the wild [16]	70.4	32.4	62.1	14.9	5.4	10.9	14.2	2.7	79.2	21.3	64.6	44.1	4.2	70.4	8.0	7.3	0.0	3.5	0.0	27.1
	CDA [34]	74.9	22.0	71.7	6.0	11.9	8.4	16.3	11.1	75.7	13.3	66.5	38.0	9.3	55.2	18.8	18.9	0.0	16.8	14.6	28.9
	ST [38]	83.8	17.4	72.1	14.6	2.9	16.5	16.0	6.8	81.4	24.2	47.2	40.7	7.6	71.7	10.2	7.6	0.5	11.1	0.9	28.1
	CBST [38]	90.4	50.8	72.0	18.3	9.5	27.2	**28.6**	14.1	82.4	25.1	70.8	42.6	14.5	76.9	5.9	12.5	1.2	14.0	**28.6**	36.1
	CyCADA [15]	85.2	37.2	76.5	21.8	15.0	23.8	22.9	21.5	80.5	31.3	60.7	50.5	9.0	76.9	17.1	28.2	4.5	9.8	0.0	35.4
	AdaptSegNet [30]	87.3	29.8	78.6	21.1	18.2	22.5	21.5	11.0	79.7	29.6	71.3	46.8	6.5	80.1	23.0	26.9	0.0	10.6	0.3	35.0
	SIBAN [19]	83.4	13.0	77.8	20.4	17.5	24.6	22.8	9.6	81.3	29.6	77.3	42.7	10.9	76.0	22.8	17.9	5.7	14.2	2.0	34.2
	CLAN [20]	88.0	30.6	79.2	23.4	20.5	26.1	23.0	14.8	81.6	**34.5**	72.0	45.8	7.9	80.5	**26.6**	29.9	0.0	10.7	0.0	36.6
	AdaptPatch [31]	87.3	35.7	79.5	32.0	14.5	21.5	24.8	13.7	80.4	32.0	70.5	50.5	16.9	81.0	20.8	28.1	4.1	15.5	4.1	37.5
	ADVENT [32]	86.9	28.7	78.7	28.5	25.2	17.1	20.3	10.9	80.0	26.4	70.2	47.1	8.4	81.5	26.0	17.2	**18.9**	11.7	1.6	36.1
	Source only	35.4	13.2	72.1	16.7	11.6	20.7	22.5	13.1	76.0	7.6	66.1	41.1	19.0	69.8	15.2	16.3	0.0	16.2	4.7	28.3
	Baseline (feat. only) [30]	85.7	22.8	77.6	24.8	10.6	22.2	19.7	10.8	79.7	27.8	64.8	41.5	18.4	79.7	19.9	21.8	0.5	16.2	4.2	34.1
	FADA	**92.3**	**51.1**	**83.7**	**33.1**	**29.1**	**28.5**	28.0	**21.0**	**82.6**	32.6	**85.3**	**55.2**	**28.8**	**83.5**	24.4	**37.4**	0.0	**21.1**	15.2	**43.8**
ResNet-101	AdaptSegNet [30]	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
	SIBAN [19]	88.5	35.4	79.5	26.3	24.3	28.5	32.5	18.3	81.2	**40.0**	76.5	58.1	25.8	82.6	30.3	34.4	3.4	21.6	21.5	42.6
	CLAN [20]	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
	AdaptPatch [31]	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
	ADVENT [32]	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
	Source only	65.0	16.1	68.7	18.6	16.8	21.3	31.4	11.2	83.0	22.0	78.0	54.4	33.8	73.9	12.7	30.7	**13.7**	28.1	19.7	36.8
	Baseline (feat. only) [30]	83.7	27.6	75.5	20.3	19.9	27.4	28.3	27.4	79.0	28.4	70.1	55.1	20.2	72.9	22.5	35.7	8.3	20.6	23.0	39.3
	FADA	**92.5**	47.5	85.1	37.6	**32.8**	33.4	33.8	18.4	85.3	37.7	83.5	63.2	39.7	**87.5**	32.9	**47.8**	1.6	**34.9**	**39.5**	49.2
	FADA-MST	91.0	50.6	**86.0**	**43.4**	29.8	**36.8**	43.4	25.0	**86.8**	38.3	**87.4**	**64.0**	**38.0**	85.2	31.6	46.1	6.5	25.4	37.1	**50.1**

Experiments

- Class Center Distance (CCD) (Modify the Cluster Center Distance)

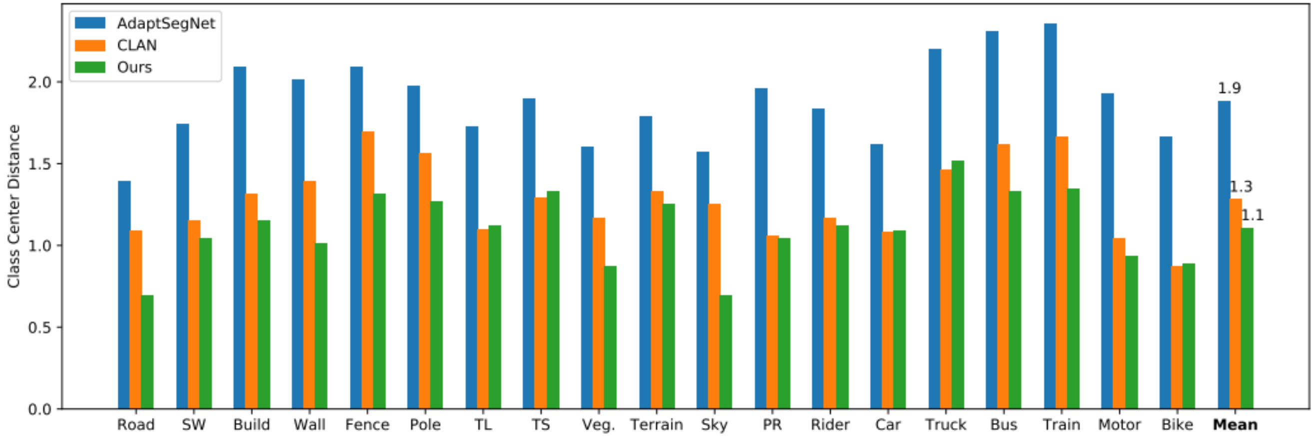


Fig. 4: Quantitative analysis of the feature joint distributions. For each class, we show the Class Center Distance as defined in Equation 9. Our FADA shows a better aligned structure in class-level compared with other state-of-the-art methods.

intra-class

$$CCD(i) = \frac{1}{K-1} \sum_{j=1, j\neq i}^{K} \frac{\frac{1}{|S_i|} \sum_{x \in S_i} \|\boldsymbol{x} - \mu_i\|^2}{\|\mu_i - \mu_j\|}$$

inter-class

References

- S. Richter, V. Vineet et al., Playing for Data: Ground Truth from Computer Games, ECCV 2016, https://download.visinf.tu-darmstadt.de/data/from_games/
- E. Tzeng et al., Adversarial Discriminative Domain Adaptation, CVPR 2017, https://arxiv.org/abs/1702.05464
- F. Qiao et al., Learning to Learn Single Domain Generalization, CVPR 2020, https://arxiv.org/abs/2003.13216