Domain Adaptation on Wheels: Closing the Gap to the Open-world

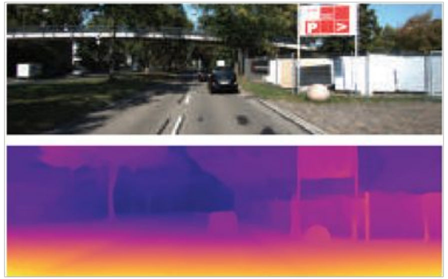
Tuan-Hung Vu
Research scientist
valeo.ai

Dengxin Dai
Director of Research
Huawei Zurich Research Center

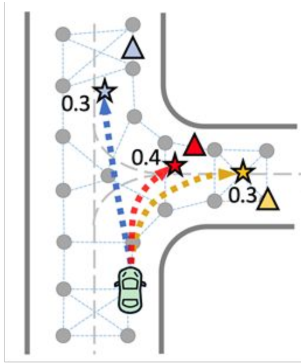
ICCV23
PARIS

Main Perception Tasks for Autonomous Driving

Image Semantic Segmentation

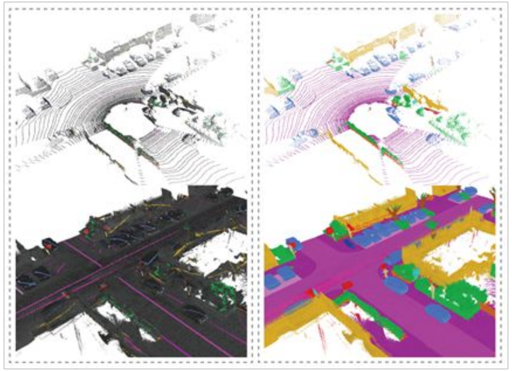


Depth Estimation

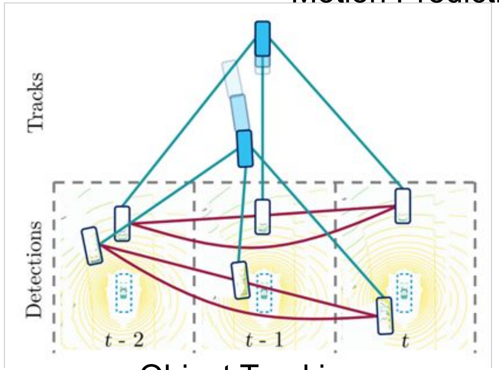


Motion Prediction

3D Object Detection

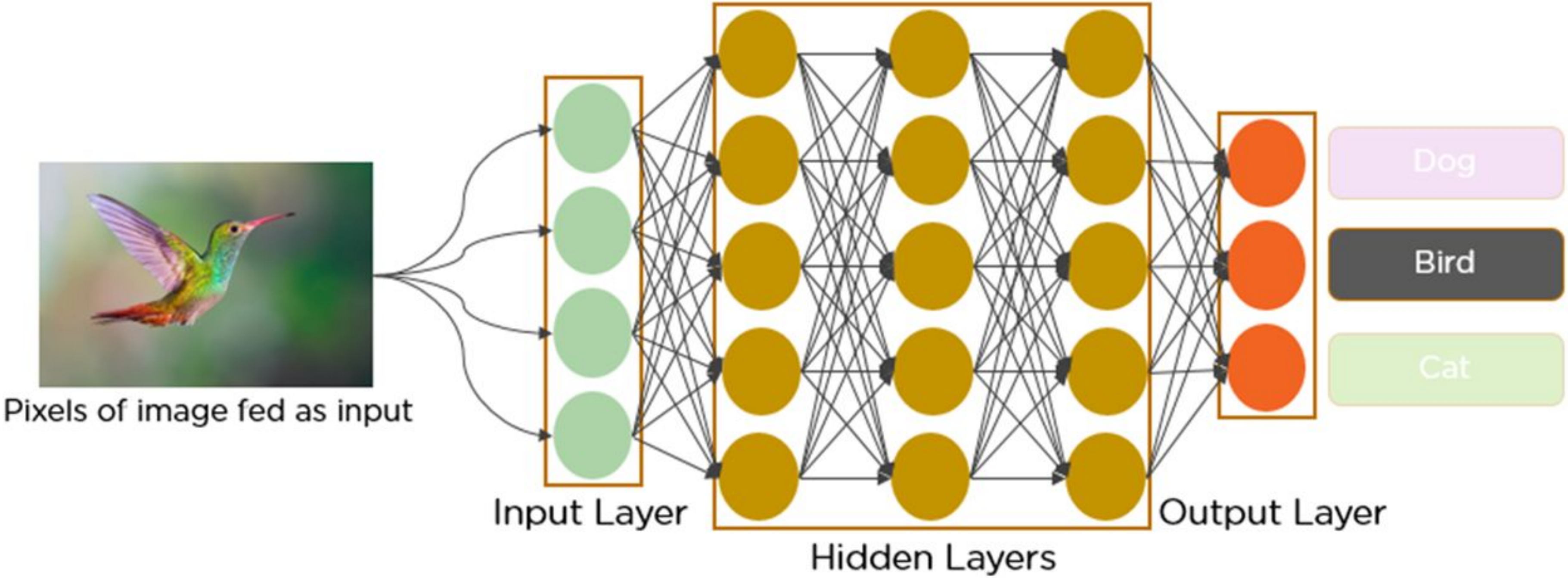


LiDAR Semantic Segmentation



Object Tracking

Perception with Neural Networks



ImageNet Classification

Image Classification on ImageNet

ImageNet Classification

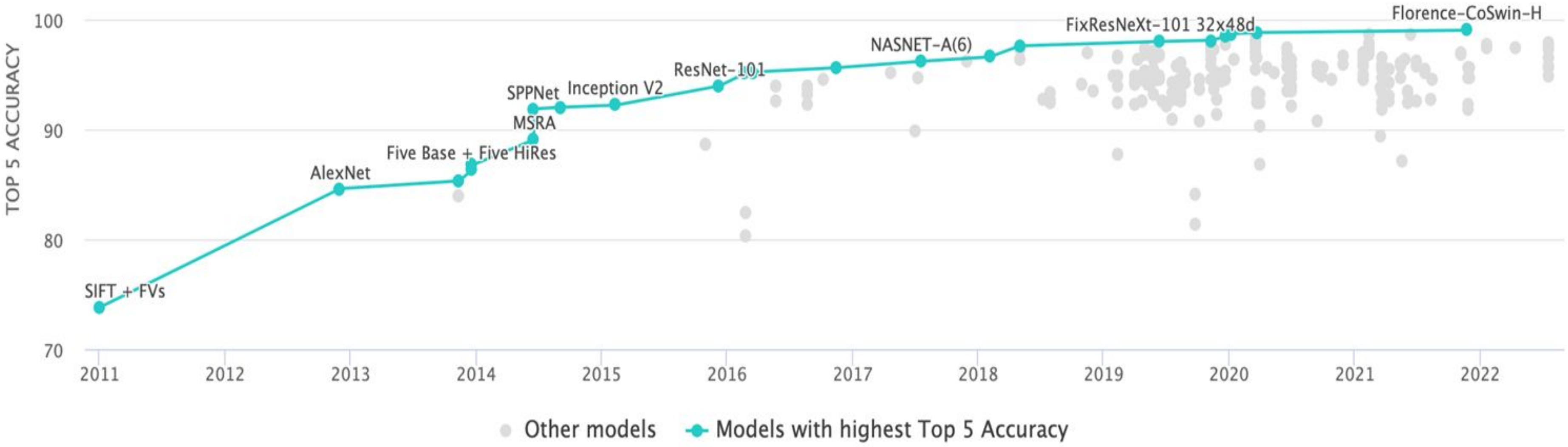
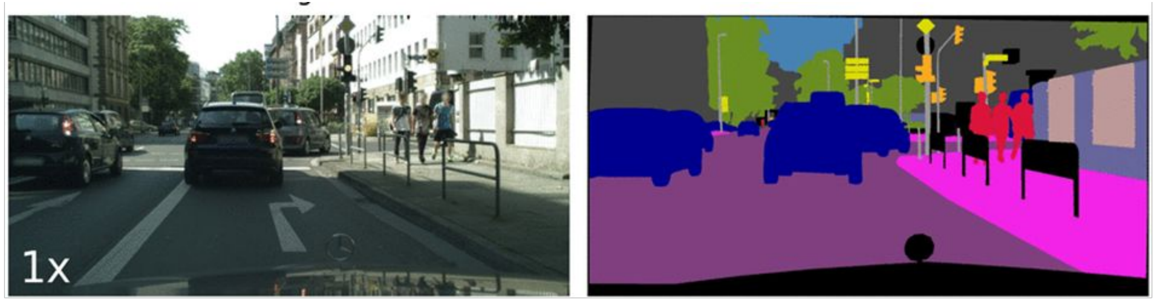


Image Classification on ImageNet

Semantic Segmentation on Cityscapes Dataset

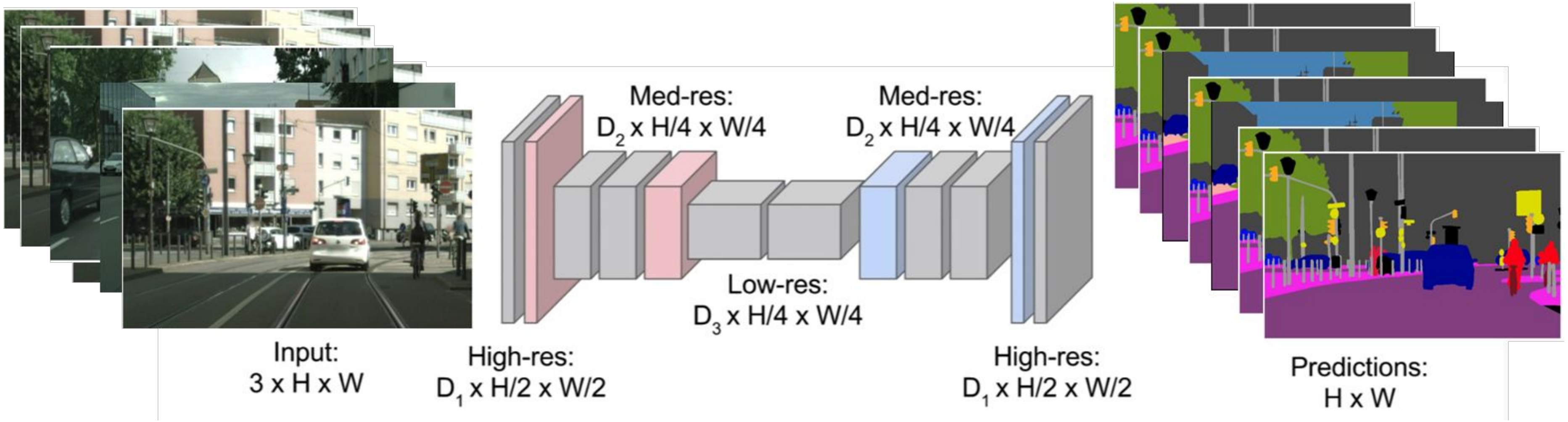


	name	fine	coarse	16-bit	depth	video	sub	IoU class	iIoU class	IoU category	iIoU category
	LeapAI	yes	yes	no	no	no	no	86.4	70.9	93.2	84.2
	MYBank-AIoT	yes	yes	no	no	no	no	86.3	72.9	93.3	85.8
	SAIT SeeThroughNet	yes	yes	no	no	no	no	86.2	71.5	93.2	85.7

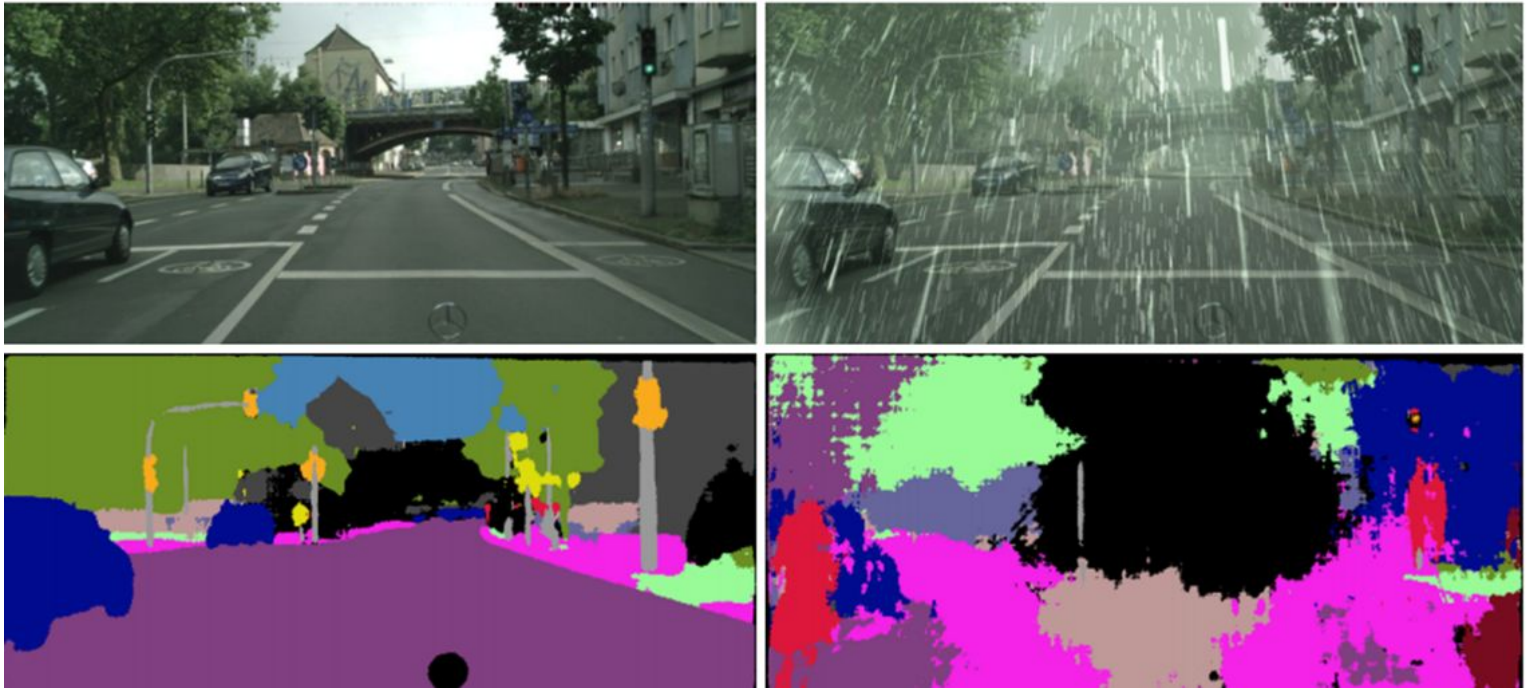
Semantic Segmentation on Cityscapes

Have we solved all perception tasks?

Semantic Segmentation: training and validation



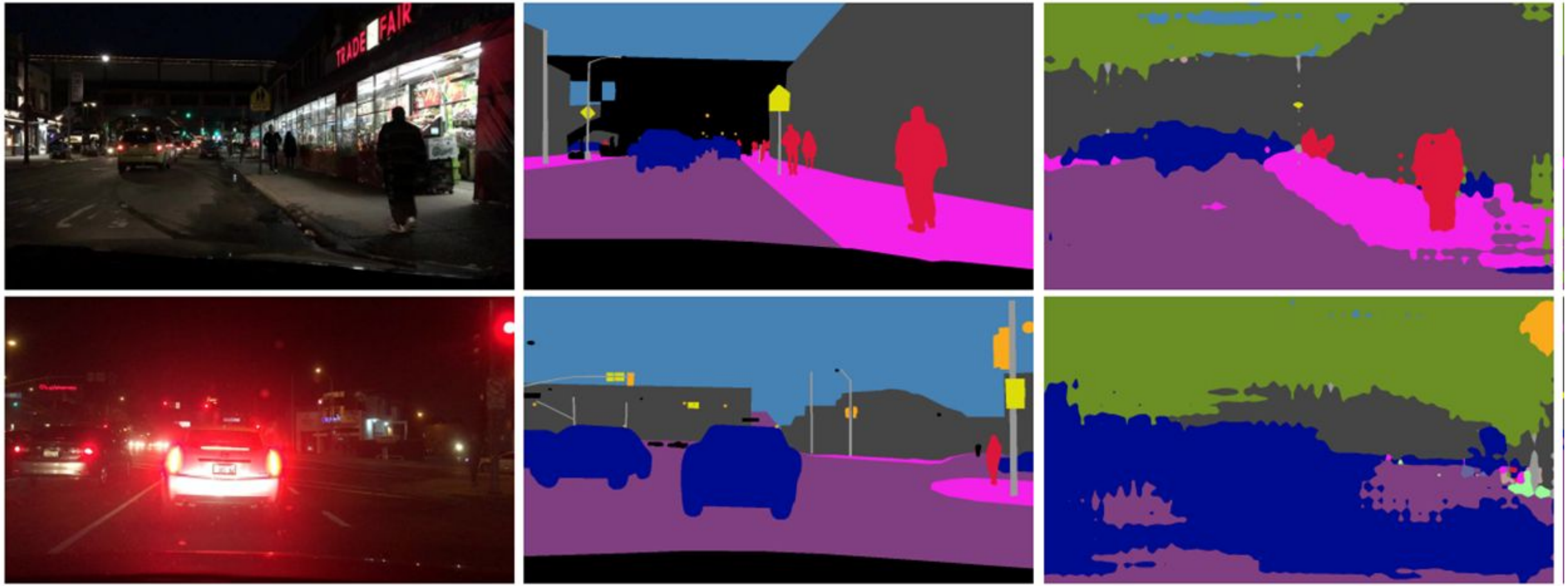
Dataset Bias or Domain Discrepancy



Clear weather Rain

Physics-Based Rendering for Improving Robustness to Rain, Halder, Lalonde, and Charette, ICCV 2019

Dataset Bias or Domain Discrepancy

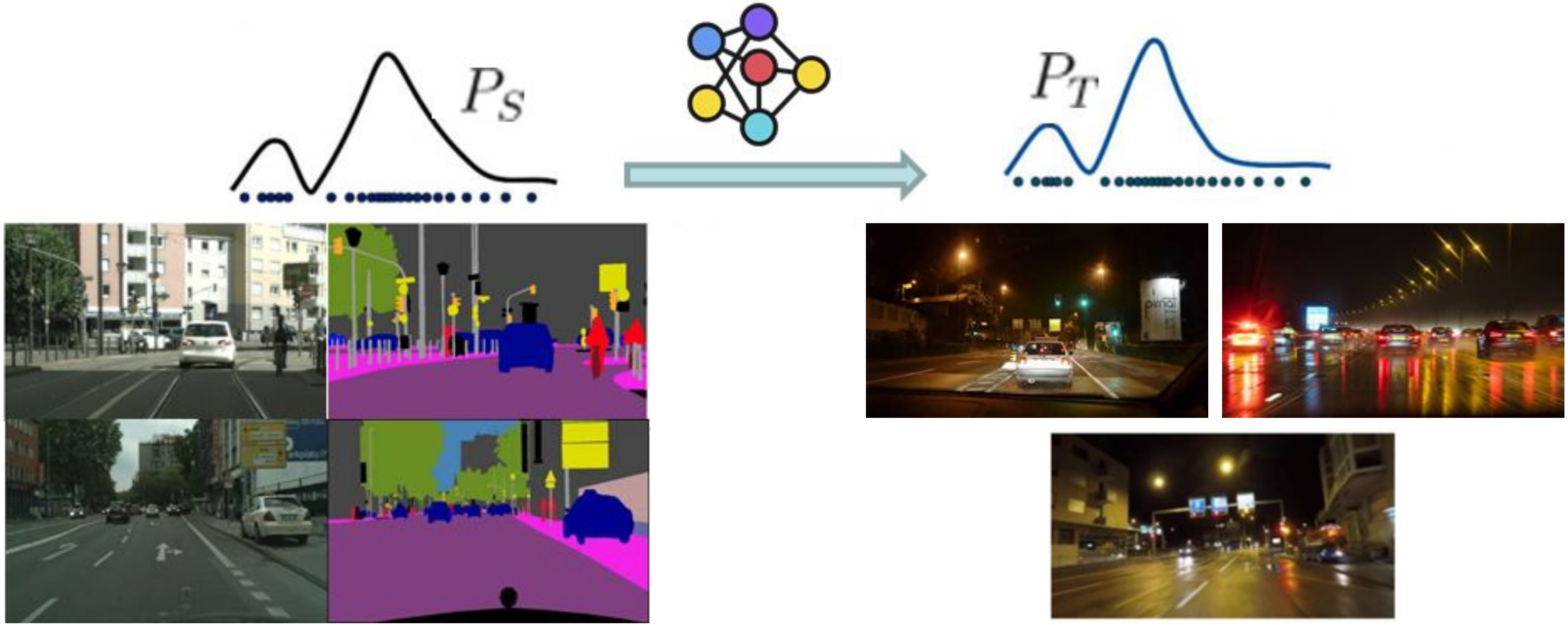


Nighttime Image   Human Annotation   Prediction

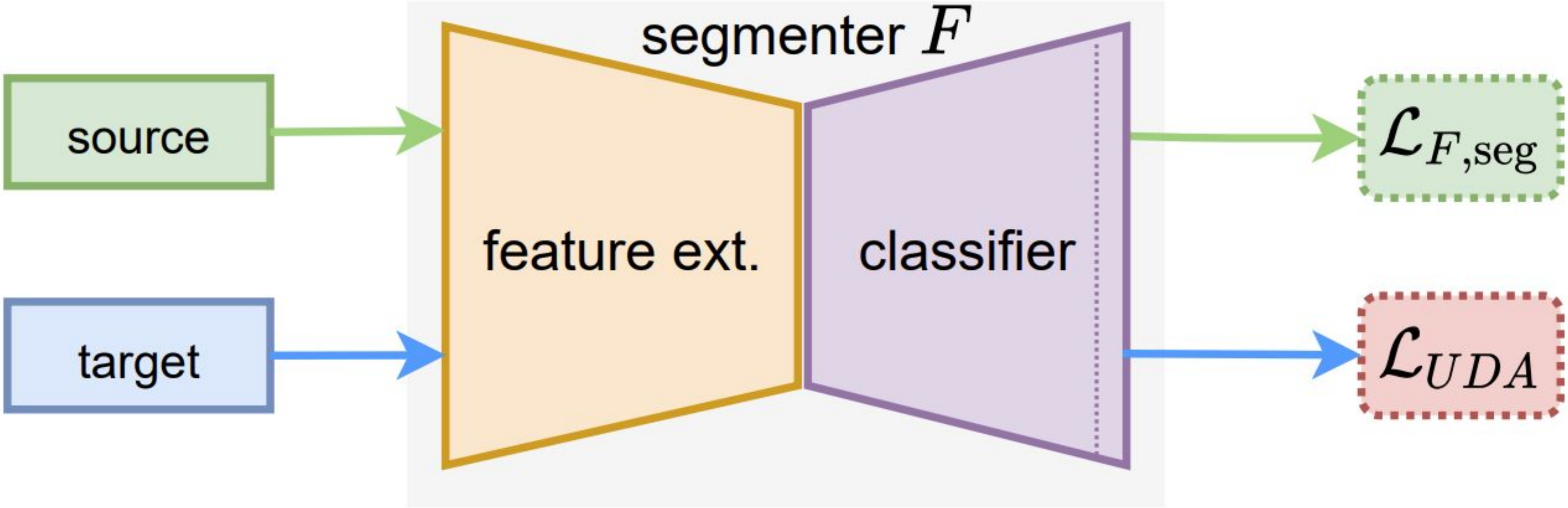
Map-Guided Curriculum Domain Adaptation and Uncertainty-Aware Evaluation for Semantic Nighttime Image Segmentation, Sakaridis, Dai, Van Gool, T-PAMI, 2020

What can we do to generalize?

1. Unsupervised Domain Adaptation: Learning Target Distribution with Unlabeled Samples

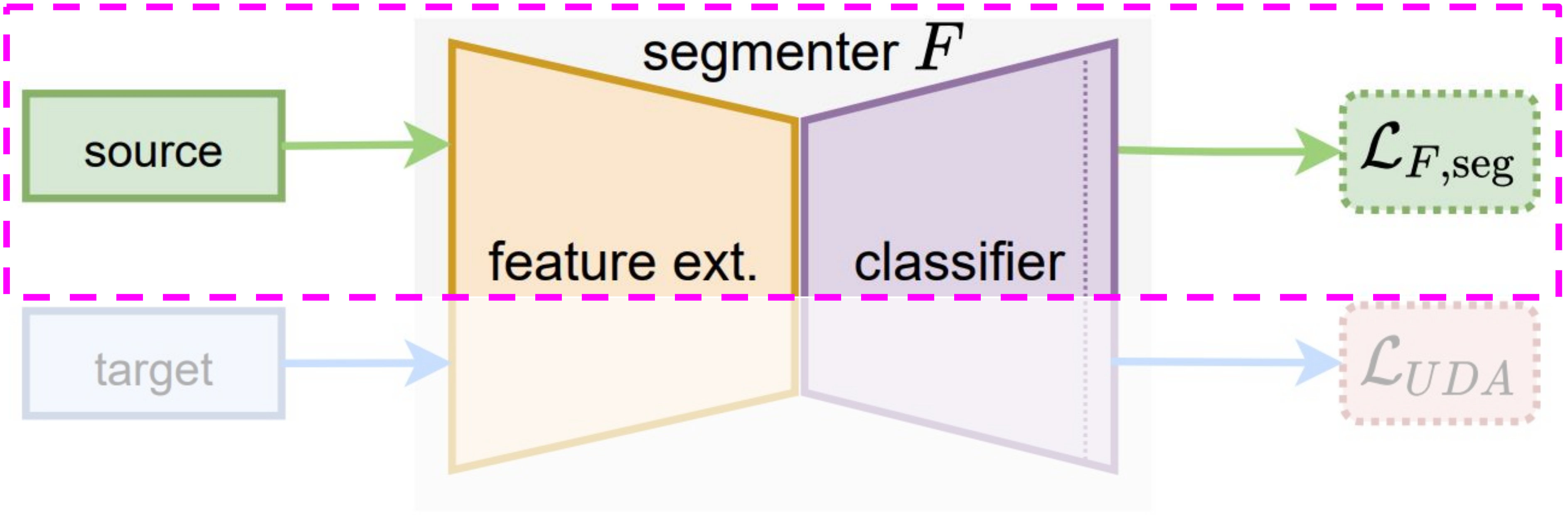


UDA in Semantic Segmentation



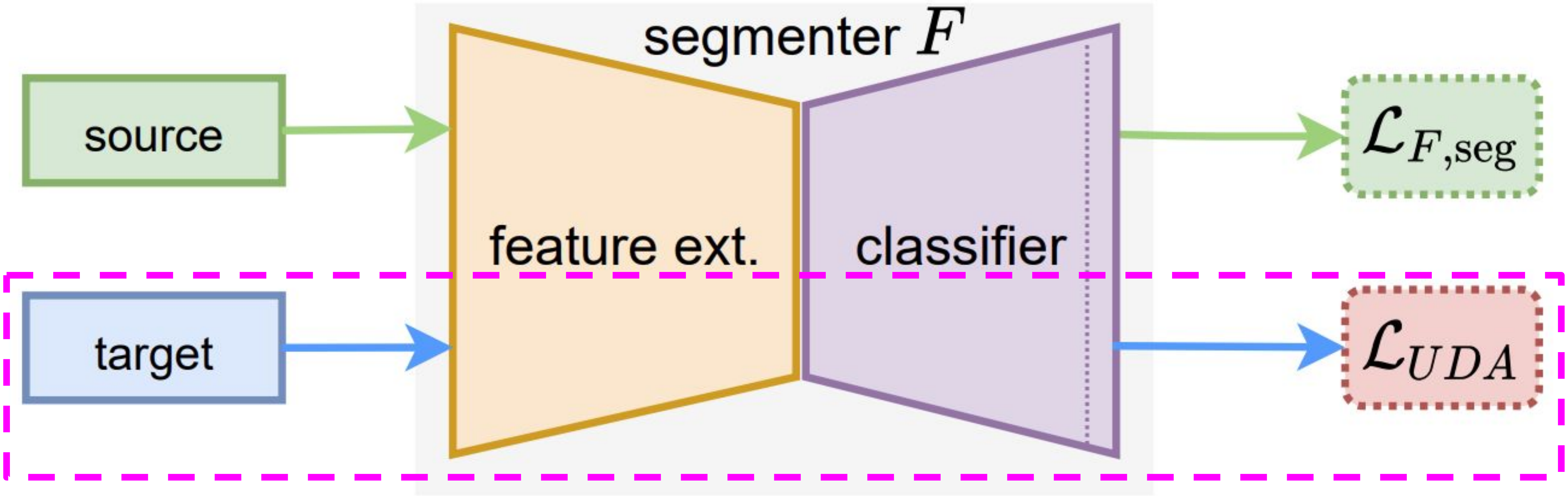
- A general UDA pipeline in segmentation

UDA in Semantic Segmentation



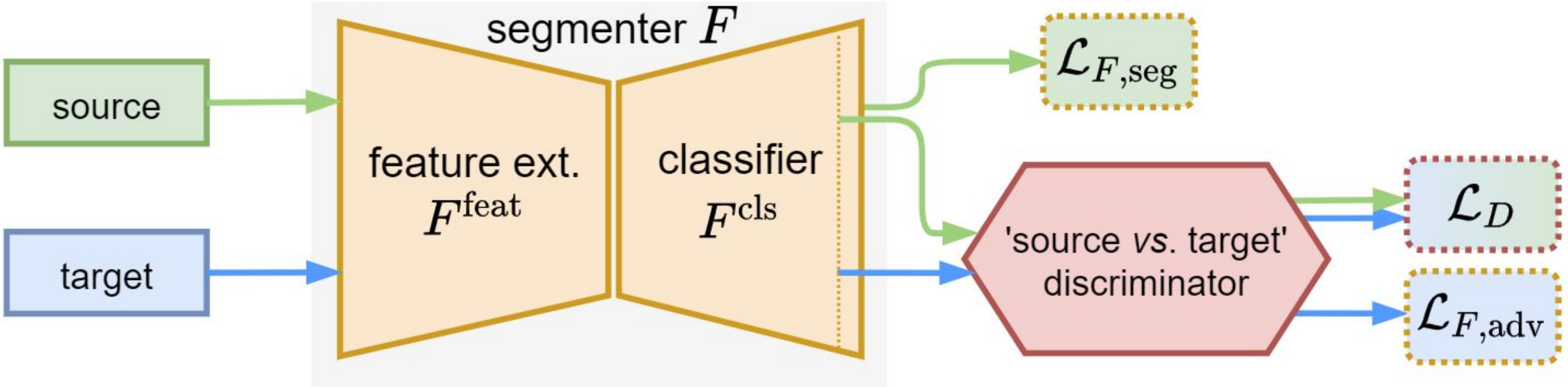
- Supervised training on source

UDA in Semantic Segmentation



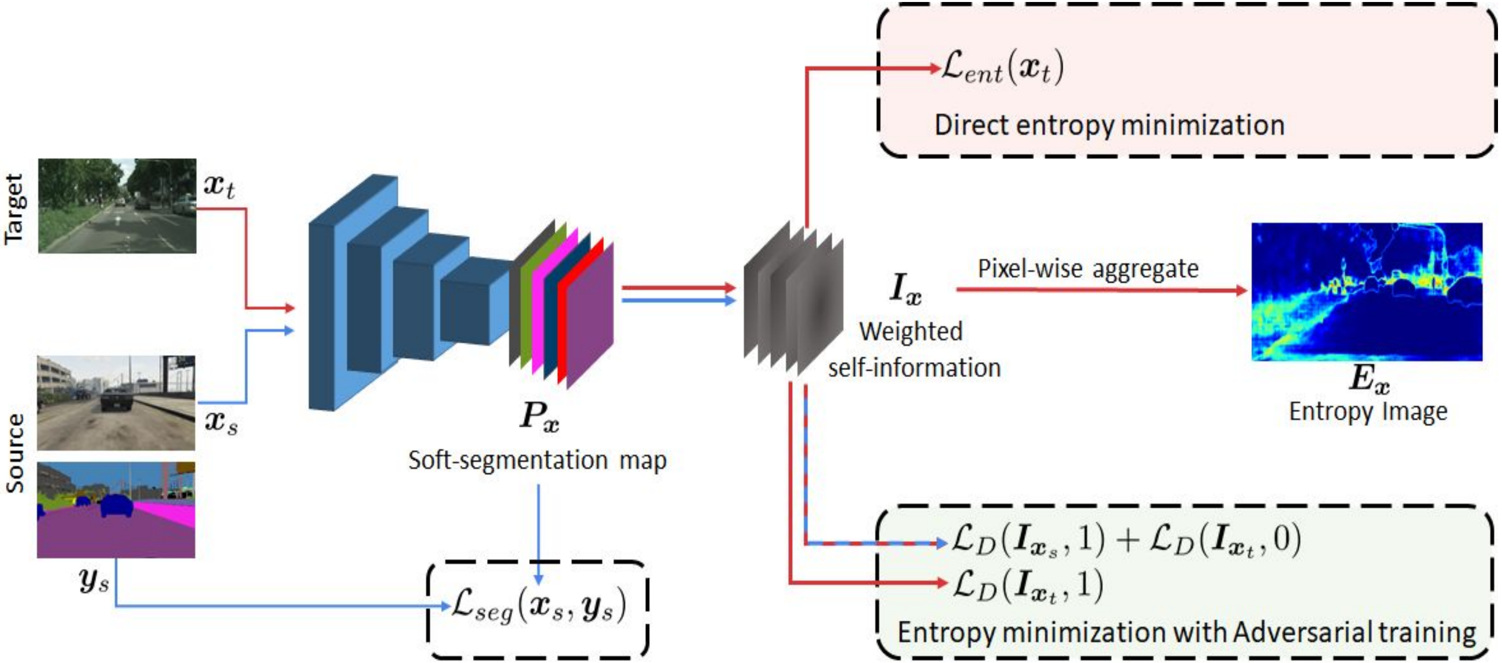
- Different UDA techniques ~ different UDA losses

Adversarial UDA framework in Segmentation



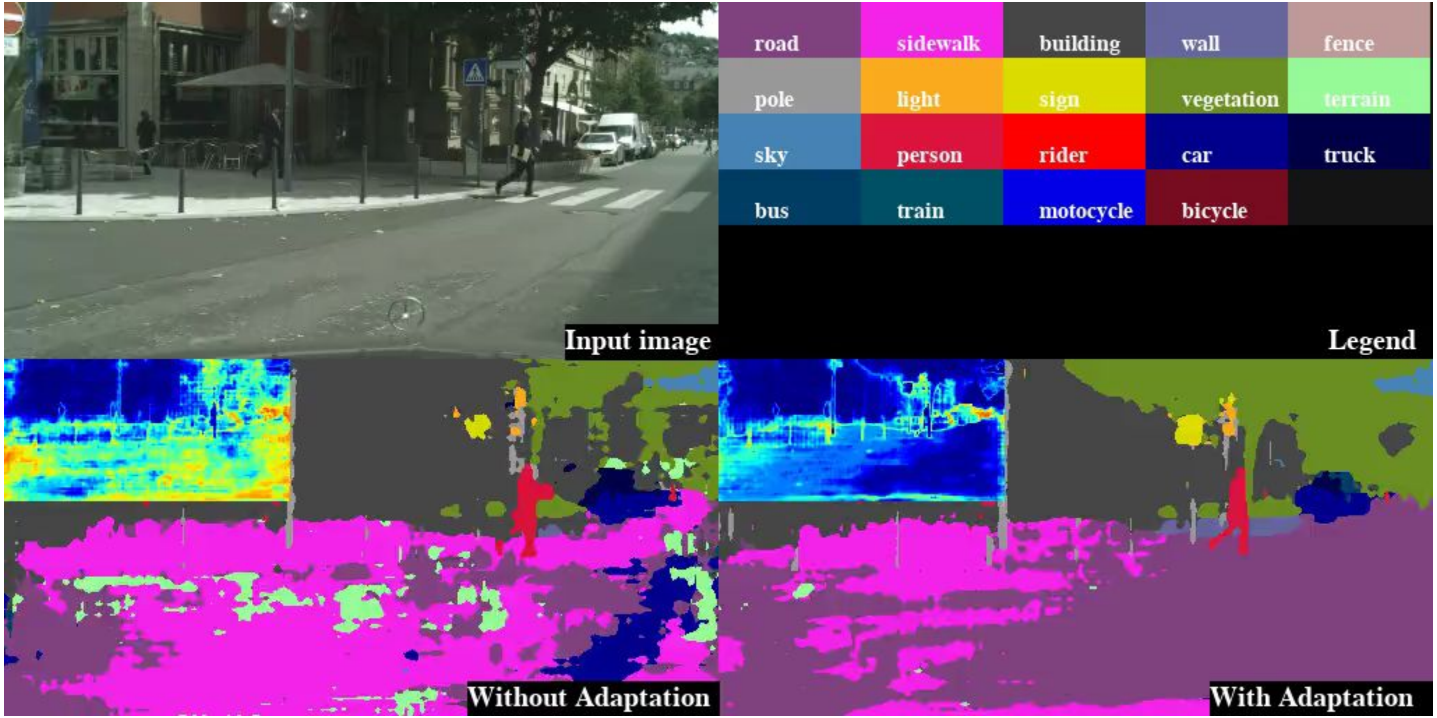
FCNs in the Wild: Pixel-level Adversarial and Constraint-based Adaptation, Hoffman et al. ICLR'17
Learning to Adapt Structured Output Space for Semantic Segmentation, Tsai et al. CVPR'18
ADVENT: Adversarial Entropy Minimization for Domain Adaptation in Semantic Segmentation, Vu et al. CVPR'19

ADVENT: adversarial UDA + entropy minimization



ADVENT: Adversarial Entropy Minimization for Domain Adaptation in Semantic Segmentation, Vu et al. CVPR’19

ADVENT

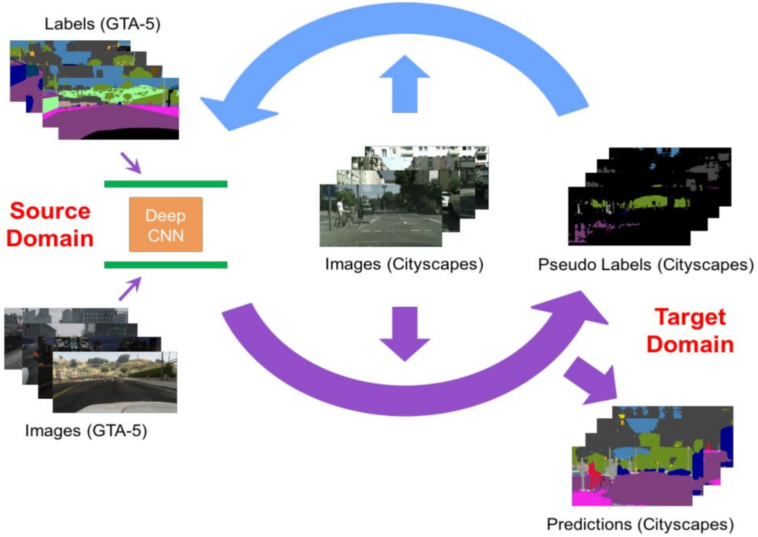


ADVENT: Adversarial Entropy Minimization for Domain Adaptation in Semantic Segmentation, Vu et al. CVPR’19

What did we learn?

- Adversarial training is great but difficult to train

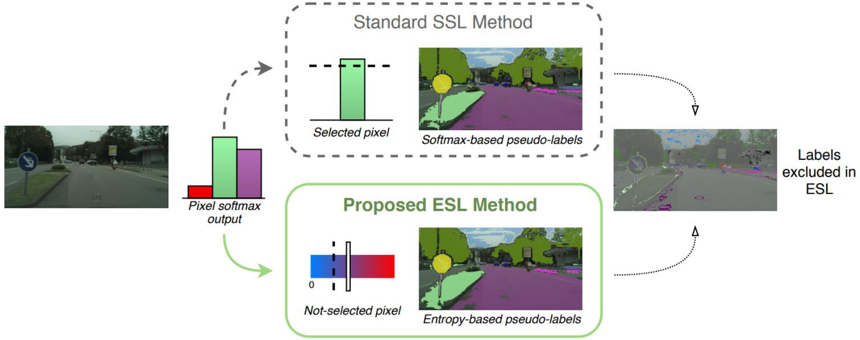
- Self-training with entropy minimization works
  - Similar finding in other works
  - Self-training with pseudo-labelling
    - High-scoring predictions
    - Training with noisy labels



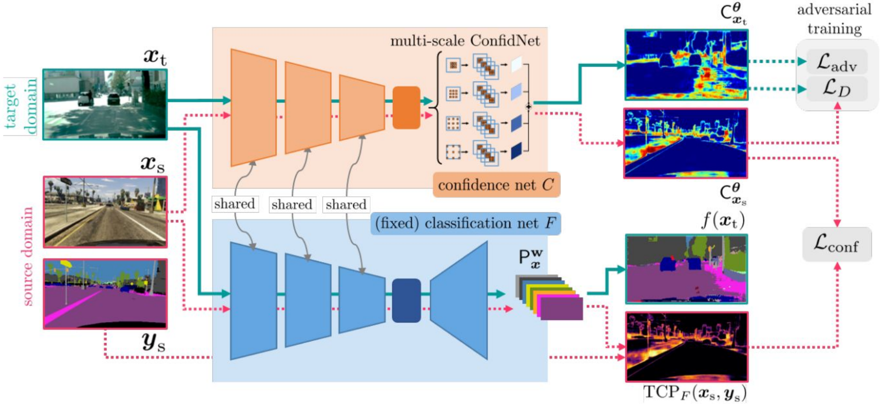
Unsupervised domain adaptation for semantic segmentation via class balanced self-training, Zou et al. ECCV’18
Bidirectional Learning for Domain Adaptation of Semantic Segmentation, Li et al., CVPR’19

Self-training for UDA

- ESL: Entropy-based criterion for pseudo-labelling
  - Low-entropy predictions as pseudo-labels
- ConDA: learnable confidence network for semantic failure detection
  - High confidence predictions as pseudo-labels



ESL: Entropy-guided Self-supervised Learning for Domain Adaptation in Semantic Segmentation, Saporta et al., CVPRW'2020



Confidence Estimation via Auxiliary Models, Corbiere et al., TPAMI'2021

Self-training for UDA

- ESL: Entropy-based criterion for pseudo-labelling
  - ► Low-entropy predictions as pseudo-labels

- ConDA: learnable confidence network for semantic failure detection
  - ► High confidence predictions as pseudo-labels

GTA5 ▷ Cityscapes

Method	Self-Train.	road	sidewalk	building	wall	fence	pole	light	sign	veg	terrain	sky	person	rider	car	truck	bus	train	mbike	bike	mIoU
AdaptSegNet [50]		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
CyCADA [49]		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
DISE [64]		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
AdvEnt [51]		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
CBST [54]	✓	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
MRKLD [55]	✓	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
BDL [21]	✓	91.0	44.7	84.2	34.6	**27.5**	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
ESL [53]	✓	90.2	43.9	84.7	35.9	28.5	31.2	37.9	34.0	84.5	42.2	83.9	59.0	32.2	81.8	36.7	49.4	1.8	30.6	34.1	48.6
ConDA	✓	**93.5**	**56.9**	**85.3**	**38.6**	26.1	34.3	36.9	29.9	**85.3**	40.6	**88.3**	58.1	30.3	**85.8**	**39.8**	**51.0**	0.0	28.9	37.8	**49.9**

Self-training for UDA

- Other self-training strategies:
  - Prototype-based pseudo-labelling: CAG_UDA [Zheng et al. NeurIPS'19], ProDA [Zhang et al. CVPR'21]
  - Inspired by the success of prototype-based approach to deal with noisy data [Han et al. ICCV'19]
  - Prototypes treat different classes equally regardless of their occurrence frequency

ADVENT [58]	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
BDL [35]	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
FADA [61]	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
CBST [75]	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
MRKLD [76]	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
CAG_UDA [69]	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
Seg-Uncertainty [73]	90.4	31.2	85.1	36.9	25.6	37.5	48.8	48.5	85.3	34.8	81.1	64.4	36.8	86.3	34.9	52.2	1.7	29.0	44.6	50.3
ProDA	87.8	**56.0**	79.7	**46.3**	**44.8**	**45.6**	**53.5**	**53.5**	**88.6**	**45.2**	82.1	**70.7**	**39.2**	**88.8**	**45.5**	**59.4**	1.0	**48.9**	**56.4**	**57.5**

What is missing?

SOTA methods still use “out-dated” network architectures and “low-res” input images

Transformer for UDA

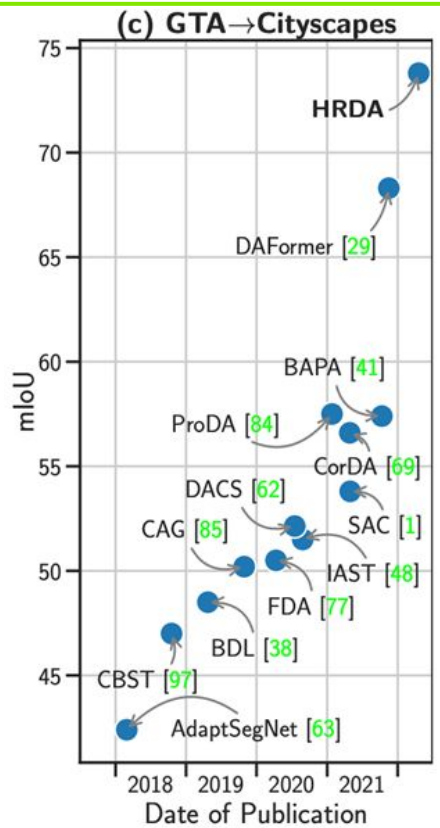
DAFormer: Improving Network Architectures and Training Strategies for Domain-Adaptive Semantic Segmentation", Hoyer, Dai, and Van Gool, CVPR 2022

HRDA: Context-Aware High-Resolution Domain-Adaptive Semantic Segmentation", Hoyer, Dai, and Van Gool, ECCV 2022

- Harness the robustness of SegFormer [Xie et al. NeurIPS 2021]

Method	Clean	Blur				Noise				Digital				Weather			
		Motion	Defoc	Glass	Gauss	Gauss	Impul	Shot	Speck	Bright	Contr	Satur	JPEG	Snow	Spatt	Fog	Frost
DLv3+ (MBv2)	72.0	53.5	49.0	45.3	49.1	6.4	7.0	6.6	16.6	51.7	46.7	32.4	27.2	13.7	38.9	47.4	17.3
DLv3+ (R50)	76.6	58.5	56.6	47.2	57.7	6.5	7.2	10.0	31.1	58.2	54.7	41.3	27.4	12.0	42.0	55.9	22.8
DLv3+ (R101)	77.1	59.1	56.3	47.7	57.3	13.2	13.9	16.3	36.9	59.2	54.5	41.5	37.4	11.9	47.8	55.1	22.7
DLv3+ (X41)	77.8	61.6	54.9	51.0	54.7	17.0	17.3	21.6	43.7	63.6	56.9	51.7	38.5	18.2	46.6	57.6	20.6
DLv3+ (X65)	78.4	63.9	59.1	52.8	59.2	15.0	10.6	19.8	42.4	65.9	59.1	46.1	31.4	19.3	50.7	63.6	23.8
DLv3+ (X71)	78.6	64.1	60.9	52.0	60.4	14.9	10.8	19.4	41.2	68.0	58.7	47.1	40.2	18.8	50.4	64.1	20.2
ICNet	65.9	45.8	44.6	47.4	44.7	8.4	8.4	10.6	27.9	41.0	33.1	27.5	34.0	6.3	30.5	27.3	11.0
FCN8s	66.7	42.7	31.1	37.0	34.1	6.7	5.7	7.8	24.9	53.3	39.0	36.0	21.2	11.3	31.6	37.6	19.7
DilatedNet	68.6	44.4	36.3	32.5	38.4	15.6	14.0	18.4	32.7	52.7	32.6	38.1	29.1	12.5	32.3	34.7	19.2
ResNet-38	77.5	54.6	45.1	43.3	47.2	13.7	16.0	18.2	38.3	60.0	50.6	46.9	14.7	13.5	45.9	52.9	22.2
PSPNet	78.8	59.8	53.2	44.4	53.9	11.0	15.4	15.4	34.2	60.4	51.8	30.6	21.4	8.4	42.7	34.4	16.2
GSCNN	80.9	58.9	58.4	41.9	60.1	5.5	2.6	6.8	24.7	75.9	61.9	70.7	12.0	12.4	47.3	67.9	32.6
SegFormer-B5	**82.4**	**69.1**	**68.6**	**64.1**	**69.8**	**57.8**	**63.4**	**52.3**	**72.8**	**81.0**	**77.7**	**80.1**	**58.8**	**40.7**	**68.4**	**78.5**	**49.9**

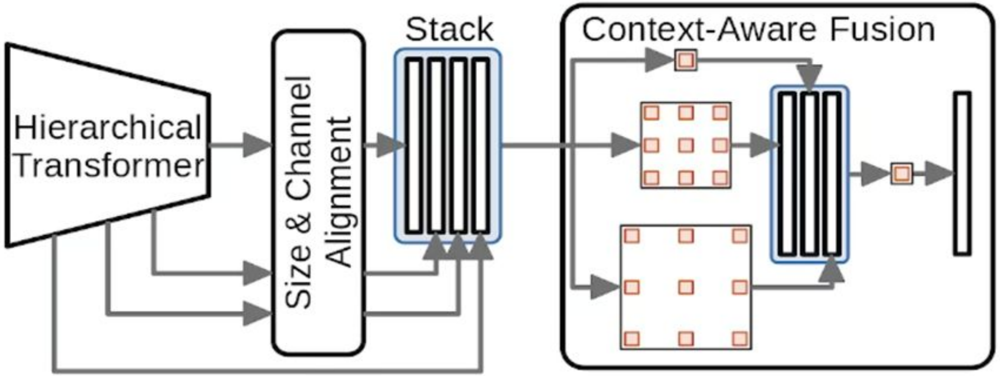
- Enable learning high-reso details and low-reso context at the same time



Transformer for UDA - DAFormer

Design of an architecture tailored for UDA

- Hierarichal Transformer encoder [4]
- Context-aware multi-level feature fusion decoder

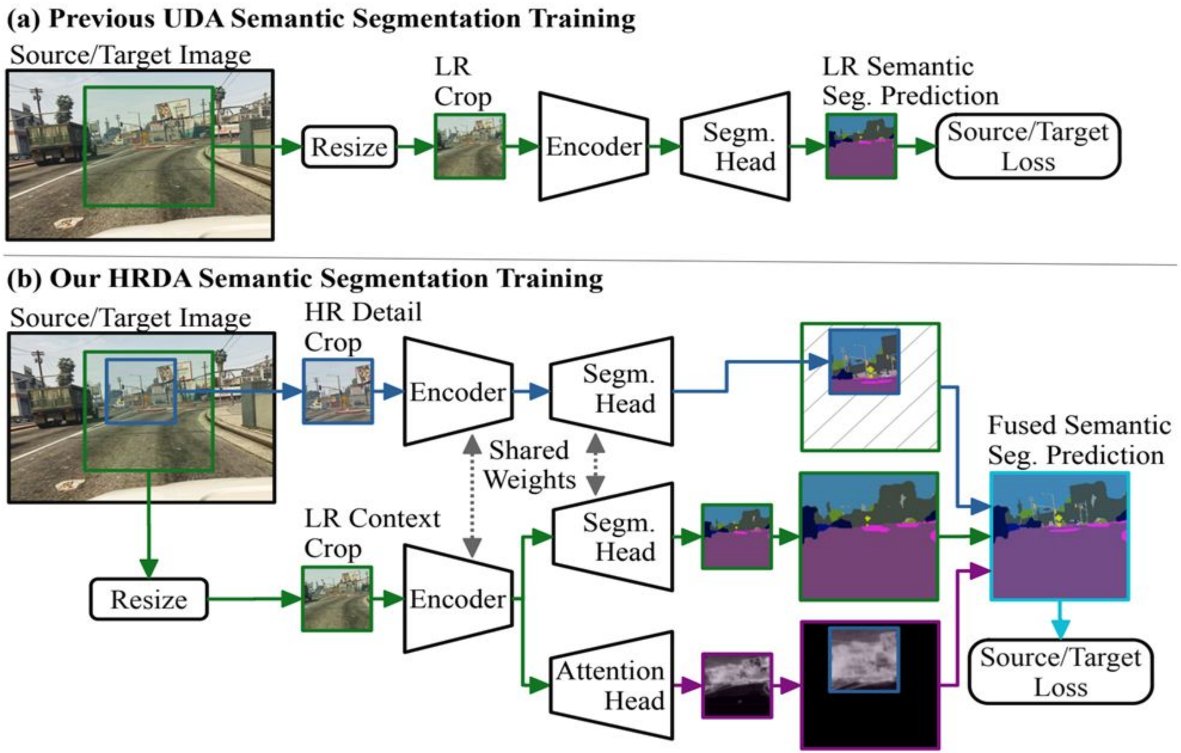


Network Architecture	UDA	Oracle	UDA / Oracle
DeepLabV2	56.0	72.1	77.7%
DAFormer	68.3	77.6	$\boxed{88.0\%}$

Reduced performance gap between UDA and supervised oracle

DAFormer: Improving Network Architectures and Training Strategies for Domain-Adaptive Semantic Segmentation", Hoyer, Dai, and Van Gool, CVPR 2022

Transformer for UDA - HRDA



HRDA: Context-Aware High-Resolution Domain-Adaptive Semantic Segmentation", Hoyer, Dai, and Van Gool, ECCV 2022

SoTA in 2023

Method	Road	S.walk	Build.	Wall	Fence	Pole	Tr.Light	Sign	Veget.	Terrain	Sky	Person	Rider	Car	Truck	Bus	Train	M.bike	Bike	mIoU
Synthetic-to-Real: GTA→Cityscapes (Val.)																				
ADVENT [76]	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
DACS [72]	89.9	39.7	87.9	30.7	39.5	38.5	46.4	52.8	88.0	44.0	88.8	67.2	35.8	84.5	45.7	50.2	0.0	27.3	34.0	52.1
ProDA [89]	87.8	56.0	79.7	46.3	44.8	45.6	53.5	53.5	88.6	45.2	82.1	70.7	39.2	88.8	45.5	59.4	1.0	48.9	56.4	57.5
DAFormer [30]	95.7	70.2	89.4	53.5	48.1	49.6	55.8	59.4	89.9	47.9	92.5	72.2	44.7	92.3	74.5	78.2	65.1	55.9	61.8	68.3
HRDA [31]	96.4	74.4	91.0	**61.6**	51.5	57.1	63.9	69.3	91.3	48.4	94.2	79.0	52.9	93.9	84.1	85.7	75.9	63.9	67.5	73.8
MIC (HRDA)	**97.4**	**80.1**	**91.7**	61.2	**56.9**	**59.7**	**66.0**	**71.3**	**91.7**	**51.4**	**94.3**	**79.8**	**56.1**	**94.6**	**85.4**	**90.3**	**80.4**	**64.5**	**68.5**	**75.9**
Synthetic-to-Real: Synthia→Cityscapes (Val.)																				
ADVENT [76]	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
DACS [72]	80.6	25.1	81.9	21.5	2.9	37.2	22.7	24.0	83.7	–	90.8	67.6	38.3	82.9	–	38.9	–	28.5	47.6	48.3
ProDA [89]	**87.8**	45.7	84.6	37.1	0.6	44.0	54.6	37.0	**88.1**	–	84.4	74.2	24.3	88.2	–	51.1	–	40.5	45.6	55.5
DAFormer [30]	84.5	40.7	88.4	41.5	6.5	50.0	55.0	54.6	86.0	–	89.8	73.2	48.2	87.2	–	53.2	–	53.9	61.7	60.9
HRDA [31]	85.2	47.7	88.8	**49.5**	4.8	57.2	65.7	60.9	85.3	–	92.9	79.4	52.8	89.0	–	**64.7**	–	63.9	**64.9**	65.8
MIC (HRDA)	86.6	**50.5**	**89.3**	47.9	**7.8**	**59.4**	**66.7**	**63.4**	87.1	–	**94.6**	**81.0**	**58.9**	**90.1**	–	61.9	–	**67.1**	64.3	**67.3**

MIC: Masked Image Consistency for Context-Enhanced Domain Adaptation, Hoyer CVPR'23

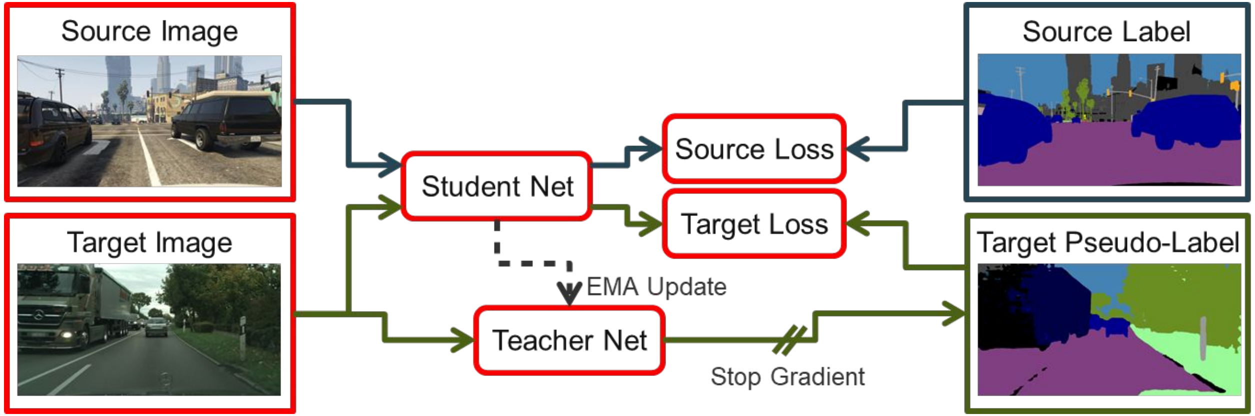
SoTA in 2023

Day-to-Nighttime: Cityscapes→DarkZurich (Test)																				
ADVENT [87]	85.8	37.9	55.5	27.7	14.5	23.1	14.0	21.1	32.1	8.7	2.0	39.9	16.6	64.0	13.8	0.0	58.8	28.5	20.7	29.7
MGCDA† [76]	80.3	49.3	66.2	7.8	11.0	41.4	38.9	39.0	64.1	18.0	55.8	52.1	53.5	74.7	66.0	0.0	37.5	29.1	22.7	42.5
DANNet† [92]	90.0	54.0	74.8	41.0	21.1	25.0	26.8	30.2	**72.0**	26.2	**84.0**	47.0	33.9	68.2	19.0	0.3	66.4	38.3	23.6	44.3
DAFormer [32]	93.5	65.5	73.3	39.4	19.2	53.3	44.1	44.0	59.5	34.5	66.6	53.4	52.7	82.1	52.7	9.5	89.3	50.5	38.5	53.8
HRDA [33]	90.4	56.3	72.0	39.5	19.5	57.8	**52.7**	43.1	59.3	29.1	70.5	60.0	58.6	**84.0**	**75.5**	11.2	90.5	51.6	40.9	55.9
MIC (HRDA)	**94.8**	**75.0**	**84.0**	**55.1**	**28.4**	**62.0**	35.5	**52.6**	59.2	**46.8**	70.0	**65.2**	**61.7**	82.1	64.2	**18.5**	**91.3**	**52.6**	**44.0**	**60.2**
Clear-to-Adverse-Weather: Cityscapes→ACDC (Test)																				
ADVENT [87]	72.9	14.3	40.5	16.6	21.2	9.3	17.4	21.2	63.8	23.8	18.3	32.6	19.5	69.5	36.2	34.5	46.2	26.9	36.1	32.7
MGCDA† [76]	73.4	28.7	69.9	19.3	26.3	36.8	53.0	53.3	75.4	32.0	84.6	51.0	26.1	77.6	43.2	45.9	53.9	32.7	41.5	48.7
DANNet† [92]	84.3	54.2	77.6	38.0	30.0	18.9	41.6	35.2	71.3	39.4	86.6	48.7	29.2	76.2	41.6	43.0	58.6	32.6	43.9	50.0
DAFormer [32]	58.4	51.3	84.0	42.7	35.1	50.7	30.0	57.0	74.8	52.8	51.3	58.3	32.6	82.7	58.3	54.9	82.4	44.1	50.7	55.4
HRDA [33]	88.3	57.9	88.1	**55.2**	36.7	56.3	**62.9**	65.3	74.2	57.7	85.9	68.8	45.7	88.5	**76.4**	82.4	87.7	52.7	60.4	68.0
MIC (HRDA)	**90.8**	**67.1**	**89.2**	54.5	**40.5**	**57.2**	62.0	**68.4**	**76.3**	**61.8**	**87.0**	**71.3**	**49.4**	**89.7**	75.7	**86.8**	**89.1**	**56.9**	**63.0**	**70.4**

MIC: Masked Image Consistency for Context-Enhanced Domain Adaptation, Hoyer CVPR’23

What did we learn?

- Self-training

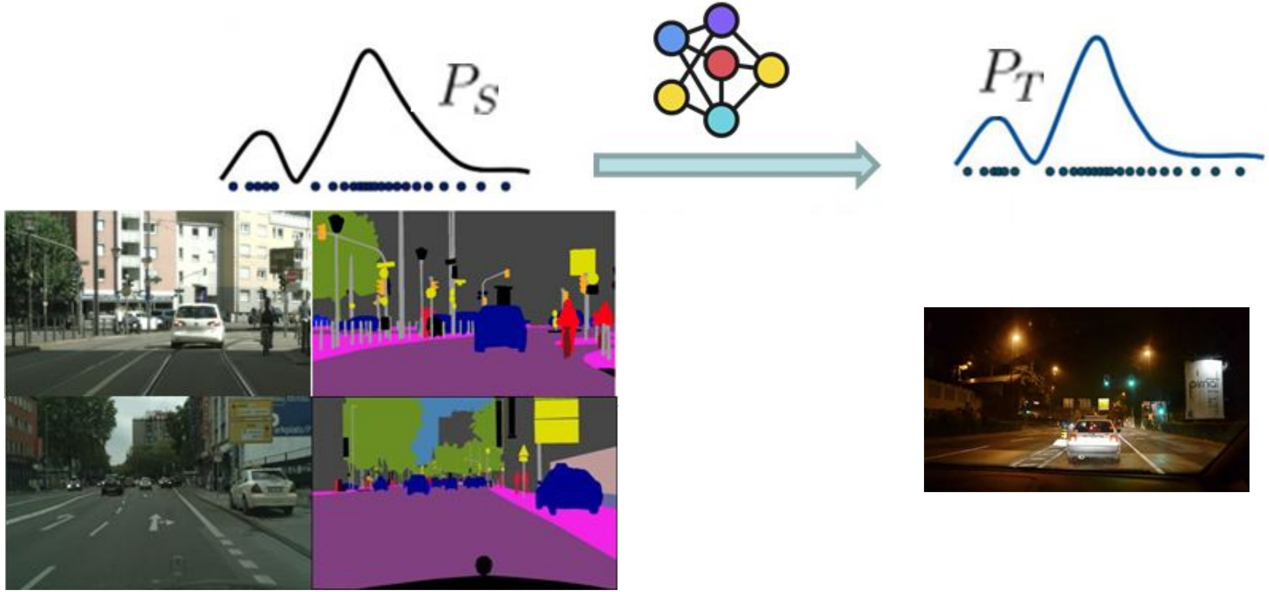


- Robust encoder architecture, e.g. SegFormer

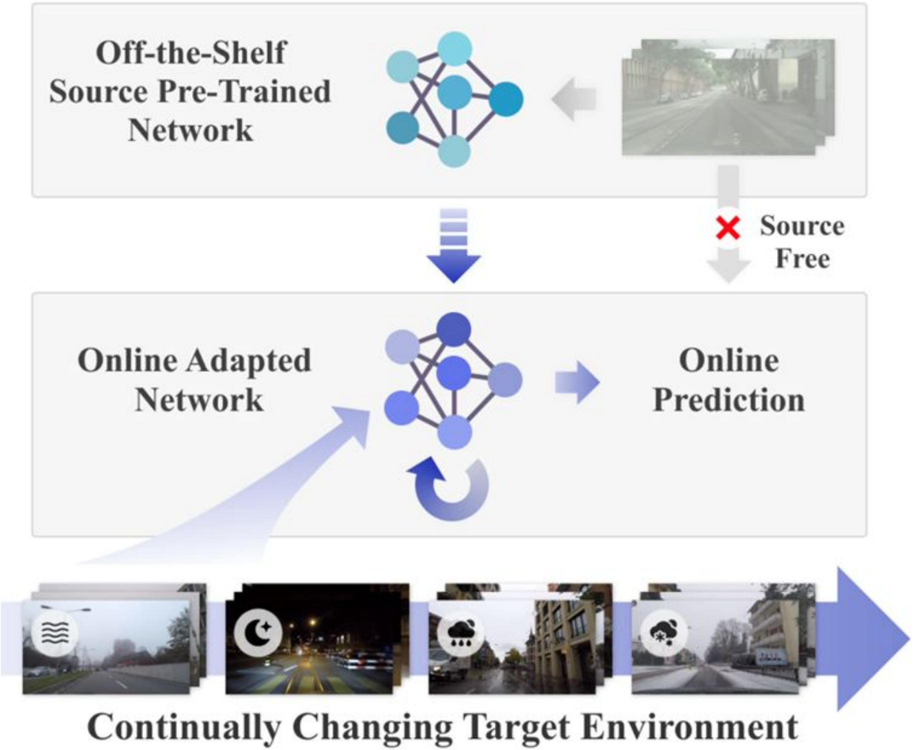
- High-resolution recognition, e.g. HRDA

What can we do to generalize?

1. Unsupervised Domain Adaptation: Learning Target Distribution with Unlabeled Samples
2. Test-time Adaptation: Learning Target Distribution at Test Time from a Single Sample

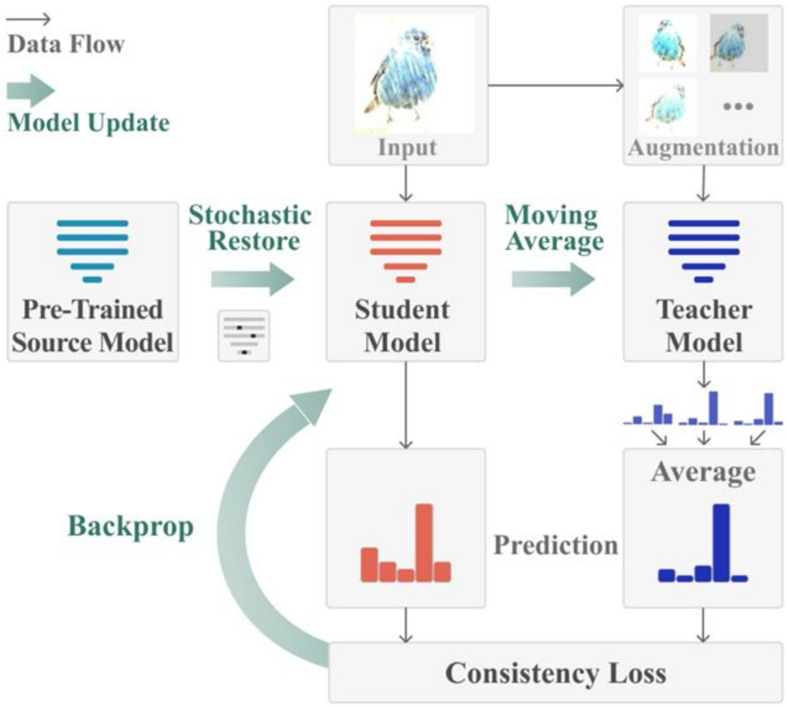


Continual Test-Time Domain Adaptation



Continual Test-Time Domain Adaptation, Wang, Fink, Van Gool, Dai, CVPR, 2022.

Continual Test-Time Domain Adaptation



- Self-training with (better) predictions by a **teacher network**
- Self-training with (better) **Augmentation-Averaged Pseudo-Labels**
- **Stochastic Weights Restoration** to avoid catastic forgetting

Continual Test-Time Domain Adaptation, Wang, Fink, Van Gool, Dai, CVPR, 2022.

Continual Test-Time Domain Adaptation

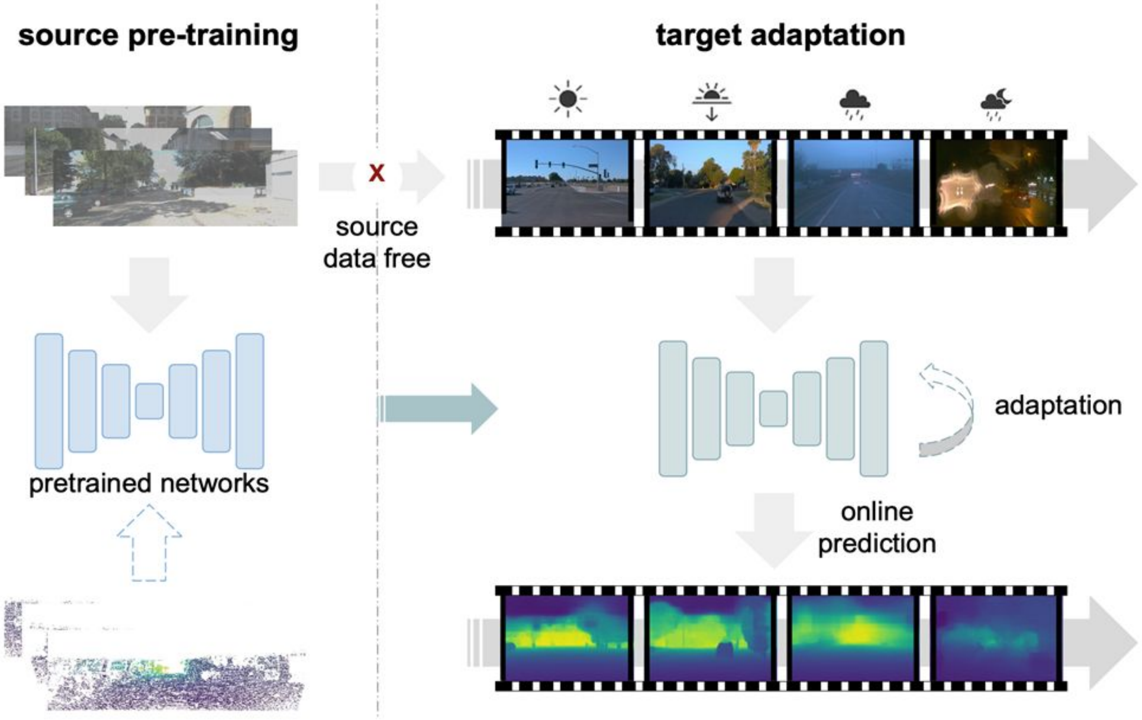
Table 2. Classification error rate (%) for the standard CIFAR10-to-CIFAR10C online continual test-time adaptation task. Tesults ar evaluated on WideResNet-28 with the largest corruption severity level 5. * denotes the requirement on additional domain information.

$t \longrightarrow$

Method	Weight-avg.	Aug-avg.	Stochastic Restore	Gaussian	shot	impulse	defocus	glass	motion	zoom	snow	frost	fog	brightness	contrast	elastic_trans	pixelate	jpeg	Mean
Source				72.3	65.7	72.9	46.9	54.3	34.8	42.0	25.1	41.3	26.0	9.3	46.7	26.6	58.5	30.3	43.5
BN Stats Adapt				28.1	26.1	36.3	12.8	35.3	14.2	12.1	17.3	17.4	15.3	8.4	12.6	23.8	19.7	27.3	20.4
Pseudo-label				26.7	22.1	32.0	13.8	32.2	15.3	12.7	17.3	17.3	16.5	10.1	13.4	22.4	18.9	25.9	19.8
TENT-online* [61]				24.8	23.5	33.0	12.0	31.8	13.7	10.8	15.9	16.2	13.7	7.9	12.1	22.0	17.3	24.2	18.6
TENT-continual [61]				24.8	**20.6**	28.6	14.4	31.1	16.5	14.1	19.1	18.6	18.6	12.2	20.3	25.7	20.8	24.9	20.7
CoTTA (Ours)	✓			27.2	22.8	30.8	12.1	30.1	13.9	11.9	17.2	16.0	14.3	9.4	13.1	19.9	15.4	19.9	18.3
CoTTA (Ours)	✓	✓		24.5	21.0	**26.0**	12.3	27.9	13.9	12.0	16.6	15.9	14.7	9.4	13.6	19.8	14.7	18.7	17.4
CoTTA (Ours)	✓	✓	✓	**24.3**	21.3	26.6	**11.6**	**27.6**	**12.2**	**10.3**	**14.8**	**14.1**	**12.4**	**7.5**	**10.6**	**18.3**	**13.4**	**17.3**	**16.2** (0.1)

Continual Test-Time Domain Adaptation, Wang, Fink, Van Gool, Dai, CVPR, 2022.

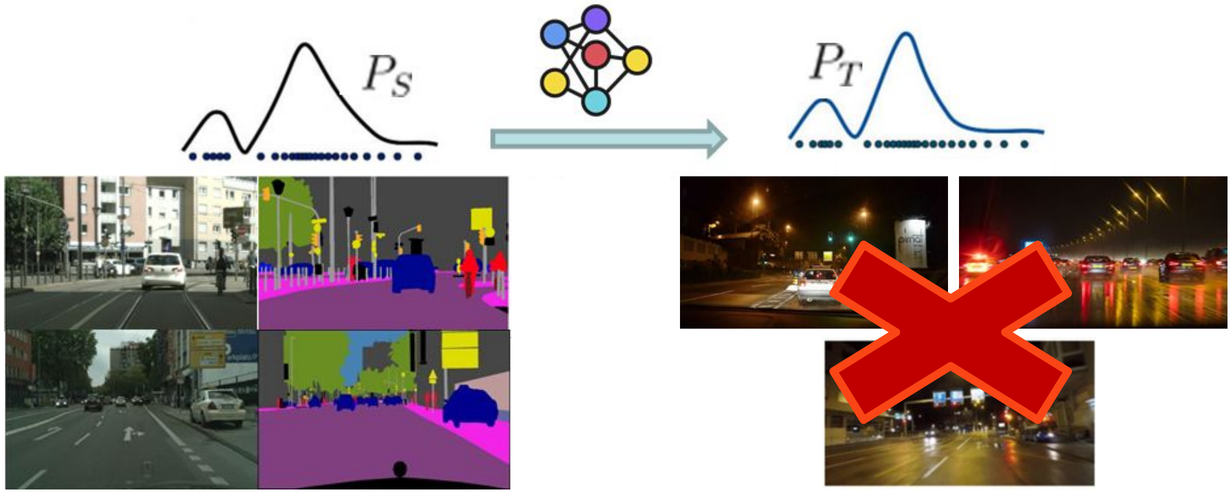
Continual Test-Time Domain Adaptation



Continual Test-Time Domain Adaptation for Monocular Depth Estimation, Li, Shi, Bernt, Dai, ICRA 2023

What can we do to generalize?

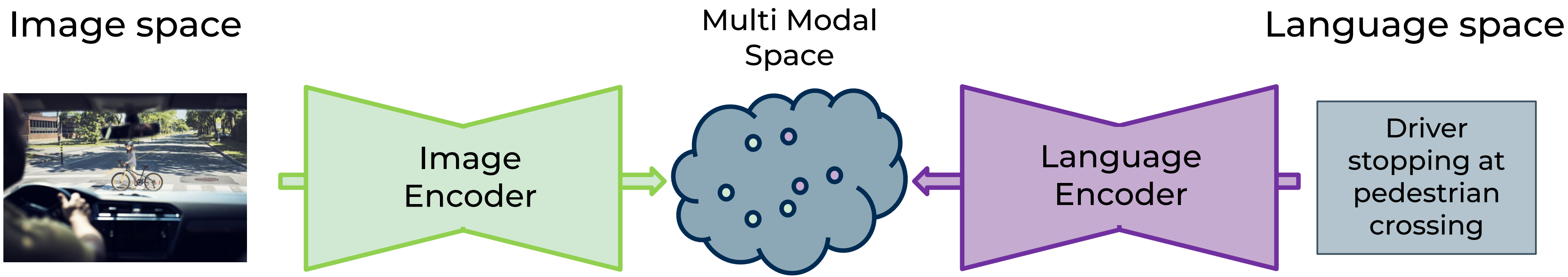
1. Unsupervised Domain Adaptation: Learning Target Distribution with Unlabeled Samples
2. Test-time Adaptation: Learning Target Distribution at Test Time from a Single Sample
3. Zero-shot Adaptation: Learning Target Distribution with Text Prompt



2022 - Foundation Models

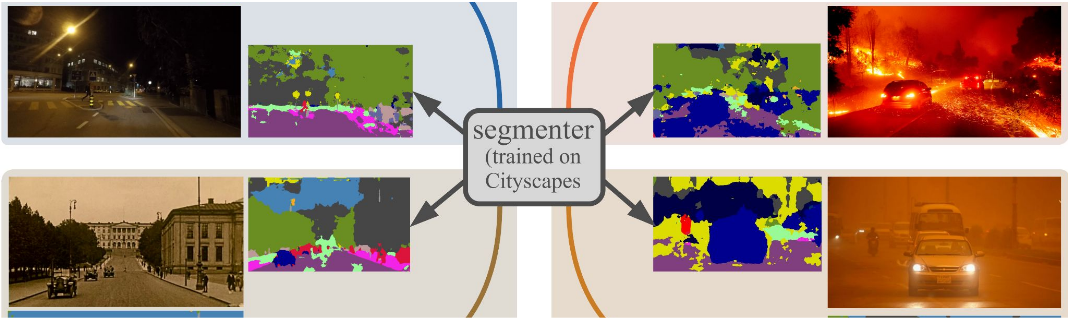
Multimodal Foundation Models

- Vision-Language Models - VLM: **CLIP / BLIP / ALIGN**



Prompt-driven Zero-shot Domain Adaptation

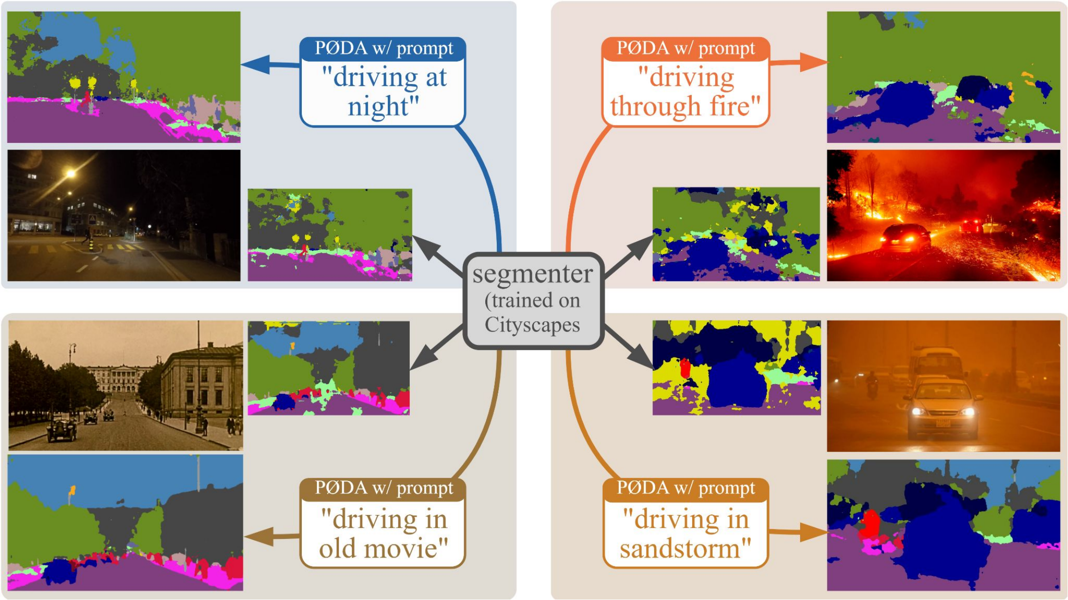
Harness foundation models for DA?



PØDA: Prompt-driven Zero-shot Domain Adaptation, Fahes et al. ICCV'23

Prompt-driven Zero-shot Domain Adaptation

Harness foundation models for DA?

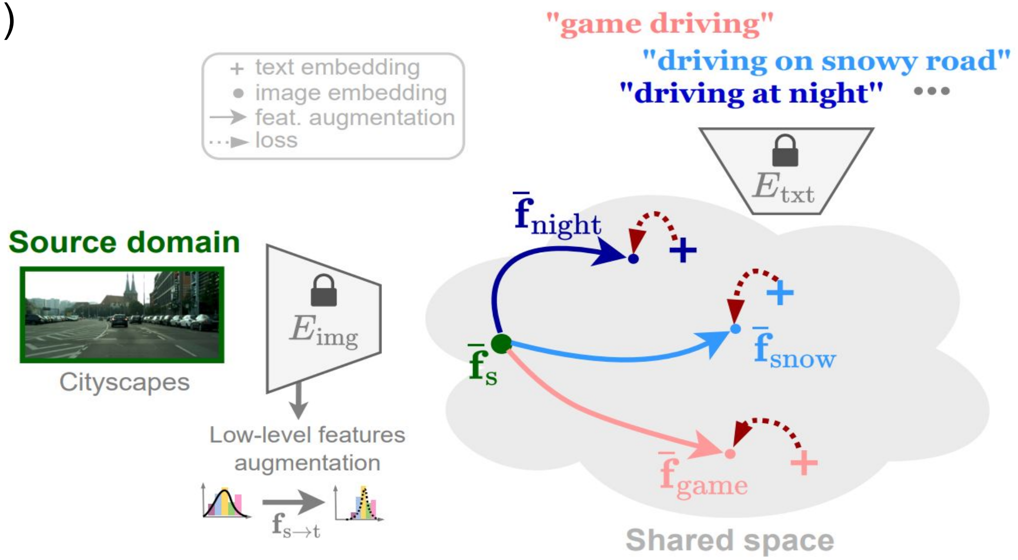


PØDA: Prompt-driven Zero-shot Domain Adaptation, Fahes et al. ICCV'23

Prompt-driven Zero-shot Domain Adaptation

Prompt-driven Instance Normalization (PIN)

- Stylize features using prompts
- Preserve semantics

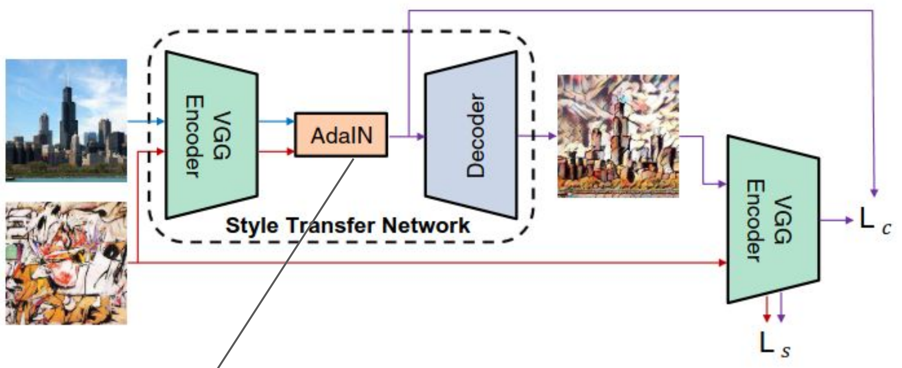


PØDA: Prompt-driven Zero-shot Domain Adaptation, Fahes et al. ICCV'23

Prompt-driven Zero-shot Domain Adaptation

Prompt-driven Instance Normalization (PIN)

- Stylize features using prompts
- Preserve semantics



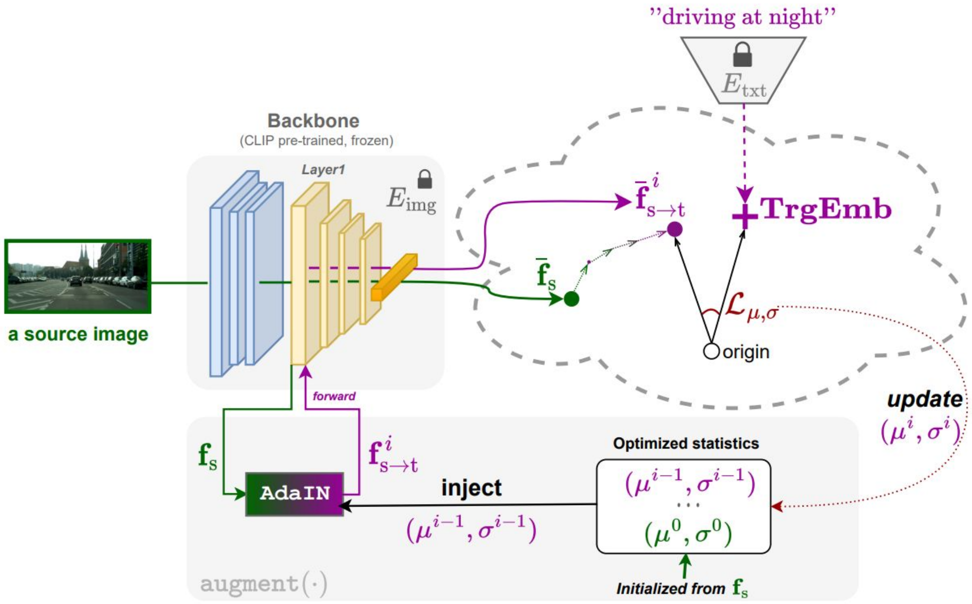
$$\mathrm{AdaIN}(x, y) = \sigma(y)\frac{x-\mu(x)}{\sigma(x)} + \mu(y)$$

[21] Huang, X. and Belongie, S., Arbitrary Style Transfer in Real-time with Adaptive Instance Normalization. ICCV 2017

Prompt-driven Zero-shot Domain Adaptation

Prompt-driven Instance Normalization (PIN)

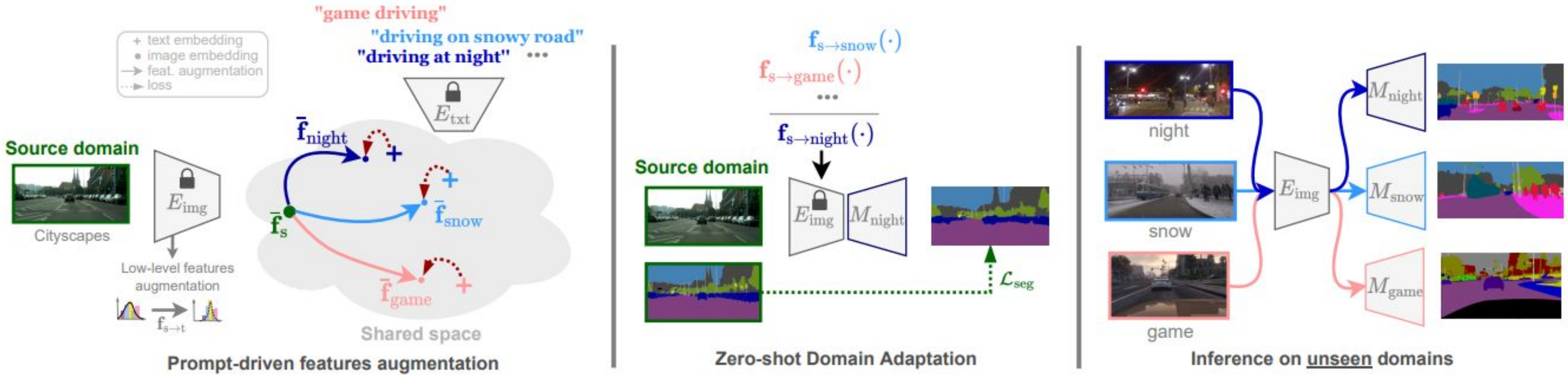
- Stylize features using prompts
- Preserve semantics



$$\mathcal{L}_{\mu,\sigma}(\bar{\mathbf{f}}_{s\to t}, \mathbf{TrgEmb}) = 1 - \frac{\bar{\mathbf{f}}_{s\to t} \cdot \mathbf{TrgEmb}}{\|\bar{\mathbf{f}}_{s\to t}\| \, \|\mathbf{TrgEmb}\|}$$

PØDA: Prompt-driven Zero-shot Domain Adaptation, Fahes et al. ICCV'23

Prompt-driven Zero-shot Domain Adaptation



PØDA: Prompt-driven Zero-shot Domain Adaptation, Fahes et al. ICCV'23

Prompt-driven Zero-shot Domain Adaptation

Source	Target eval.	Method	mIoU[%]
	TrgPrompt = "driving at night"		
CS	ACDC Night	source-only	18.31
		CLIPstyler	21.38 ±0.36
		PØDA	**25.03** ±0.48
	TrgPrompt = "driving in snow"		
	ACDC Snow	source-only	39.28
		CLIPstyler	41.09 ±0.17
		PØDA	**43.90** ±0.53
	TrgPrompt = "driving under rain"		
	ACDC Rain	source-only	38.20
		CLIPstyler	37.17 ±0.10
		PØDA	**42.31** ±0.55
	TrgPrompt = "driving in a game"		
	GTA5	source-only	39.59
		CLIPstyler	38.73 ±0.16
		PØDA	**41.07** ±0.48
	TrgPrompt = "driving"		
GTA5	CS	source-only	36.38
		CLIPstyler	31.50 ±0.21
		PØDA	**40.08** ±0.52

Figure 5. **CLIPstyler [21] stylization.** A sample Cityscapes image stylized using adhoc target prompts. Translated images exhibit visible artifacts, potentially harming adaptation *e.g.* rain in Tab. 1

PØDA: Prompt-driven Zero-shot Domain Adaptation, Fahes et al. ICCV'23

Prompt-driven Zero-shot Domain Adaptation

Method	Prior	ACDC Night
CIConv* [26]	physics	30.60 / 34.50 (Δ=3.90)
SM-PPM [56]	1 target image	13.07 / 14.60 (Δ=1.53)
CLIPstyler [25]	1 prompt	18.31 / 21.38 (Δ=3.07)
PØDA	1 prompt	18.31 / 25.03 (Δ=6.72)

* Results of CIConv are on DarkZurich, a subset of ACDC Night [45].

Table 8. **Effect of different priors for zero-shot/one-shot adaptation.** We report mIoU% for source-only / adapted models, and gain brought by adaptation (Δ in mIoU). Note that [26, 56] use a deeper backbone making results not directly comparable.

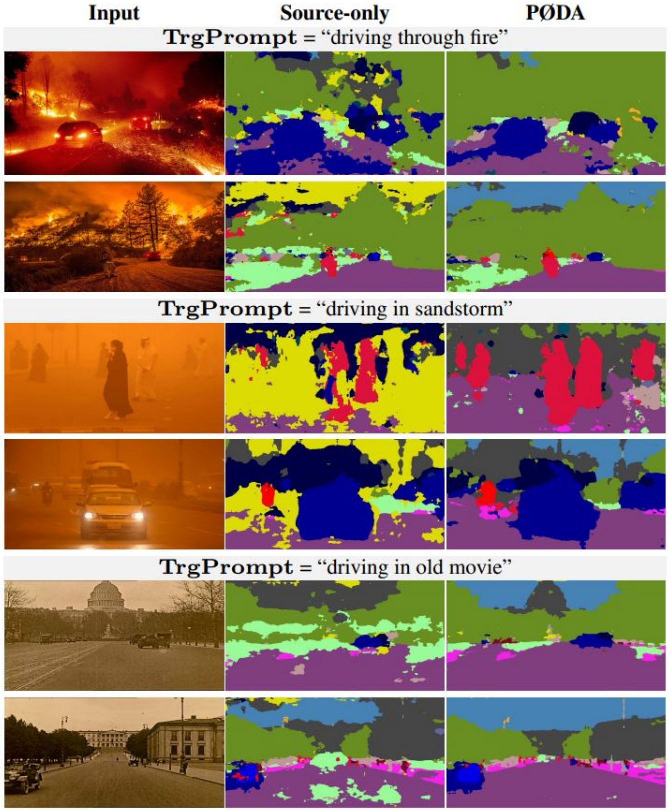
Method	ACDC Night	ACDC Snow	ACDC Rain	GTA5
Source only	18.31	39.28	38.20	39.59
Trg	"driving at night"	"driving in snow"	"driving under rain"	"driving in a game"
	25.03 ±0.48	43.90 ±0.53	42.31 ±0.55	41.07 ±0.48
ChatGPT-generated, Relevant →	"operating a vehicle after sunset"	"operating a vehicle in snowy conditions"	"operating a vehicle in wet conditions"	"piloting a vehicle in a virtual world"
	24.38 ±0.37	44.33 ±0.36	42.21 ±0.47	41.25 ±0.40
	"driving during the nighttime hours"	"driving on snow-covered roads"	"driving on rain-soaked roads"	"controlling a car in a digital simulation"
	25.22 ±0.64	43.56 ±0.62	**42.51** ±0.33	41.19 ±0.14
	"navigating the roads in darkness"	"piloting a vehicle in snowy terrain"	"navigating through rainfall while driving"	"maneuvering a vehicle in a computerized racing experience"'
	24.73 ±0.47	**44.67** ±0.18	41.11 ±0.69	40.34 ±0.49
	"driving in low-light conditions"	"driving in wintry precipitation"	"driving in inclement weather"	"operating a transport in a video game environment"
	24.68 ±0.34	43.11 ±0.56	40.68 ±0.37	41.34 ±0.42
	"travelling by car after dusk"	"travelling by car in a snowstorm"	"travelling by car during a downpour"	"navigating a machine through a digital driving simulation"
	24.89 ±0.24	43.83 ±0.17	42.05 ±0.35	**41.86** ±0.10
	24.82	*43.90*	*41.81*	*41.18*
ChatGPT-generated, ← Irrelevant	"mesmerizing northern lights display"			
	20.05 ±0.77	40.07 ±0.66	38.43 ±0.82	37.98 ±0.31
	"playful dolphins in the ocean"			
	20.11 ±0.31	39.87 ±0.26	38.56 ±0.58	37.05 ±0.31
	"breathtaking view from mountaintop"			
	20.65 ±0.33	42.08 ±0.28	40.05 ±0.52	40.09 ±0.23
	"cheerful sunflower field in bloom"			
	21.10 ±0.50	39.85 ±0.68	40.09 ±0.41	37.93 ±0.55
	"dramatic cliff overlooking the ocean"			
	20.09 ±0.98	38.20 ±0.54	38.48 ±0.37	37.57 ±0.46
	"majestic eagle in flight over mountains"			
	20.70 ±0.38	39.60 ±0.27	40.38 ±0.86	38.52 ±0.21
	20.45	*39.95*	*39.33*	*38.19*

Prompt-driven Zero-shot Domain Adaptation

Method	Target	CS→Foggy	Night Clear	Dusk Rainy	Night Rainy	Day Foggy
Backbone		ResNet-50		ResNet-101		
DA-Faster [6]	✓	32.0	-	-	-	-
ViSGA [38]	✓	43.3	-	-	-	-
NP+ [12]	✗	46.3	-	-	-	-
S-DGOD [48]	✗	-	36.6	28.2	16.6	33.5
CLIP The Gap [44]	✗	-	36.9	32.3	18.7	38.5
PØDA	✗	**47.3**	**40.3**	**37.4**	**19.0**	**41.7**

Table 7. **PØDA for object detection (mAP%)**..

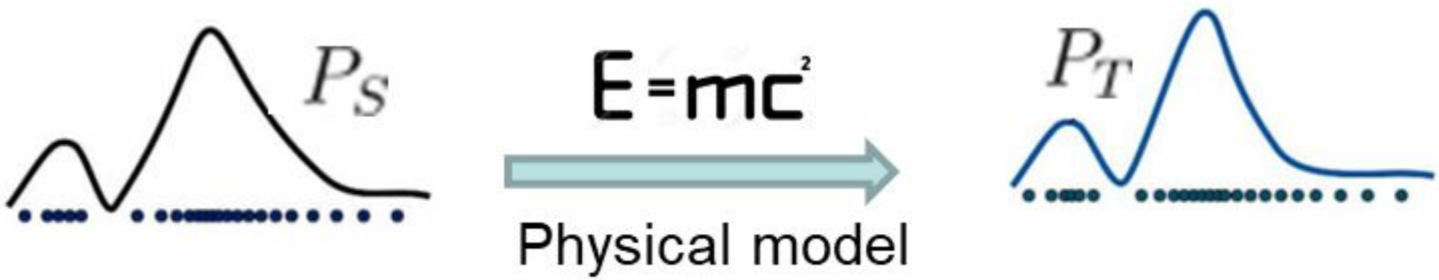
Prompt-driven Zero-shot Domain Adaptation



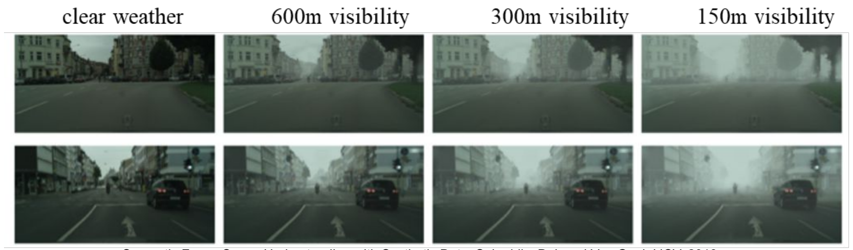
Poster on Friday

What else can we do to advance?

1. Unsupervised Domain Adaptation: Learning Target Distribution with Unlabeled Samples
2. Test-time Adaptation: Learning Target Distribution at Test Time from a Single Sample
3. Zero-shot Adaptation: Learning Target Distribution with Text Prompt
4. Data Synthesis: Simulate Target Distribution via Physics-Based Model



Data synthesis



Semantic Foggy Scene Understanding with Synthetic Data, Sakaridis, Dai, and Van Gool, IJCV, 2018


Flare7K: A Phenomenological Nighttime Flare Removal Dataset. Dai, Li, Zhou, Feng, and Loy, NeurIPS, 2022

Physics-Based Rendering for Improving Robustness to Rain. Halder, Lalonde, and Charette, ICCV 2019

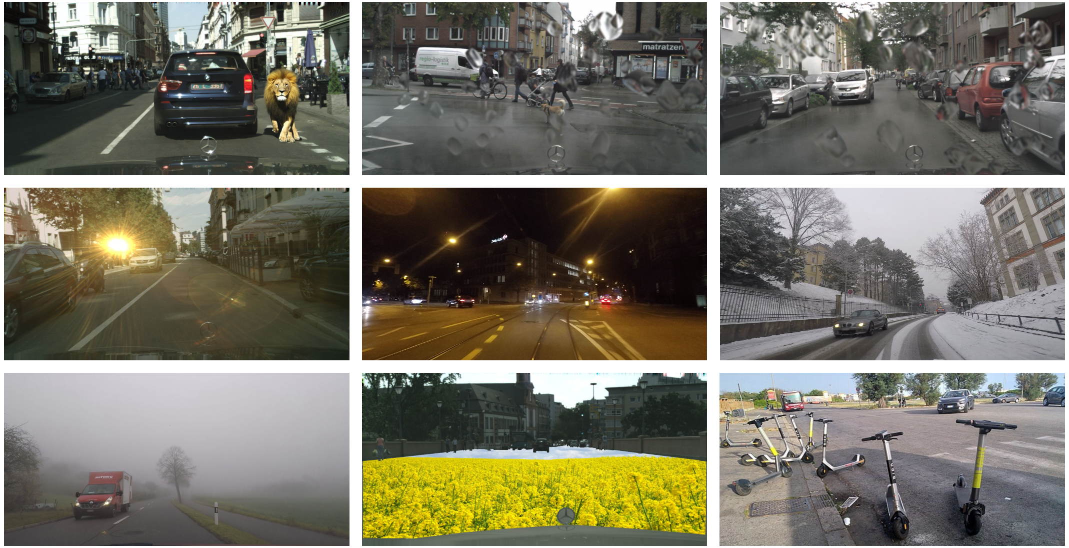
What else can we do to advance?

1. Unsupervised Domain Adaptation: Learning Target Distribution with Unlabeled Samples
2. Test-time Adaptation: Learning Target Distribution at Test Time from a Single Sample
3. Zero-shot Adaptation: Learning Target Distribution with Text Prompt
4. Data Synthesis: Simulate Target Distribution via Physics-Based Model
5. Robustness Benchmark

BRAVO Challenge

A unified robustness benchmark for vision perception in autonomous driving

- Semantic segmentation
- Two tracks: single- and multi-domain training
- 3,901 images
- 7 metrics for a comprehensive assessment
- 6 assessment modalities on the test datasets




Thank you!