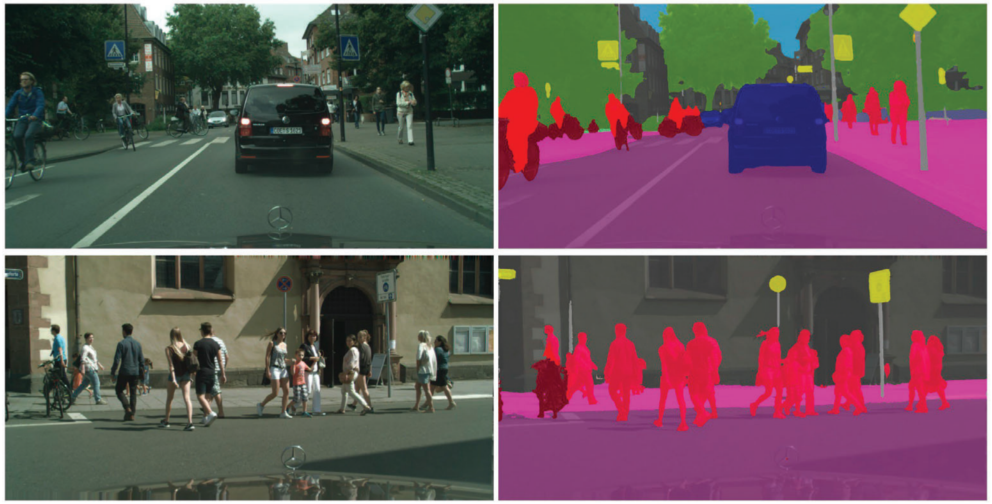
ACM Multimedia 2022
Lisbon, Portugal | 10-14 October

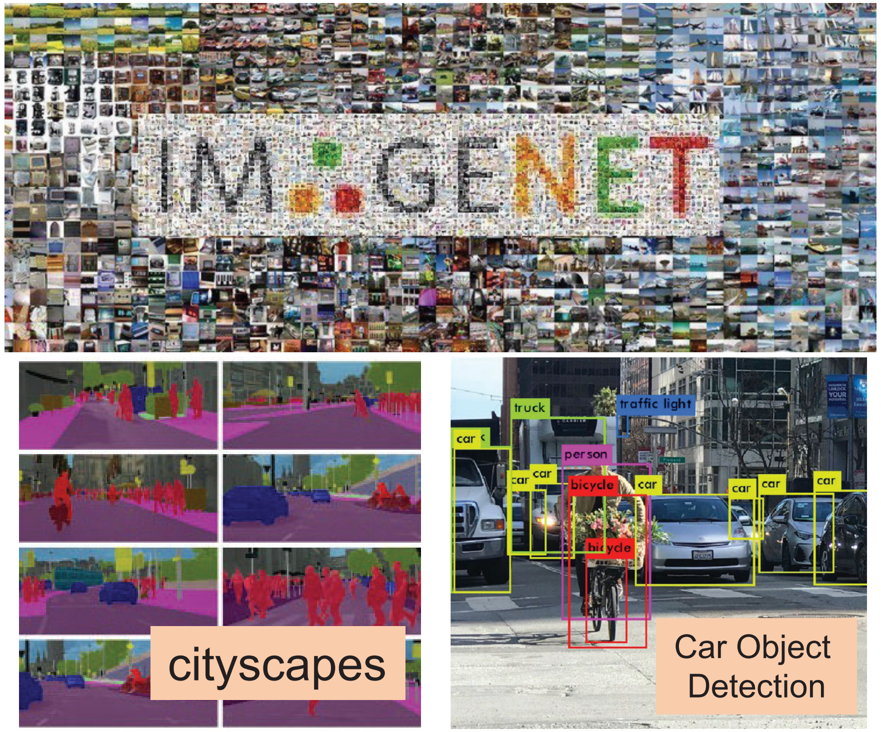
Understanding, Detection, and Retrieval in Harsh Environments

Zheng Wang
Wuhan University

Excellent Environments & Sufficient Training Samples



Excellent Environments



Sufficient Training Samples

Harsh environments

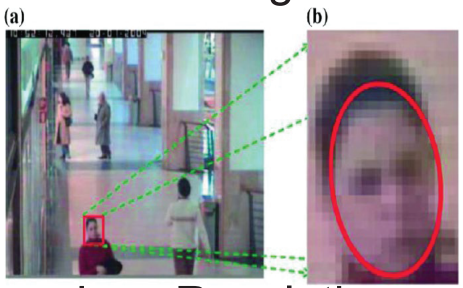

Fog


Rain


Snow


Low Light


Blur



Low Resolution

Harsh environments

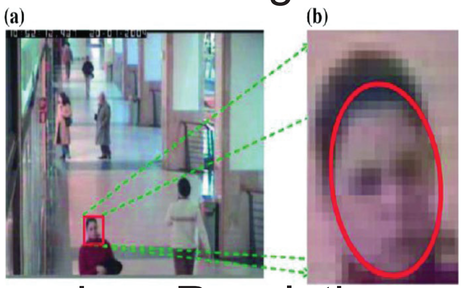
Fog

Rain

Snow

Low Light

Blur



Low Resolution

Enhancement → Understanding

super-resolution
defog, derain, deblur
inpainting
low-light enhancement

Domain Adaptation → Understanding

disentangle
pseudo label

Generation → Understanding

synthesis data

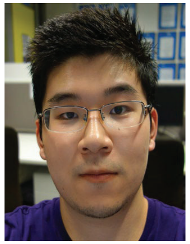
Presentations

Dr. Kui Jiang

Image enhancement: Disentanglement

Dr. Dan Xu

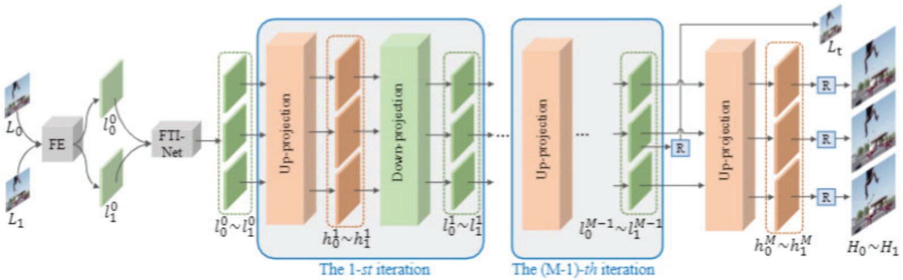
2D and 3D Scene Understanding



Dr. Zhedong Zheng

Domain Adaptation: Consistency and Uncertainty

Our Related Works



Video Super-Resolution [1-3]
[1] October 13: Oral Session 7c
[2] October 12: Poster Session 2

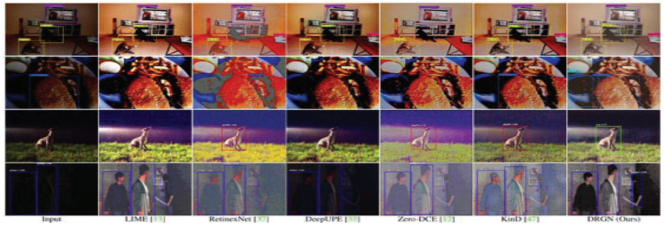
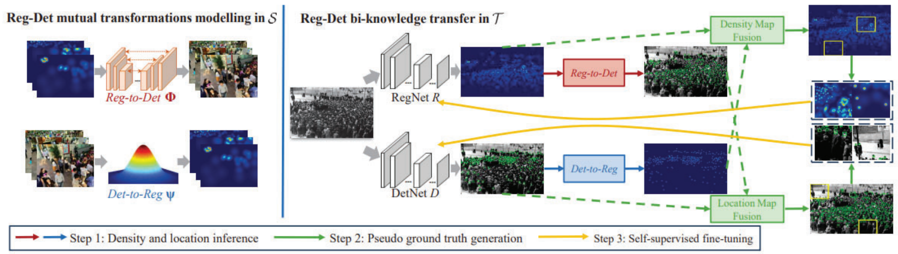


Image Processing and Enhancement [4-7]
[4] October 12: Oral Session 5c



Domain Adaptation [8-9]
[8] October 13: Poster Session 3

[1] You Only Align Once: Bidirectional Interaction for Spatial-Temporal Video Super-Resolution, **ACM MM**, 2022
[2] Progressive Spatial-temporal Collaborative Network for Video Frame Interpolation, **ACM MM**, 2022
[3] Spatial-Temporal Space Hand-in-Hand: Spatial-Temporal Video Super-Resolution via Cycle-Projected Mutual Learning, **CVPR**, 2022
[4] Magic ELF: Image Deraining Meets Association Learning and Transformer, **ACM MM**, 2022
[5] DANet: Image Deraining via Dynamic Association Learning, **IJCAI**, 2022
[6] Degrade is Upgrade: Learning Degradation for Low-light Image Enhancement, **AAAI**, 2022
[7] Image Inpainting Guided by Coherence Priors of Semantics and Textures, **CVPR**, 2021
[8] Fine-Grained Fragment Diffusion for Cross Domain Crowd Counting, **ACM MM**, 2022
[9] Towards Unsupervised Crowd Counting via Regression-Detection Bi-knowledge Transfer, **ACM MM**, 2020

Topics in this Presentation

Understanding in Harsh Environments

- Dataset [IJCAI'22]
- Factor [CVPR'22]
- Label [TIP'22]

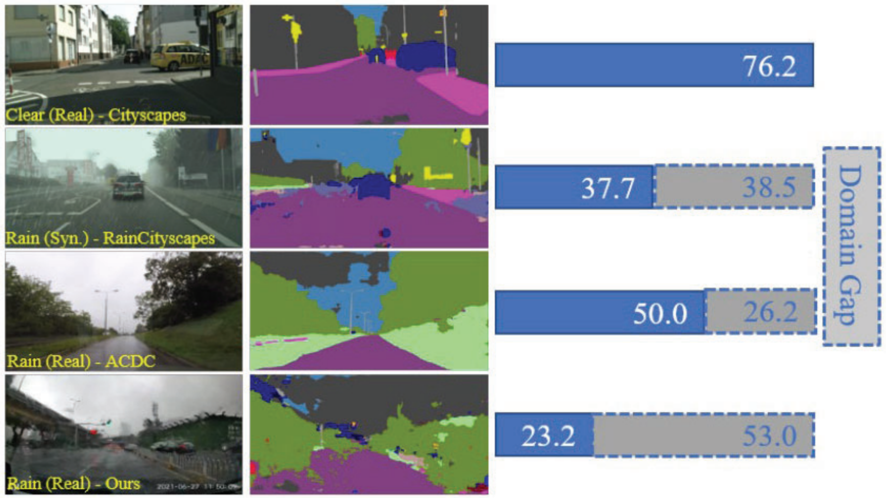
Detection in Harsh Environments

- Dataset [IJCAI'20]
- Consistency [ACM MM'21]

Retrieval in Harsh Environments

- Factor [TMM'20]
- Diffusion [TMM'22]
- Unify [CVPR'19]

[IJCAI 22] Motivation

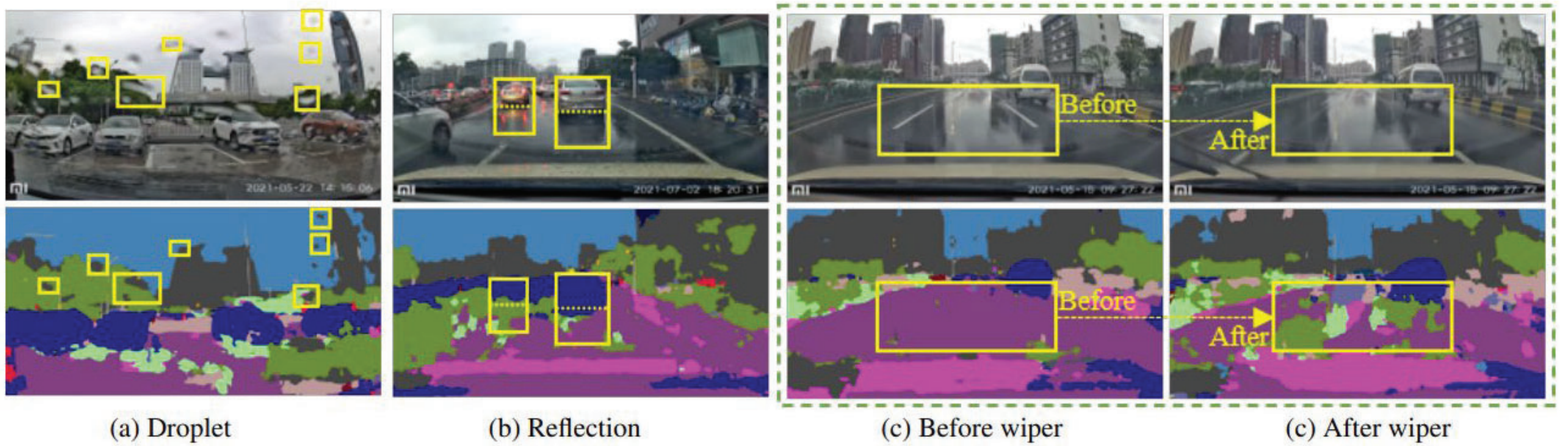


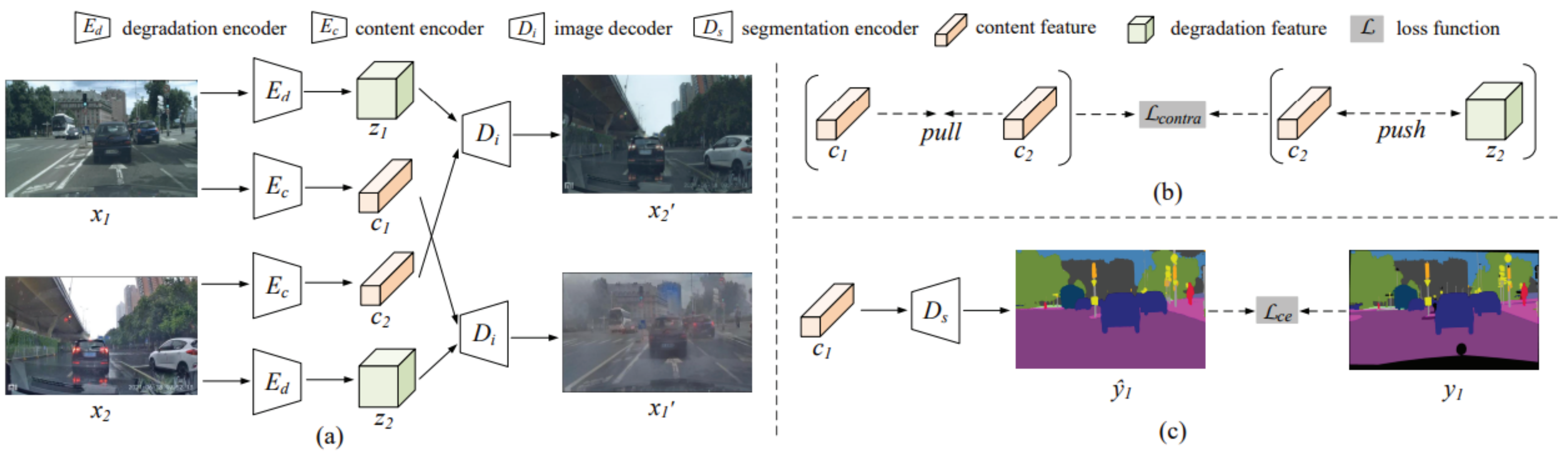
(a) Input (b) Prediction (c) mIoU

Dataset	Resolution	Label/Total	Real	Scenario					Intensity		
				Occlusion	Blur	Droplet	Reflection	Wiper	Light	Moderate	Heavy
Cityscapes [Cordts *et al.*, 2016]	2,048×1,024	0/0	×	×	×	×	×	×	×	×	×
Raincouver [Tung *et al.*, 2017]	1,280×720	285/326	✓	×	✓	✓	✓	×	✓	×	×
KITTI [Alhaija *et al.*, 2018]	1,382×512	0/0	×	×	×	×	×	×	×	×	×
RID [Li *et al.*, 2019]	Variable	0/2,495	✓	✓	✓	✓	✓	×	✓	×	×
Apolloscape [Huang *et al.*, 2020]	3,384×2,710	0/0	×	×	×	×	×	×	×	×	×
nuImages [Caesar *et al.*, 2020]	1,600×900	58/1,300	✓	×	✓	×	✓	×	✓	×	×
BDD [Yu *et al.*, 2020]	1,280×720	253/5,808	✓	✓	✓	✓	✓	×	✓	×	×
ACDC [Sakaridis *et al.*, 2021]	1,920×1,080	1,000/1,000	✓	✓	✓	×	✓	✓	✓	×	×
RainCityscapes [Hu *et al.*, 2021]	2,048×1,024	1,760/10,620	×	×	✓	✓	×	×	✓	✓	✓
RaidaR [Jin *et al.*, 2021]	1,920×1,080	5,000/58,542	✓	✓	✓	×	✓	×	✓	×	×
Rainy Wcity (Ours)	1,920×1,080	500/24,335	✓	✓	✓	✓	✓	✓	✓	✓	✓

Rainy WCity: A Real Rainfall Dataset with Diverse Conditions for Semantic Driving Scene Understanding, IJCAI, 2022

[IJCAI 22] Method



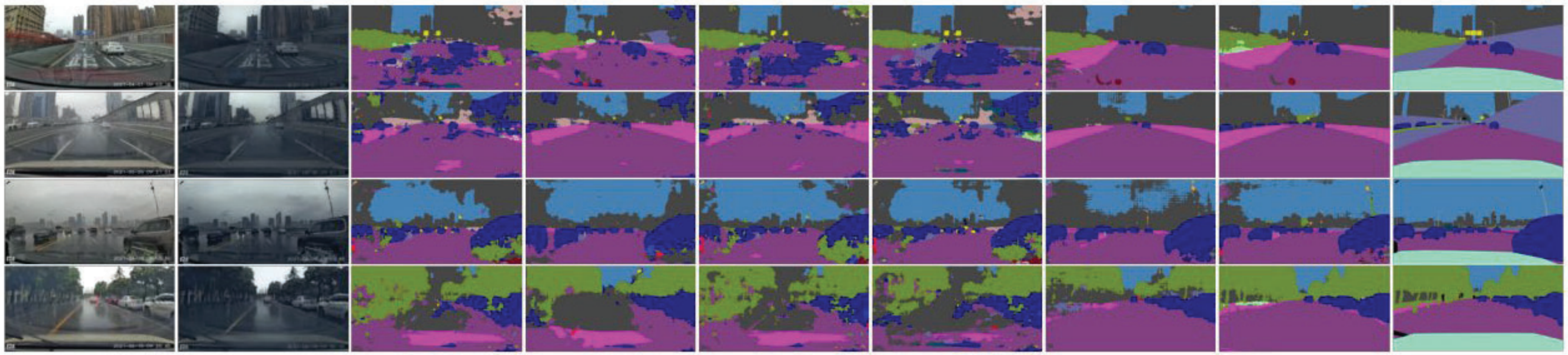


Rainy WCity: A Real Rainfall Dataset with Diverse Conditions for Semantic Driving Scene Understanding, IJCAI, 2022

[IJCAI 22] Results

Category	Method	road	sidew.	build	wall	fence	pole	light	sign	veget.	terrain	sky	person	rider	car	truck	bus	motorc.	bicycle	mIoU
Droplet	DeepLabV3+	73.5	7.5	47.4	7.4	11.5	24.6	6.9	46.5	47.0	0	78.1	5.1	0.2	43.6	1.5	0	0.1	1.5	21.2
	MPRNet	58.6	1.9	44.2	**12.4**	10.5	**25.7**	6.0	44.7	45.9	0	84.8	4.3	0.5	31.2	1.5	0.2	0.8	2.4	19.8
	DualGCN	73.8	**8.1**	50.5	6.3	7.3	23.3	6.9	46.6	46.3	0	82.3	5.4	0.7	41.8	1.0	0.5	0.1	1.3	21.1
	S2R2 (Ours)	**91.8**	6.9	**58.3**	3.2	**38.8**	18.1	**39.2**	**52.3**	**75.2**	**0.5**	**88.8**	**33.2**	**4.0**	**81.8**	**24.2**	**17.0**	**14.4**	**32.1**	**37.7**
Wiper	DeepLabV3+	63.8	0.5	47.4	6.3	0	21.9	0	25.8	42.8	0	86.6	0	0	28.9	0	3.8	0	0	17.2
	MPRNet	54.1	0	47.3	**9.6**	0	20.9	0	25.7	40.7	0	86.4	0	0	24.3	0	3.4	0	0	16.4
	DualGCN	66.3	**2.6**	48.7	3.4	0	15.8	0	28.8	40.8	0	85.5	0	0	32.4	0	5.0	0	0	17.3
	S2R2 (Ours)	**94.9**	**2.6**	**58.7**	4.7	0	**27.6**	0	**58.3**	**75.6**	0	**92.9**	0	0	**86.9**	0	**40.0**	**50.0**	0	**32.9**
Reflection	DeepLabV3+	73.5	4.8	46.2	9.8	26.0	20.0	14.5	38.7	57.0	0	86.0	7.9	0	45.7	1.1	6.0	0	0	23.0
	MPRNet	67.1	3.3	44.9	**13.2**	22.9	**20.7**	17.7	35.2	56.2	0	87.4	5.9	0	39.9	1.1	3.5	0.1	0	22.1
	DualGCN	72.2	4.6	45.5	9.0	11.2	20.2	14.6	**39.3**	54.1	0	88.6	9.2	0	43.5	1.9	4.0	0	0	22.0
	S2R2 (Ours)	**87.1**	**11.5**	**60.4**	9.0	**60.2**	20.1	**36.0**	22.2	**79.8**	0	**91.6**	**28.1**	**6.1**	**81.5**	**77.3**	**8.3**	**1.9**	0	**37.8**

Qualitative comparison with DeepLabV3+

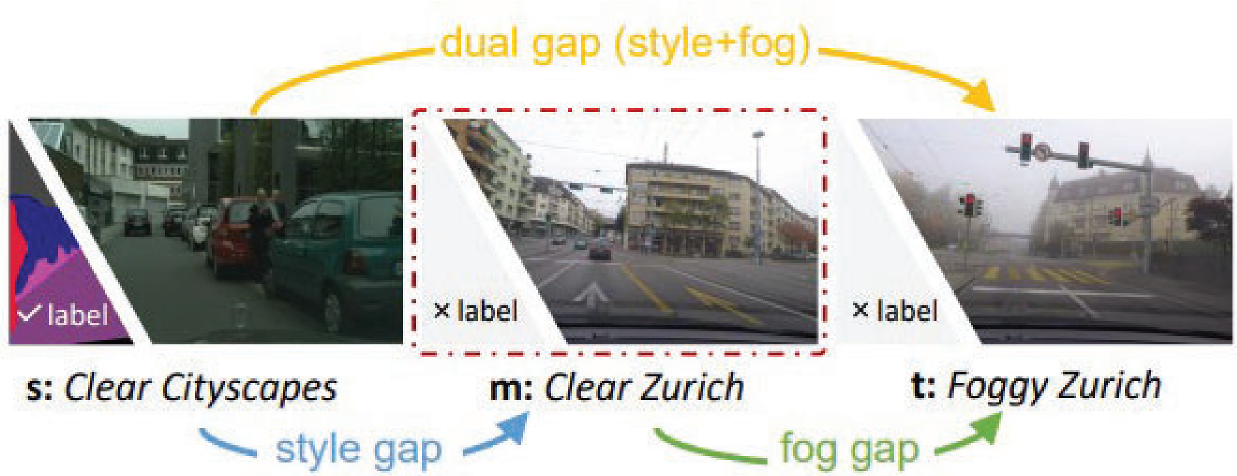


(a) Input (b) Derain (ours) (c) V3+ (d) V3+ (derain) (e) DualGCN (f) MPRNet (g) Our (derain) (h) Ours (i) Ground Truth

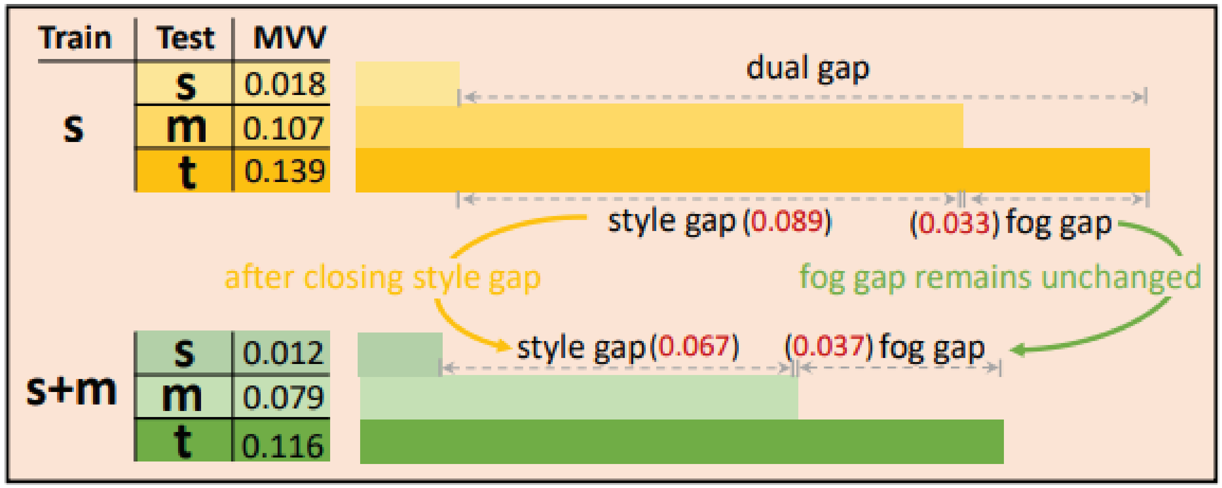
Rainy WCity: A Real Rainfall Dataset with Diverse Conditions for Semantic Driving Scene Understanding, IJCAI, 2022

[CVPR 22] Motivation

The problem and our main idea

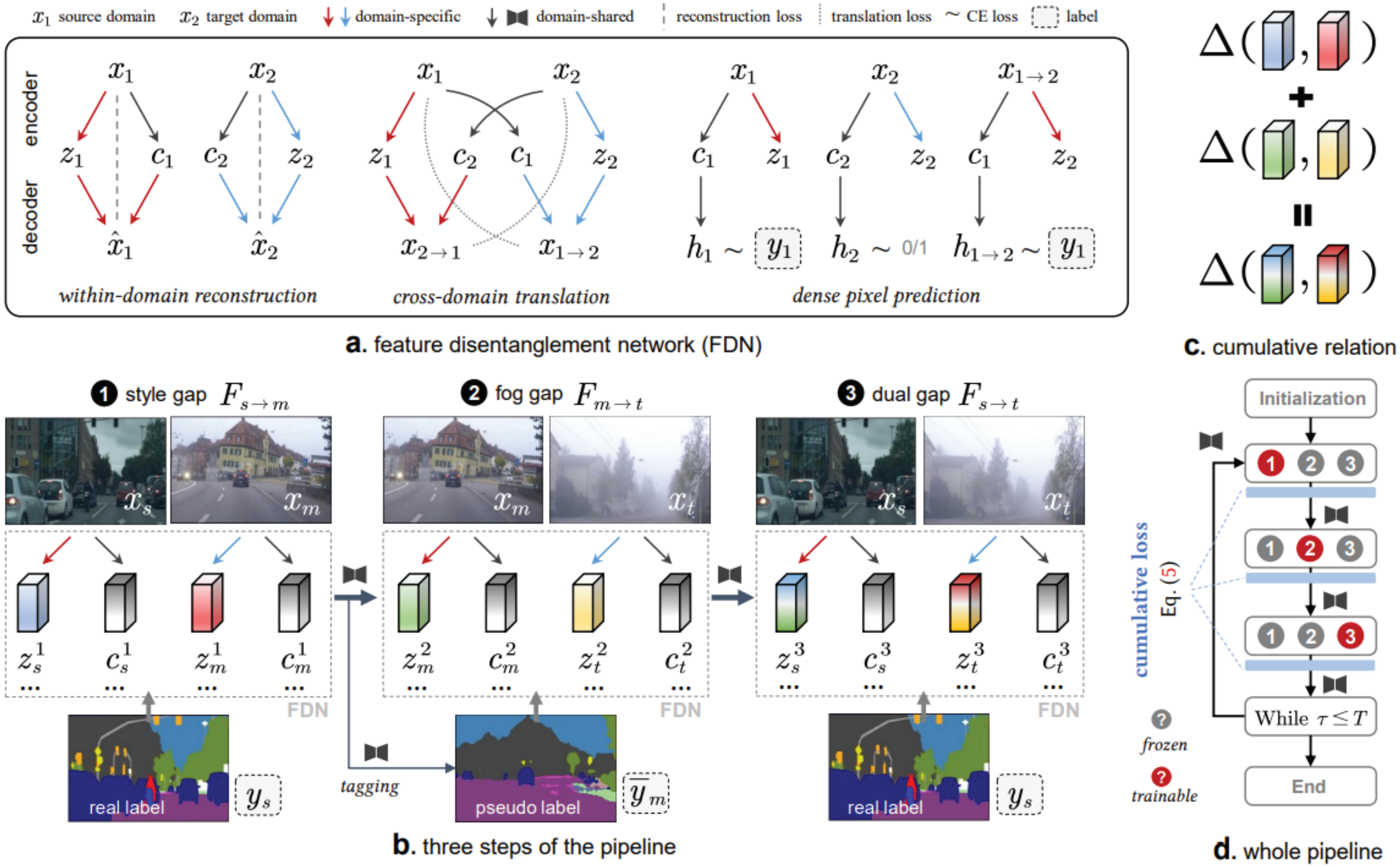


Empirical finding of the motivation



Both Style and Fog Matter: Cumulative Domain Adaptation for Semantic Foggy Scene Understanding, CVPR, 2022

[CVPR 22] Method



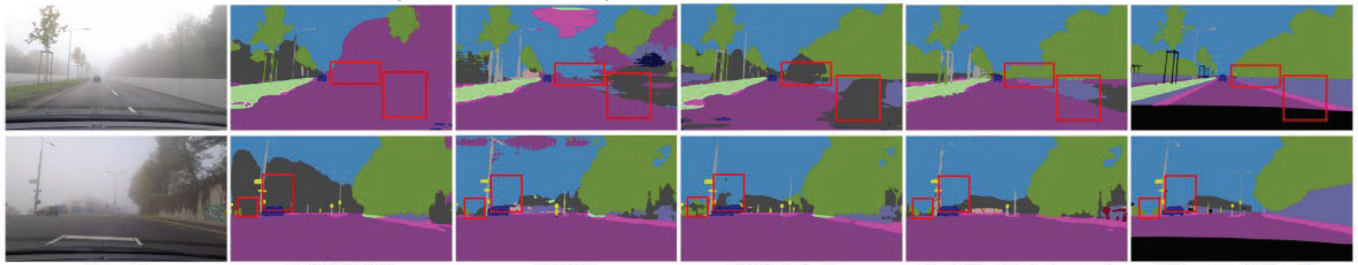
Both Style and Fog Matter: Cumulative Domain Adaptation for Semantic Foggy Scene Understanding, CVPR, 2022

[CVPR 22] Results

Performance comparison

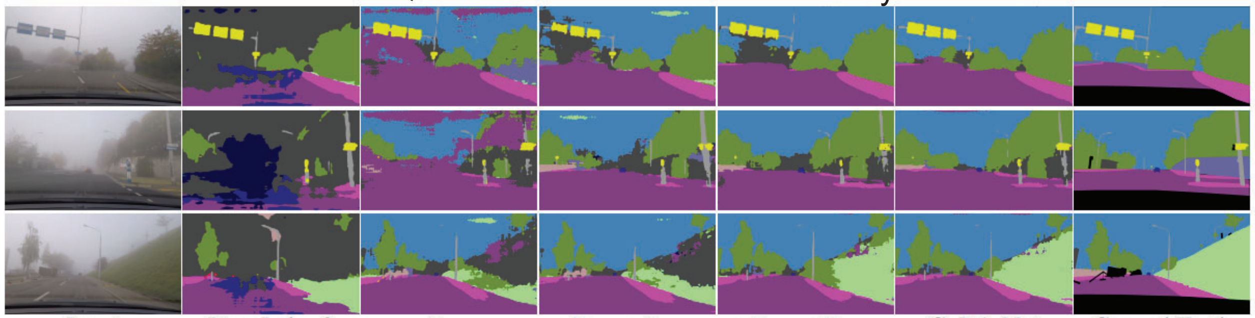
Experiment	Method	Backbone	FZ	FD
Backbone	–	DeepLab-v2	25.9	35.7
	–	RefineNet	34.6	35.8
Defogging	MSCNN [24]	RefineNet	34.4	38.3
	DCP [15]	RefineNet	31.2	33.2
	Non-local [3]	RefineNet	27.6	32.8
	GFN [25]	DeepLab-v2	27.5	37.2
	DCPDN [38]	DeepLab-v2	28.7	37.9
Domain Adaptation	Multi-task [1]	–	26.1	31.6
	AdSegNet [34]	DeepLab-v2	26.1	37.6
	ADVENT [35]	DeepLab-v2	24.5	36.1
	DISE [4]	DeepLab-v2	40.7	45.2
	CCM [19]	DeepLab-v2	35.8	42.6
	SAC [2]	DeepLab-v2	37.0	43.4
	ProDA [39]	DeepLab-v2	37.8	41.2
	DMLC [13]	DeepLab-v2	33.5	32.6
	DACS [32]	DeepLab-v2	28.7	35.0
Defogging+DA	MSCNN [24]+DISE [4]	DeepLab-v2	38.6	37.1
Ours	CuDA-Net	DeepLab-v2	**48.2**	**52.7**
Synthesis†	SFSU [28]	RefineNet	35.7	35.9
	CMAda2 [27]	RefineNet	42.9	37.3
	CycleGAN [43]	RefineNet	40.5	47.7
	MUNIT [16]	RefineNet	39.1	47.8
	AnalogicalGAN [12]	RefineNet	42.3	47.5
	CMAd[illegible]	[illegible]	[illegible]	[illegible]
Synthesis+DA	SFSU [28]+D[illegible]	[illegible]	[illegible]	[illegible]
Ours	CuDA-Net+	DeepLab-v2	**49.1**	**53.5**

The qualitative comparison with the SOTA methods



Input MSCNN DISE CMAda3+ CuDA-Net (ours) Ground Truth

Qualitative results of ablation study



Input DeepLab-v2 $+F_{s\to m}$ $+F_{s\to m}+F_{m\to t}$ $+F_{s\to m}+F_{m\to t}+F_{s\to t}$ CuDA-Net Ground Truth

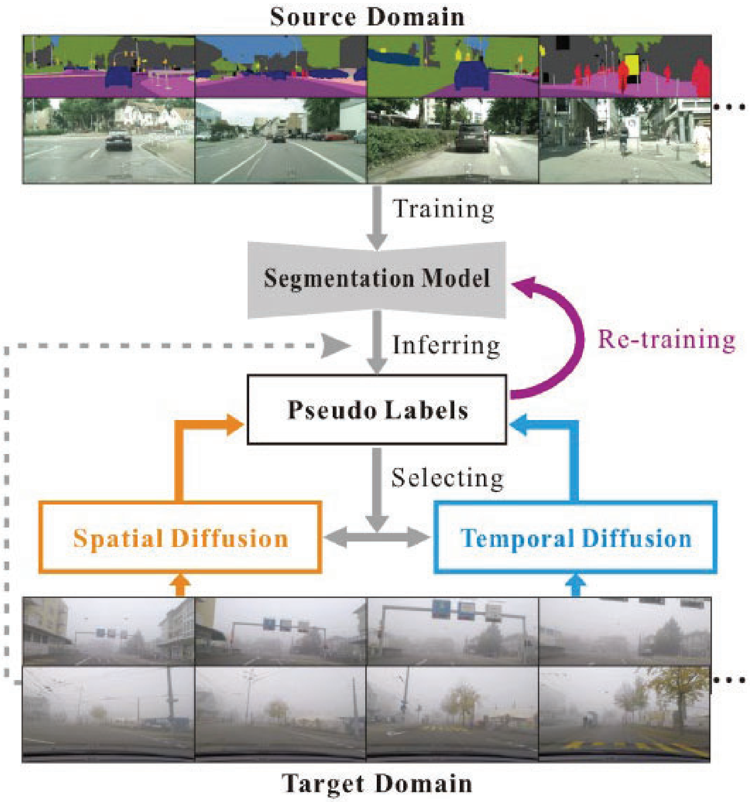
The ability of

Input GFN Ours ($F_{m\to t}$)

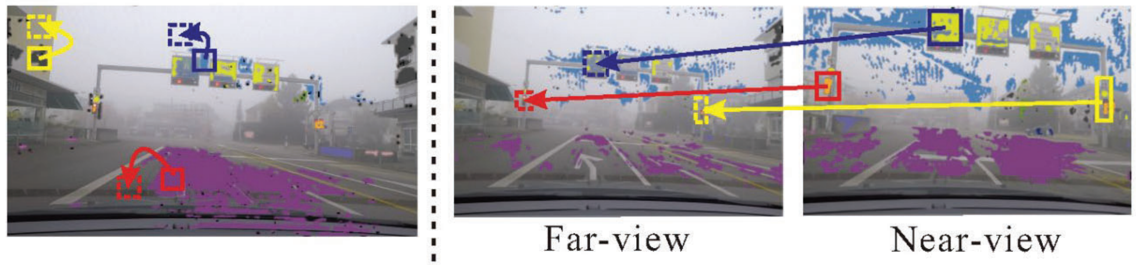
focus on the degradation factors in harsh environments

Both Style and Fog Matter: Cumulative Domain Adaptation for Semantic Foggy Scene Understanding, CVPR, 2022

[TIP 22] Motivation



(a) Proposed TDo-Dif Scheme for Domain Adaptation



(b) Spatial Diffusion (c) Temporal Diffusion

Unsupervised Foggy Scene Understanding via Self Spatial-Temporal Label Diffusion, IEEE Trans. Image Processing, 2022

[TIP 22] Method

Illustration of the superpixel-based spatial diffusion

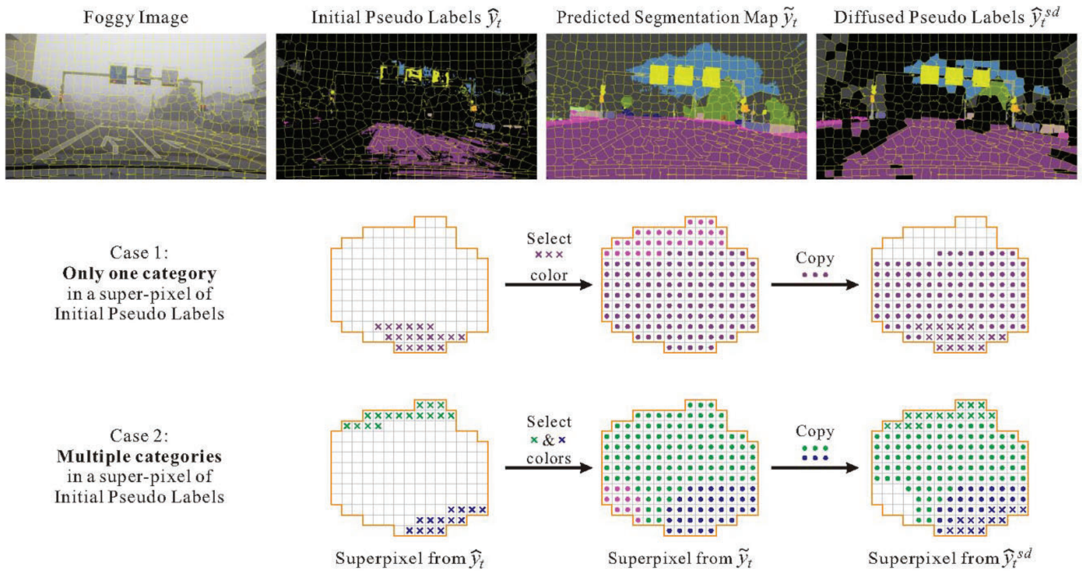
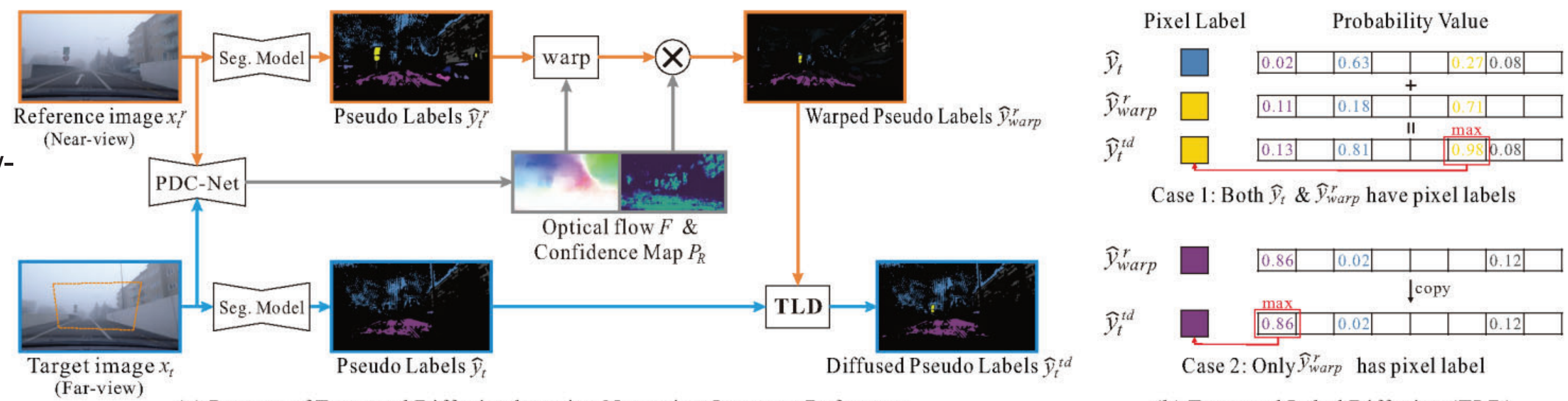


Illustration of the optical flow-based temporal diffusion



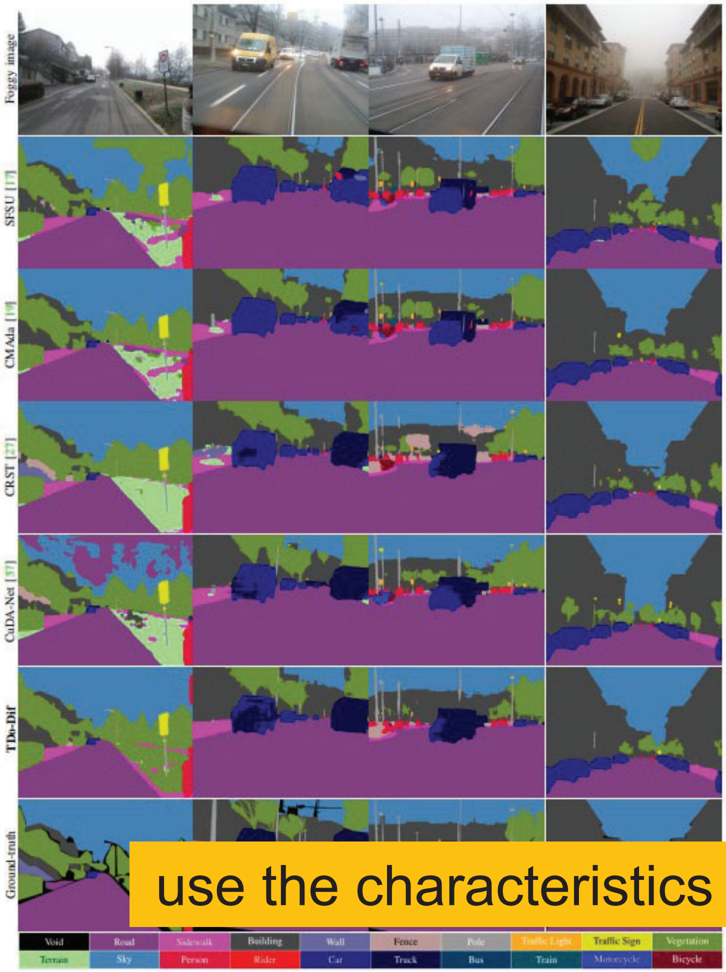
(a) Process of Temporal Diffusion by using Near-view Image as Reference

(b) Temporal Label Diffusion (TLD)

Unsupervised Foggy Scene Understanding via Self Spatial-Temporal Label Diffusion, IEEE Trans. Image Processing, 2022

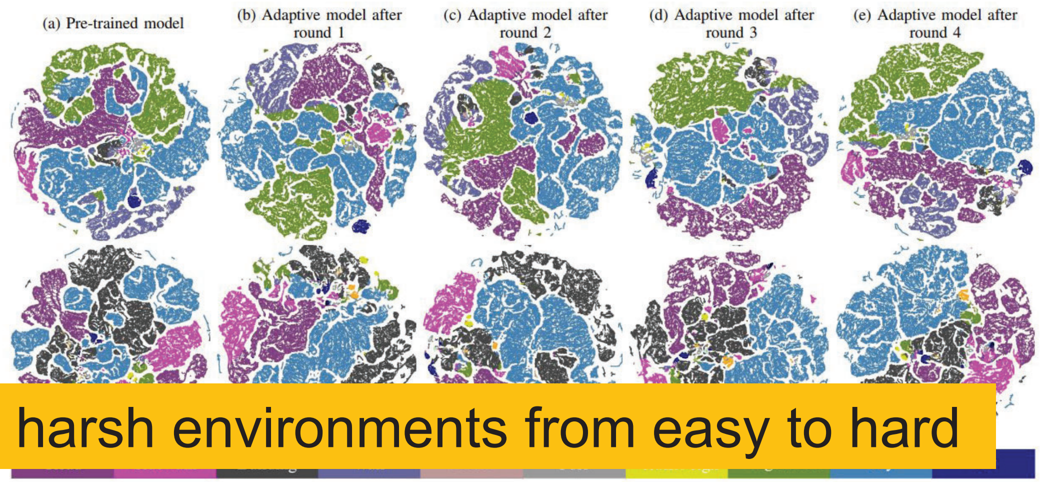
[TIP 22] Results

Foggy Driving



Method	Road	SW	Build	Wall	Fence	Pole	TL	TS	Veg.	Terrain	Sky	PR	Rider	Car	Truck	Bus	Train	Motor	Bike	mIoU
SFSU [17]	90.26	28.83	72.13	25.23	13.41	42.84	52.03	58.97	64.27	5.78	76.71	57.26	44.02	70.41	13.42	27.73	58.48	19.29	46.48	45.66
CMAda [19]	91.51	29.24	74.77	28.37	15.10	**49.36**	51.35	59.26	**74.76**	7.82	**92.29**	62.63	47.67	72.90	19.38	**32.48**	52.05	24.62	**52.81**	49.39
AdSegNet [56]	45.82	13.52	43.34	0.63	8.94	25.97	37.57	35.92	54.12	0.53	80.70	30.73	27.08	56.74	0.73	12.58	0.40	11.19	26.47	27.00
CBST [21]	91.68	31.35	68.63	25.61	15.98	48.14	49.48	**60.02**	67.85	10.37	82.18	62.22	41.62	73.30	36.96	15.69	31.69	29.90	46.95	46.82
CRST [27]	91.82	36.34	70.59	23.93	16.33	46.02	49.66	56.92	70.84	12.68	86.36	64.25	42.17	**75.07**	30.72	13.24	31.32	**35.06**	45.70	47.32
CuDA-Net [57]	90.14	**45.52**	71.47	**43.63**	**44.23**	43.83	46.30	52.24	72.63	**36.18**	91.19	59.90	47.90	72.04	**48.58**	**40.96**	32.81	33.47	44.09	**53.50**
FIFO [58]	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	50.70
CMDIT [59]	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	45.35
FogAdapt+ [60]	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	52.40
TDo-Dif† (SD→TD+SL+TL)	**92.12**	31.84	74.82	28.44	17.91	43.72	**52.45**	56.12	71.40	12.86	86.71	**64.31**	**47.98**	73.11	38.40	23.19	55.34	27.43	**53.46**	50.08
TDo-Dif† (TD→SD+SL+TL)	92.09	31.80	**74.87**	27.72	17.93	45.01	**52.77**	57.92	71.33	**13.73**	86.63	**64.62**	**48.09**	**73.69**	38.69	24.59	**60.25**	28.11	52.45	50.65
TDo-Dif* (SD+SL)	**93.03**	**39.26**	**76.72**	**33.35**	**18.77**	**48.35**	50.17	**64.41**	**79.99**	2.32	**92.66**	61.87	46.64	78.31	**44.63**	28.22	**70.78**	**41.58**	51.58	**53.84**

TDo-Dif† and **TDo-Dif*** denote the results from the model trained on **Foggy Zurich** and **Foggy Driving**, respectively.
Note that the images in **Foggy Driving** dataset are non-sequential images, thus we only use the spatial diffusion and spatial loss.



use the characteristics of data in harsh environments from easy to hard

Unsupervised Foggy Scene Understanding via Self Spatial-Temporal Label Diffusion, IEEE Trans. Image Processing, 2022

Topics in this Presentation

Understanding in Harsh Environments

- Dataset [IJCAI'22]
- Factor [CVPR'22]
- Label [TIP'22]

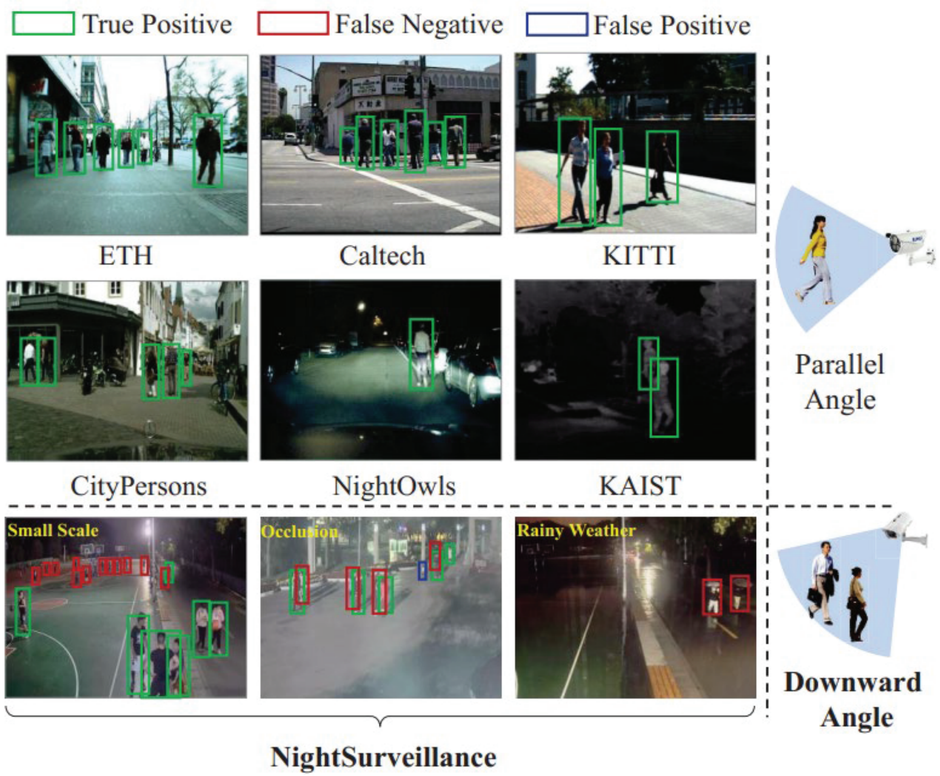
Detection in Harsh Environments

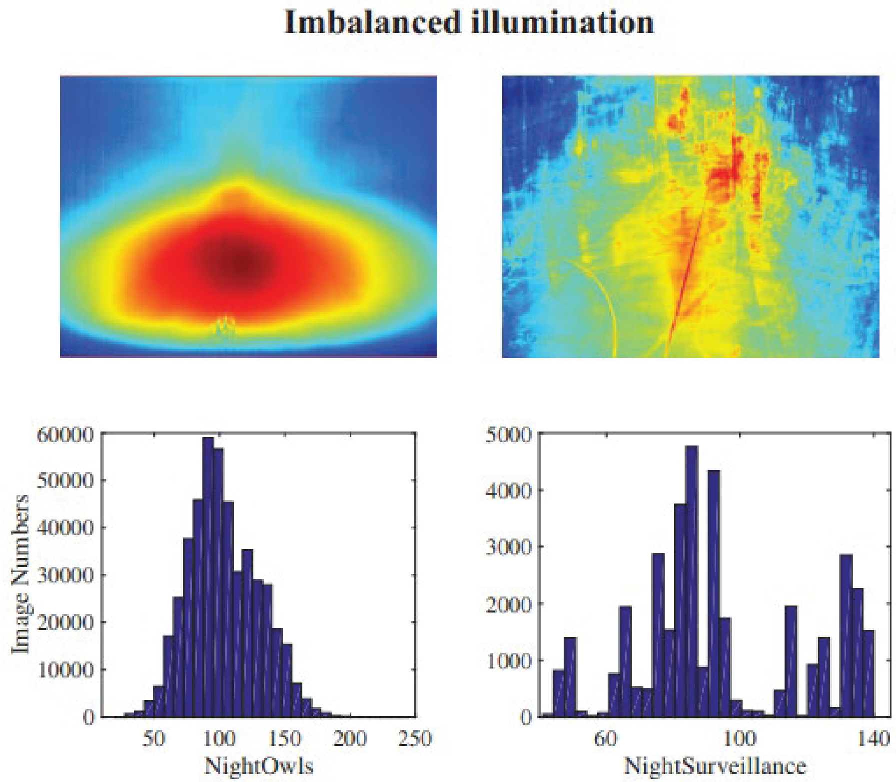
- Dataset [IJCAI'20]
- Consistency [ACM MM'21]

Retrieval in Harsh Environments

- Factor [TMM'20]
- Diffusion [TMM'22]
- Unify [CVPR'19]

[IJCAI 20] Motivation





When Pedestrian Detection Meets Nighttime Surveillance: A New Benchmark, IJCAI, 2020

[IJCAI 20] Method

The number of frames and pedestrian annotations in datasets.

	Dataset	Train	Test	All	
		#images/#bboxes	#images/#bboxes	#images	#pedestrian/frame ↑
Daytime	KITTI [Geiger *et al.*, 2012]	$7k/4k$	–	$7k$	0.6
	Daimler [Enzweiler and Gavrila, 2009]	$22k/14k$	–	$22k$	0.65
	INRIA [Dalal and Triggs, 2005]	$2k/1k$	288/589	$2k$	0.86
	Caltech [Dollar *et al.*, 2012]	$128k/153k$	$121k/132k$	$250k$	1.14
	TUD [Wojek *et al.*, 2009]	$508/1k$	–	508	2.95
	CityPersons [Zhang *et al.*, 2017]	$3k/17k$	$1.5k/14k$	$5k$	7
	ETH [Ess *et al.*, 2008]	$2k/14k$	–	$2k$	7.85
Nighttime	NightOwls [Neumann *et al.*, 2018]	$128k/38k$	$103k/8k$	$231k$	0.20
	KAIST [Hwang *et al.*, 2015]	$17k/17k$	$16k/12k$	$33k$	0.86
	NightSurveillance	$19k/26k$	$19k/26k$	$38k$	2.46

Comparison of the annotation attributes.

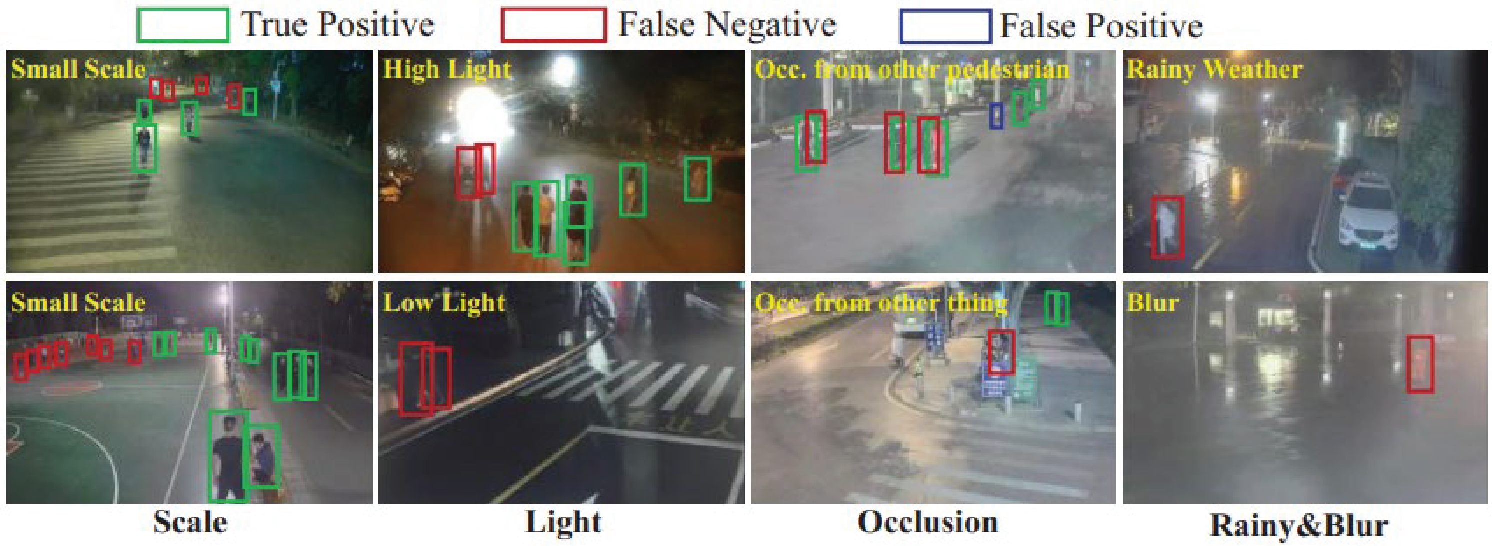
Dataset	ImageSize	Data Diversity				
		Occlusion	Scale	Blur	Rainy	Lighting
KITTI	1392×512	✓				
Caltech	640×480	✓	✓			
CityPersons	2048×1024	✓	✓			
KAIST	640×480		✓			
NightOwls	1024×640	✓	✓			✓
NightSurveillance	1920×1080	✓	✓	✓	✓	✓

The proportion of pedestrians with different settings in NightSurveillance dataset

Setting	#Occlusion	#Scale			#Lighting			#Blur	#Rainy	#All
		#Small	#Medium	#Large	#Low	#Medium	#High			
Train	$12k$(25%)	$22k$(47%)	$12k$(26%)	$13k$(28%)	$9k$(20%)	$30k$(63%)	$8k$(17%)	$1k$(2%)	$2k$(4%)	$47k$
Test	$12k$(26%)	$21k$(46%)	$12k$(26%)	$13k$(28%)	$8k$(17%)	$30k$(65%)	$8k$(18%)	$1k$(2%)	$2k$(4%)	$46k$
All	$24k$(26%)	$43k$(46%)	$24k$(26%)	$26k$(28%)	$17k$(18%)	$60k$(65%)	$16k$(17%)	$2k$(2%)	$4k$(4%)	$93k$

When Pedestrian Detection Meets Nighttime Surveillance: A New Benchmark, IJCAI, 2020

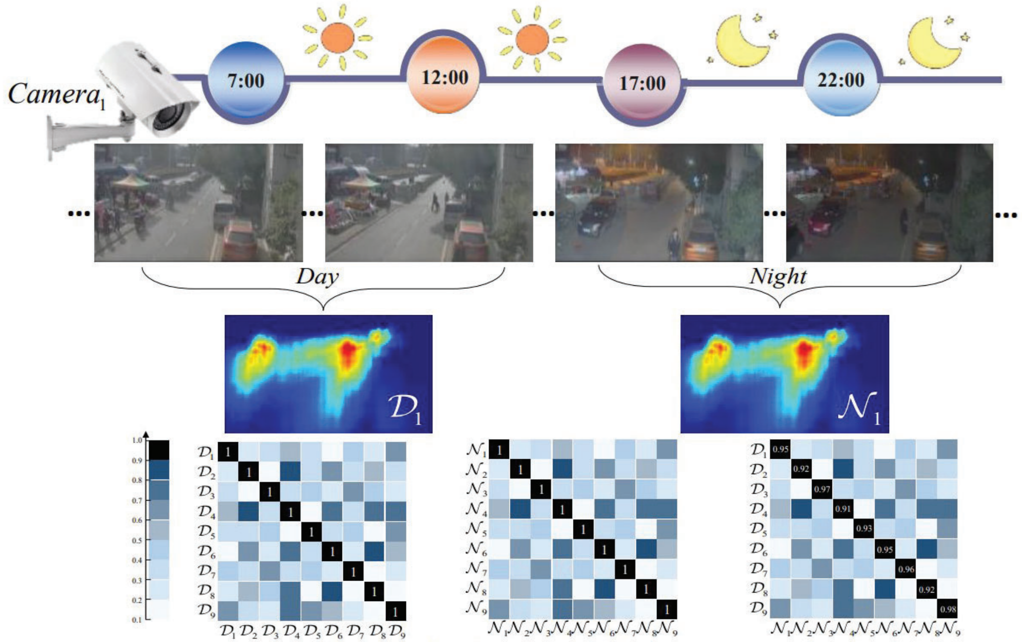
[IJCAI 20] Results



Methods	KITTI	Caltech	CityPersons	NightOwls	*NightSurveillance*
	mAP (%)	Average Miss Rate (%)			
ACF [Dollar *et al.*, 2014]	47.29	27.63	33.10	51.68	89.34
RPN+BF [Zhang *et al.*, 2016a]	61.29	9.58	7.31	23.26	86.34
Vanilla Faster [Ren *et al.*, 2017] R-CNN	65.91	20.98	23.46	20.00	26.55
Adapted Faster R-CNN [Zhang *et al.*, 2017]	66.72	10.27	12.81	18.81	24.84
SDS R-CNN [Brazil *et al.*, 2017]	63.05	7.36	13.26	17.80	23.62
S3D [Wang *et al.*, 2019]	65.60	9.28	11.24	14.32	21.73

When Pedestrian Detection Meets Nighttime Surveillance: A New Benchmark, IJCAI, 2020

[ACM MM 21] Motivation

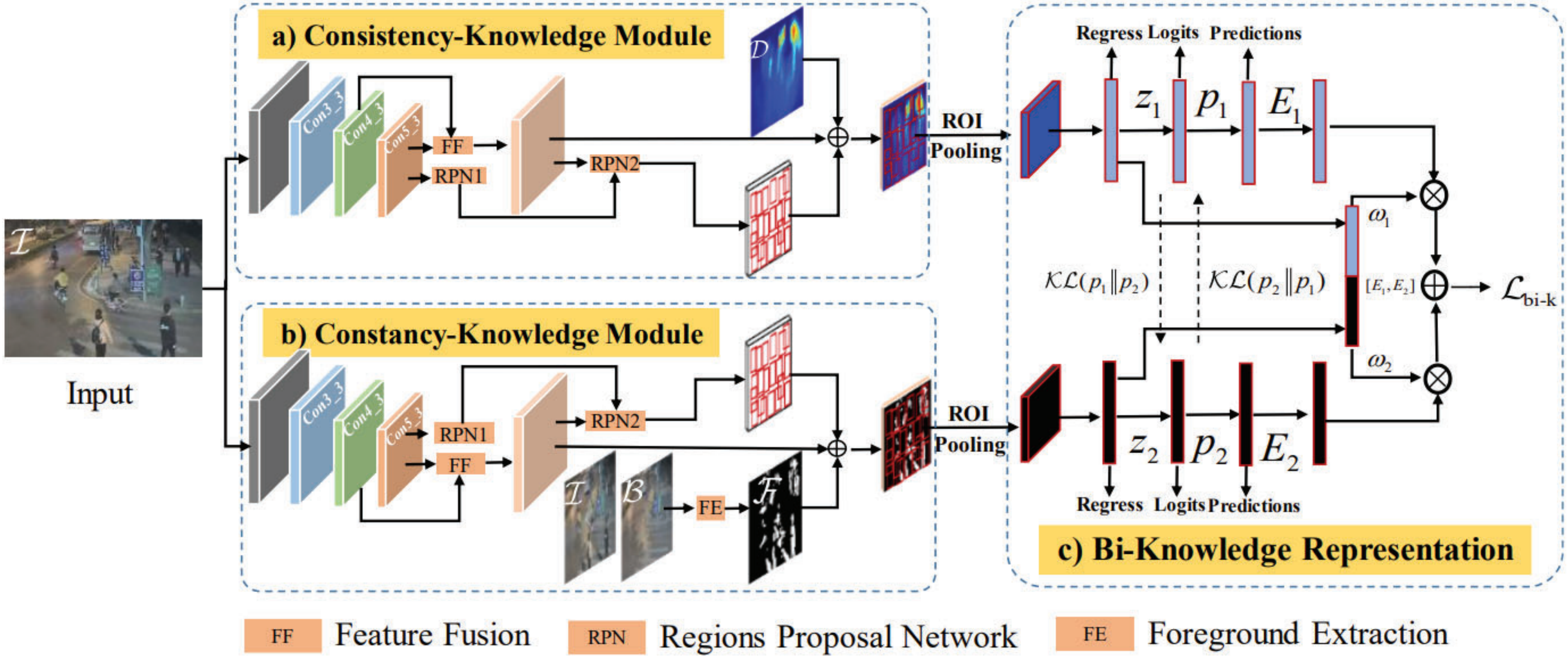


(a) Distribution Cross-Time Consistency

(b) Background Cross-Frame Constancy

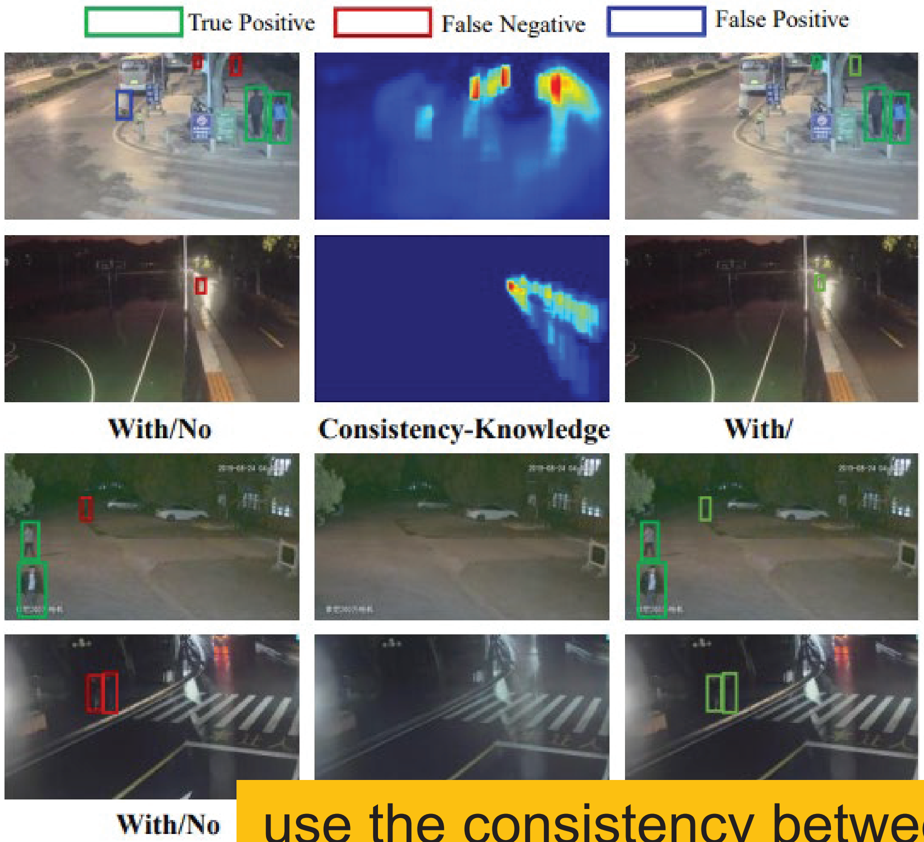
Consistency-Constancy Bi-Knowledge Learning for Pedestrian Detection in Night Surveillance, ACM MM, 2021

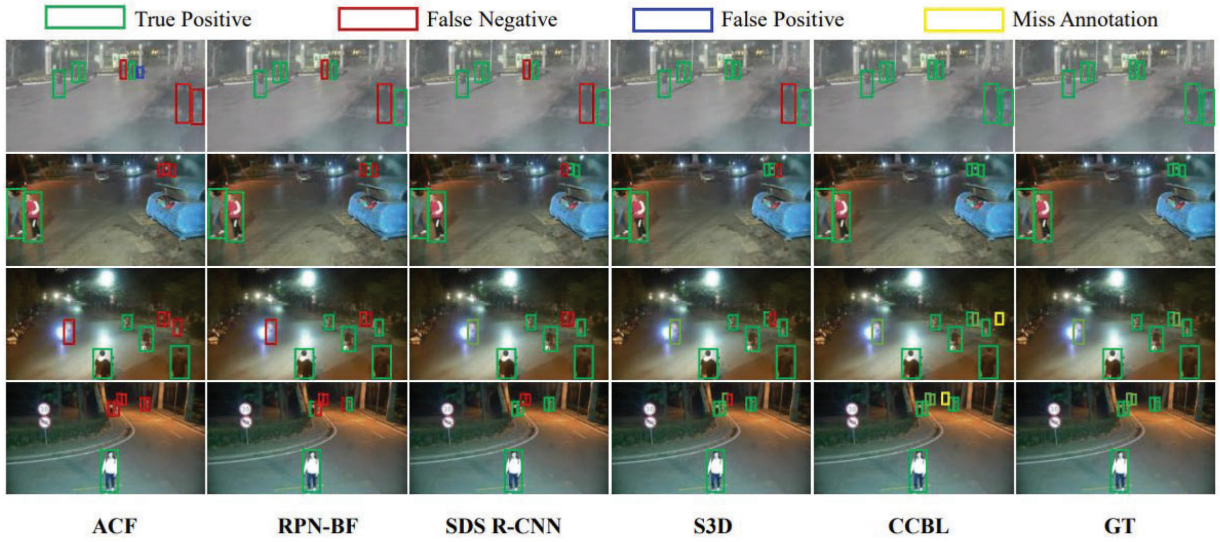
[ACM MM 21] Method



Consistency-Constancy Bi-Knowledge Learning for Pedestrian Detection in Night Surveillance, ACM MM, 2021

[ACM MM 21] Results





Methods	Scale			Light			Occlusion	Rain	Blur	Mean
	Larger	Medium	Small	High	Medium	Low				
ACF [33]	82.00	89.94	89.94	87.19	80.52	84.84	88.64	89.37	89.90	86.93
RPN-BF [43]	76.78	82.69	73.69	78.63	69.69	82.95	87.19	85.82	84.94	80.26
Vanilla Faster R-CNN [37]	21.36	43.54	68.16	40.57	47.02	49.79	50.78	62.28	62.93	49.60
Adapted Faster R-CNN [25]	9.02	41.12	62.10	46.02	15.92	34.56	30.11	51.19	50.18	37.80
SDS R-CNN [56]	8.26	23.31	41.19	25.93	11.94	29.80	25.54	34.94	45.05	27.33
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	32.61	43.23	25.28
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	**28.91**	**38.43**	**22.42**

use the consistency between excellent and harsh environments

Consistency-Constancy Bi-Knowledge Learning for Pedestrian Detection in Night Surveillance, ACM MM, 2021

Topics in this Presentation

Understanding in Harsh Environments

- Dataset [IJCAI’22]
- Factor [CVPR’22]
- Label [TIP’22]

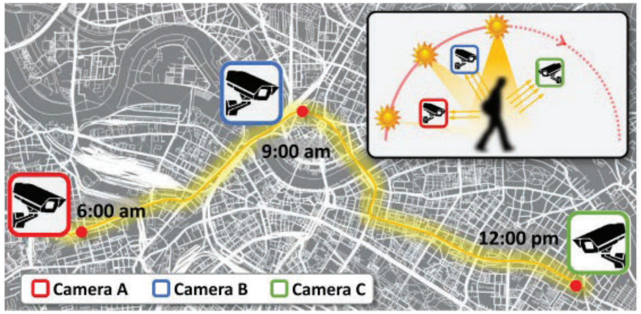
Detection in Harsh Environments

- Dataset [IJCAI’20]
- Consistency [ACM MM’21]

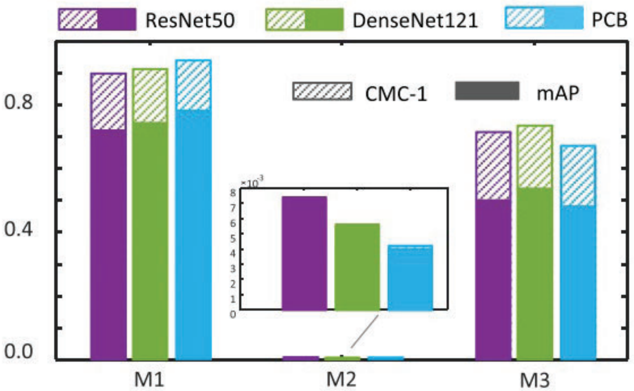
Retrieval in Harsh Environments

- Factor [TMM’20]
- Diffusion [TMM’22]
- Unify [CVPR’19]

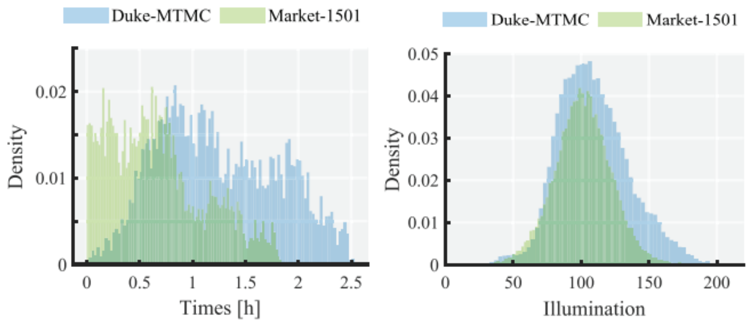
[TMM 20] Motivation

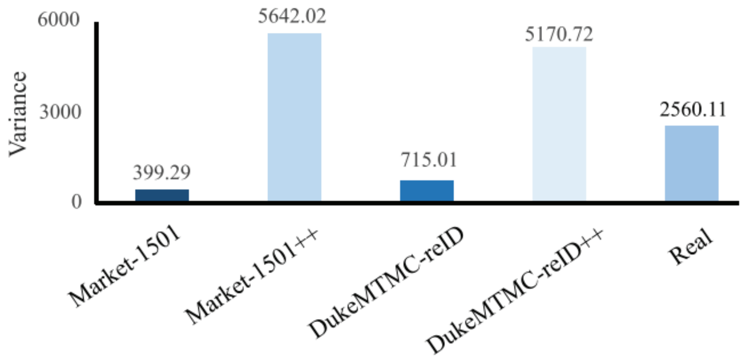


(a)



(b)

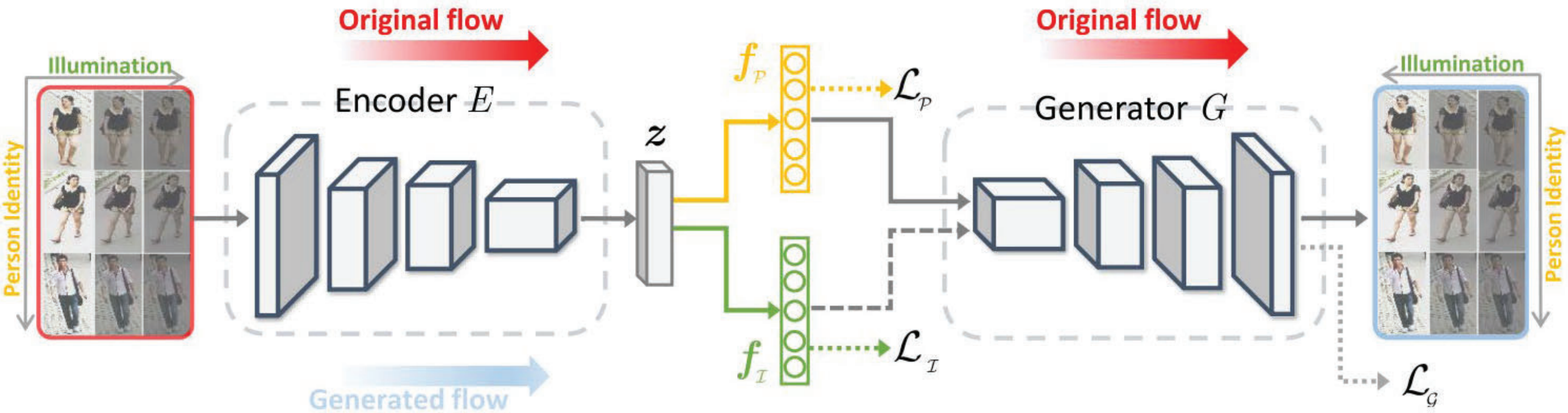




(a) Market-1501++ (b) DukeMTMC-reID++

Illumination-Adaptive Person Re-identification, IEEE Trans. Multimedia, 2020

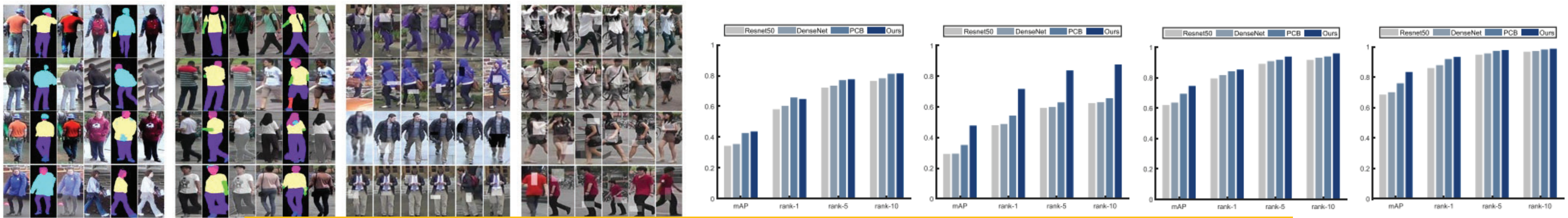
[TMM 20] Method



Illumination-Adaptive Person Re-identification, IEEE Trans. Multimedia, 2020

[TMM 20] Method

Method	Market-1501++				DukeMTMC-reID++			
	CMC-1	CMC-5	CMC-10	mAP	CMC-1	CMC-5	CMC-10	mAP
DenseNet121 [44]	0.74	2.29	3.53	0.73	1.21	2.74	4.13	0.80
DenseNet121 w/ Train	70.60	85.36	89.66	49.79	64.45	77.82	82.45	45.12
PCB [45]	0.56	1.69	2.91	0.54	0.72	2.15	3.23	0.49
PCB w/ Train	72.55	85.22	90.08	53.11	65.98	77.93	82.21	45.15
ResNet50 [46]	0.42	1.16	2.05	0.39	0.54	1.97	3.14	0.50
ResNet50 w/ Train (Baseline)	66.18	81.97	87.02	47.71	62.07	75.54	88.08	42.63
IID	**73.37**	**86.55**	**91.01**	**56.22**	**68.11**	**79.75**	**91.27**	**49.20**
Improvement over baseline	7.19↑	4.58↑	3.99↑	8.51↑	6.04↑	4.21↑	3.19↑	6.57↑



(a) Foreground Duke

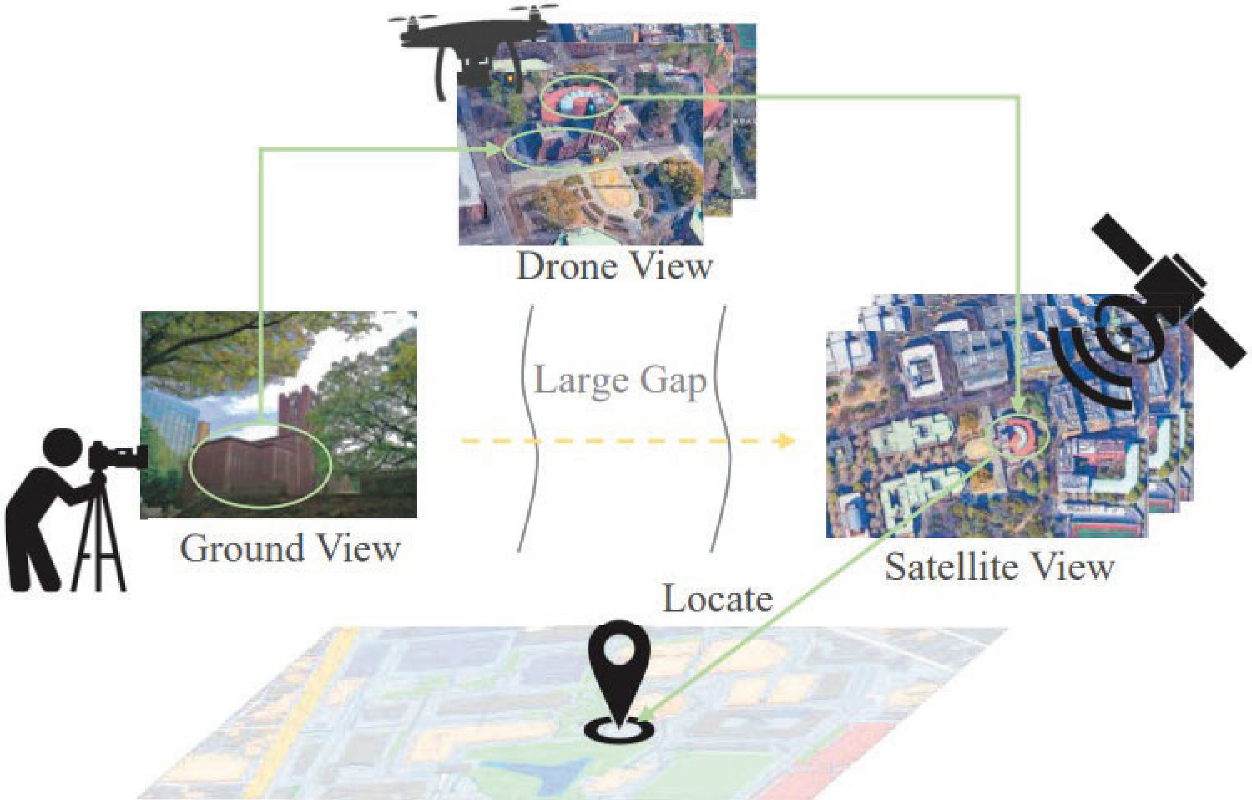
(b) Foreg

(d) Patch Market

disentangle the influence factor from harsh environments

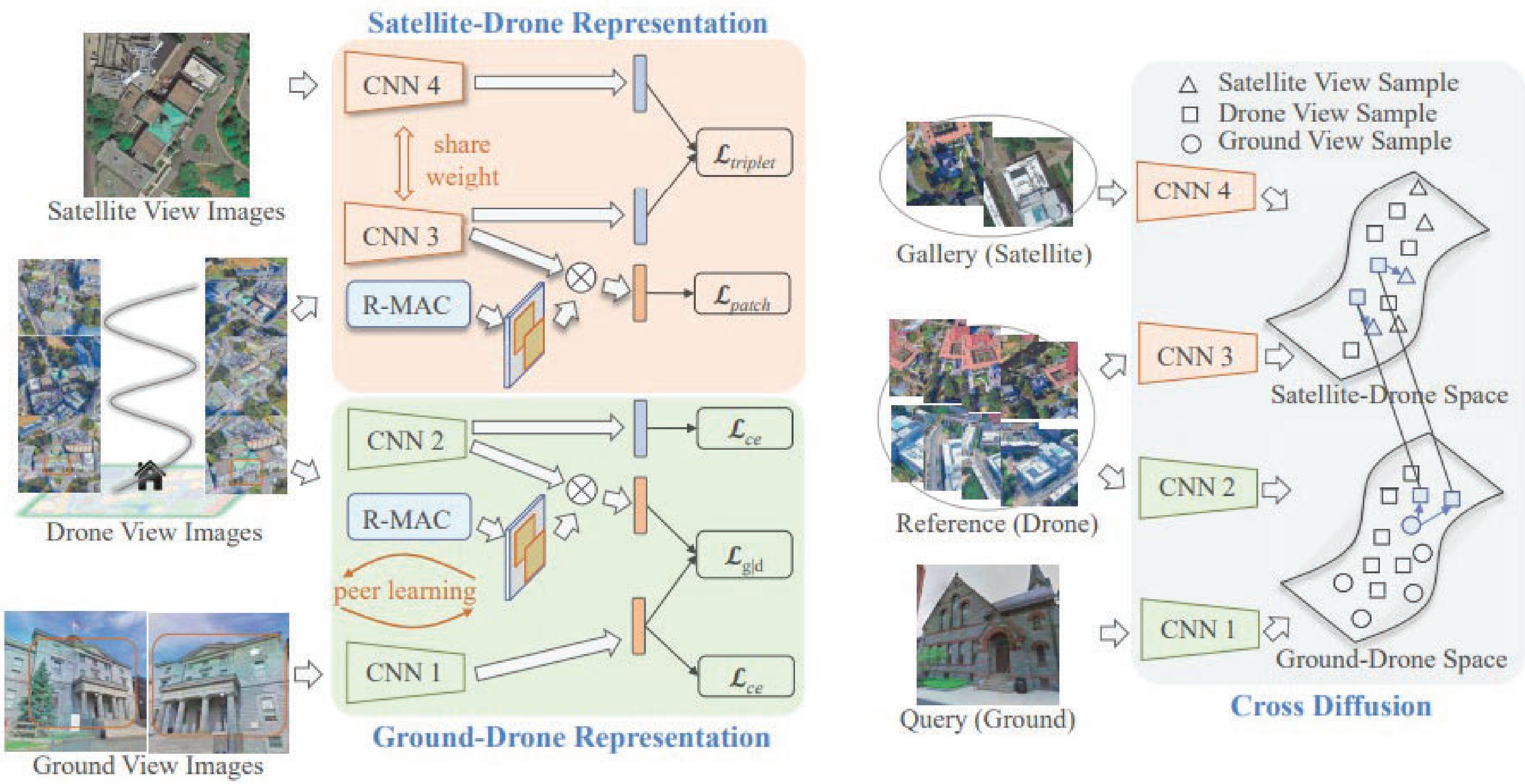
Illumination-Adaptive Person Re-identification, IEEE Trans. Multimedia, 2020

[TMM 22] Motivation



Geo-Localization via Ground-to-Satellite Cross-View Image Retrieval, IEEE Trans. Multimedia, 2022

[TMM 22] Method



Geo-Localization via Ground-to-Satellite Cross-View Image Retrieval, IEEE Trans. Multimedia, 2022

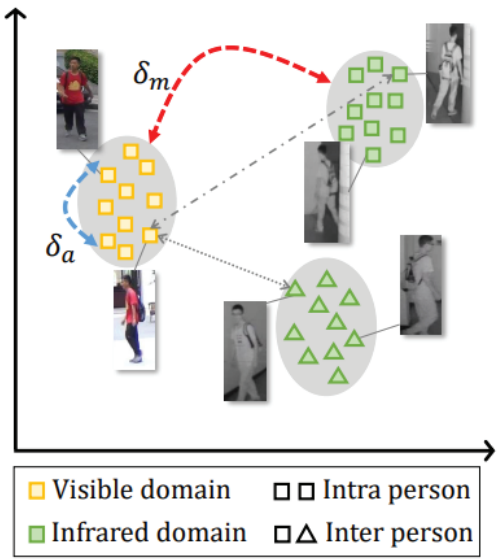
[TMM 22] Results

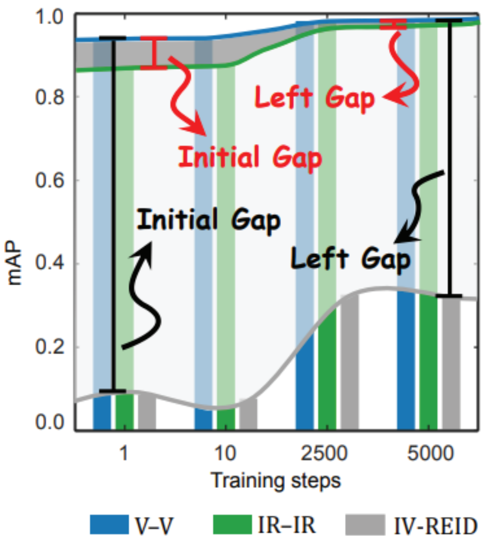
Method	University-Earth				
	CMC@1	CMC@5	CMC@10	CMC@1%	mAP
DELF [36] *w/o* $\mathcal{D}$	0.01	0.39	0.66	0.74	0.60
DELF [36]	0.12	0.50	0.93	0.93	0.87
R-MAC [41] *w/o* $\mathcal{D}$	1.09	3.61	6.59	6.94	2.19
R-MAC [41]	1.09	3.84	6.67	7.06	2.22
Str-CNN [10] *w/o* $\mathcal{D}$	0.74	2.79	4.85	5.66	1.70
Str-CNN [10]	1.01	3.22	6.01	6.63	2.08
Str-CNN [10] + Multi-loss	1.51	5.39	9.77	10.55	3.12
CVM-Net [9] *w/o* $\mathcal{D}$	0.35	1.05	2.09	2.29	0.88
CVM-Net [9]	1.78	4.69	8.61	9.42	3.18
Siam-FCANet50 [12] *w/o* $\mathcal{D}$	0.39	2.02	3.84	4.23	1.34
Siam-FCANet50 [12]	1.20	4.07	7.25	7.68	2.46
LPN [26] *w/o* $\mathcal{D}$	0.16	0.78	1.82	2.06	0.65
LPN [26]	0.74	3.10	4.69	5.04	1.70
Instance Loss [42] *w/o* $\mathcal{D}$	0.62	-	5.51	-	1.60
Instance Loss [42]	1.20	-	7.56	-	2.52
PLCD (Ours)	**9.15**	**27.66**	**38.83**	**40.87**	**14.16**

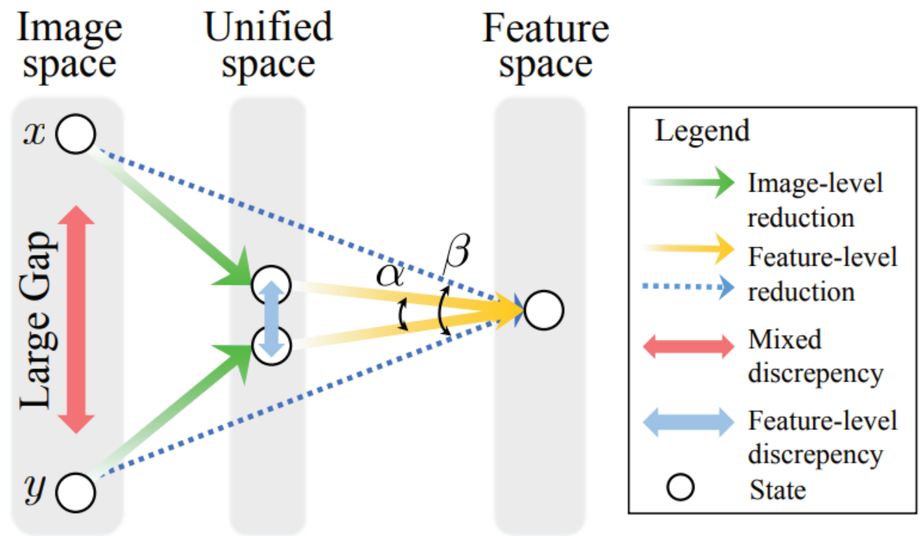
diffuse the results in harsh environments from good to better

Geo-Localization via Ground-to-Satellite Cross-View Image Retrieval, IEEE Trans. Multimedia, 2022

[CVPR 19] Motivation

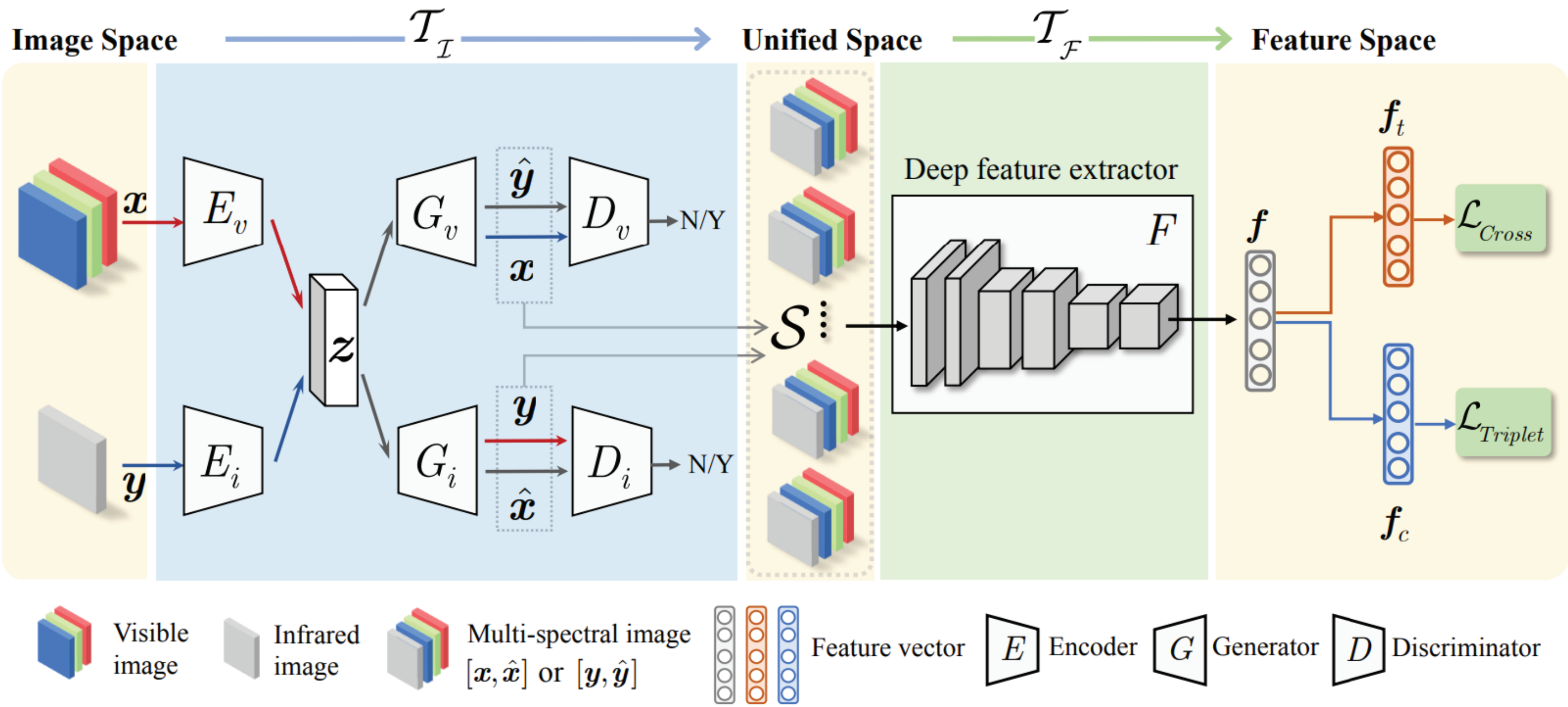






Learning to Reduce Dual-level Discrepancy for Infrared-Visible Person Re-identification, CVPR, 2019

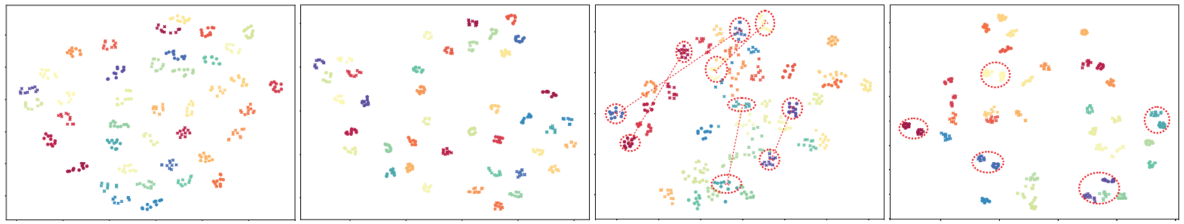
[CVPR 19] Method



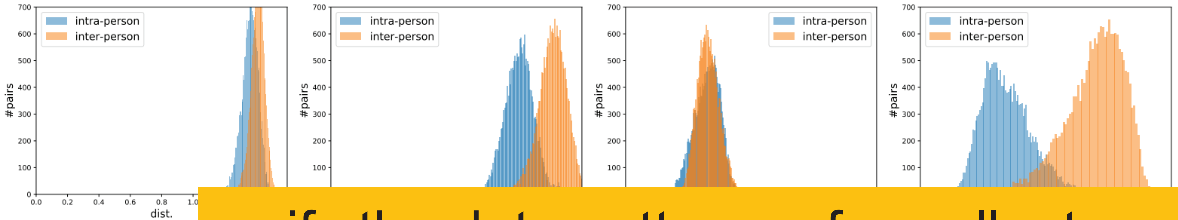
Learning to Reduce Dual-level Discrepancy for Infrared-Visible Person Re-identification, CVPR, 2019

[CVPR 19] Results

Approach	Constraints		RegDB				SYSU-MM01			
	Feature-level	Image-level	CMC-1	CMC-10	CMC-20	mAP	CMC-1	CMC-10	CMC-20	mAP
LOMO [8]	✗	✗	0.85	2.47	4.10	2.28	1.75	14.14	26.63	3.48
MLBP [9]	✗	✗	2.02	7.33	10.90	6.77	2.12	16.23	28.32	3.86
HOG [3]	✗	✗	13.49	33.22	43.66	10.31	2.76	18.25	31.91	4.24
GSM [10]	✗	✗	17.28	34.47	45.26	15.06	5.29	33.71	52.95	8.00
One-stream [21]	✓	✗	13.11	32.98	42.51	14.02	12.04	49.68	66.74	13.67
Two-stream [21]	✓	✗	12.43	30.36	40.96	13.42	11.65	47.99	65.50	12.85
Zero-Padding [21]	✓	✗	17.75	34.21	44.35	18.90	14.80	54.12	71.33	15.95
TONE [22]	✓	✗	16.87	34.03	44.10	14.92	12.52	50.72	68.60	14.42
HCML [22]	✓	✗	24.44	47.53	56.78	20.80	14.32	53.16	69.17	16.16
BDTR [23]	✓	✗	33.47	58.42	67.52	31.83	17.01	55.43	71.96	19.66
*cm*GAN [2]	✓	✗	–	–	–	–	26.97	67.51	80.56	27.80
Proposed D^{2}RL	✓	✓	**43.4**	**66.1**	**76.3**	**44.1**	**28.9**	**70.6**	**82.4**	**29.2**



(a) BDTR initial (b) BDTR best (c) Ours initial (d) Ours best



(e) BDTR initi

(a) RegDB (b) SYSU-MM01

unify the data patterns of excellent and harsh environments

Learning to Reduce Dual-level Discrepancy for Infrared-Visible Person Re-identification, CVPR, 2019

Summary

Construct Real Dataset

- rainy [IJCAI'22]
- low-light [IJCAI'20]

Use the consistent knowledge between source and target domains

- consistency [ACM MM'21]

Use the features of target domain

- diffusion [TMM'22]
- label [TIP'22]

Focus on the harsh factor in target domain

- fog factor [CVPR'22]
- light factor [TMM'20]

Unify the data status

- unify [CVPR'19]

Thank You!