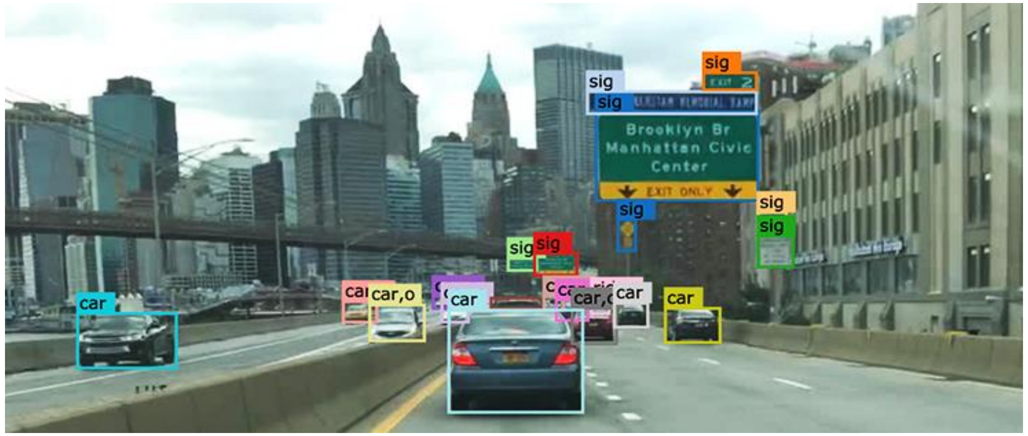
Progressive Disentanglement for Instance-Invariant Domain Adaptive Object Detection

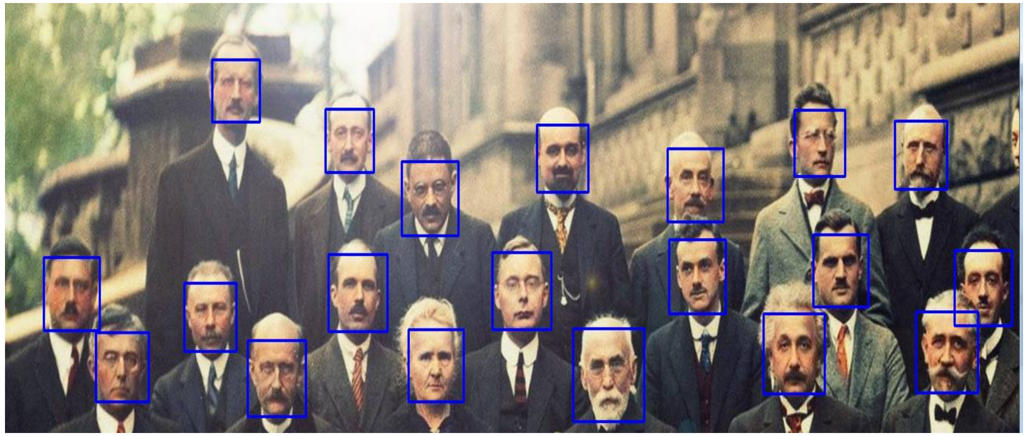
Aming Wu

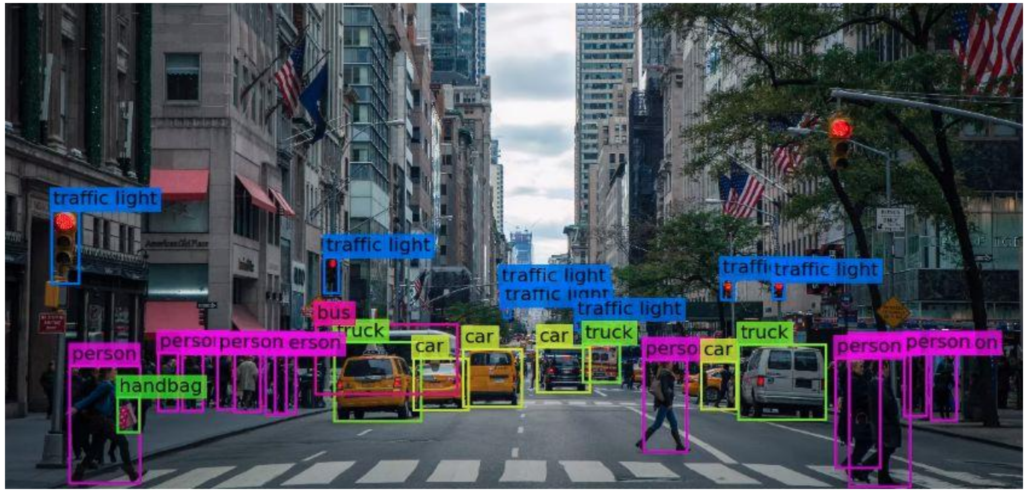
amwu@xidian.edu.cn

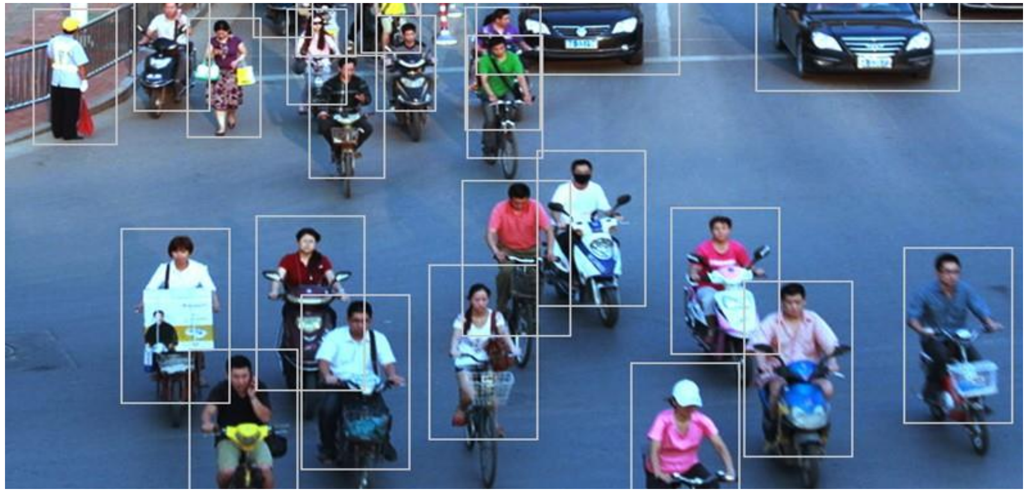
June 9th, 2021

Applications of Object Detection

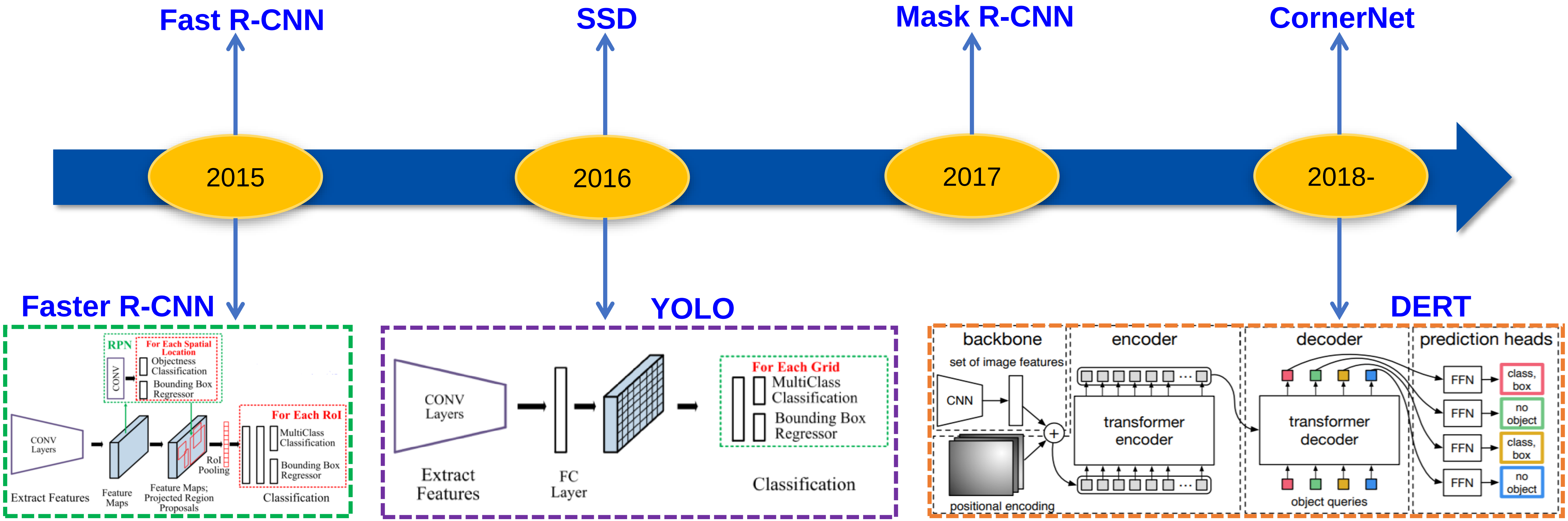








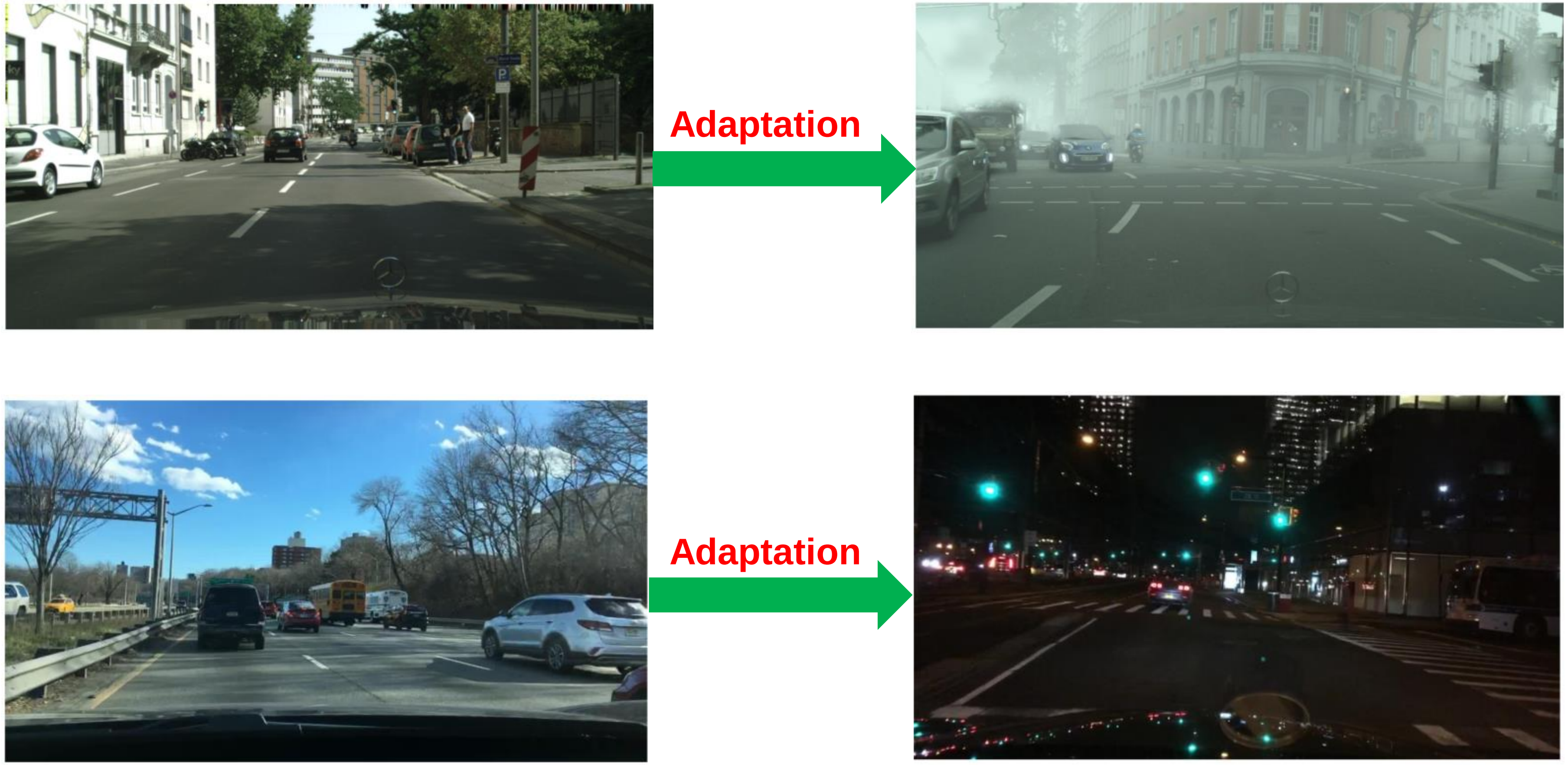
Progresses on Object Detection



[1] Ren S, He K, Girshick R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks. NeurIPS 2015
[2] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection. CVPR 2016
[3] Carion N, Massa F, Synnaeve G, et al. End-to-end object detection with transformers. ECCV 2020

Domain-Shift Impact on Object Detection

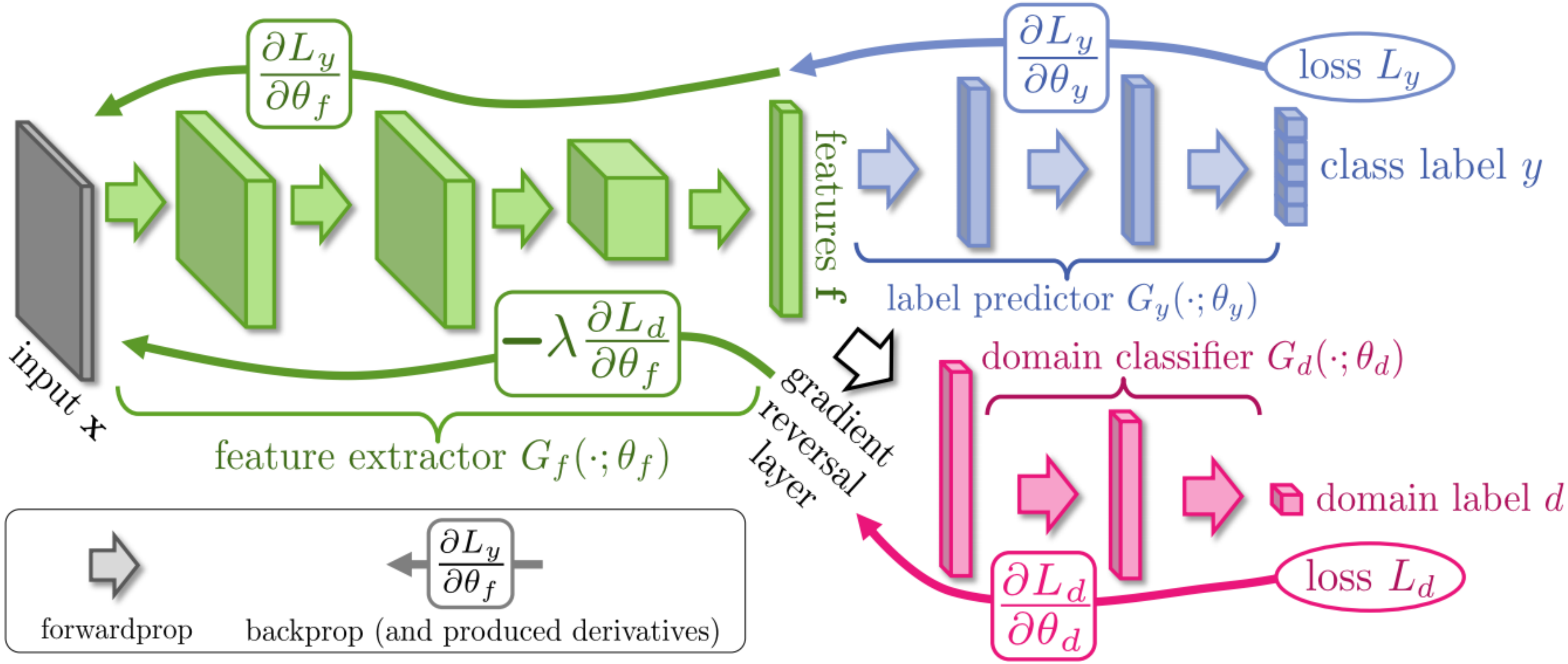
- **Training data and test data are from different domains**




BDD100K
CLEAR

Adversarial Feature Learning

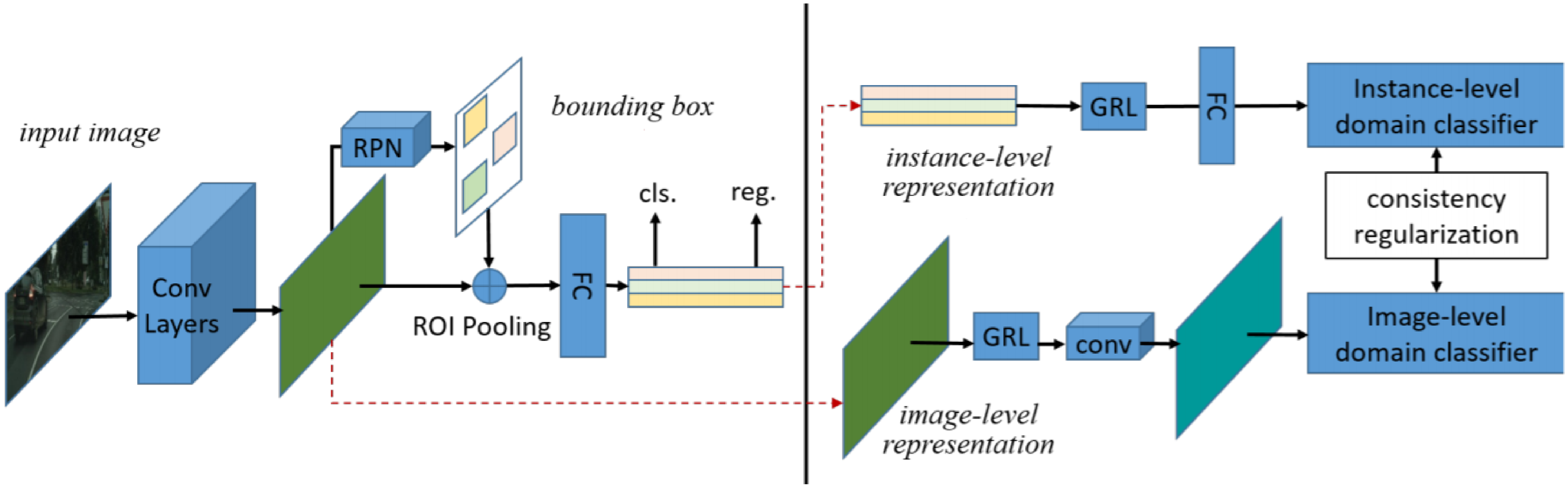
- **Feature alignment through gradient reversal**



[1] Ganin Y, Ustinova E, Ajakan H, et al. Domain-adversarial training of neural networks. The journal of machine learning research, 2016

Adversarial Feature Learning

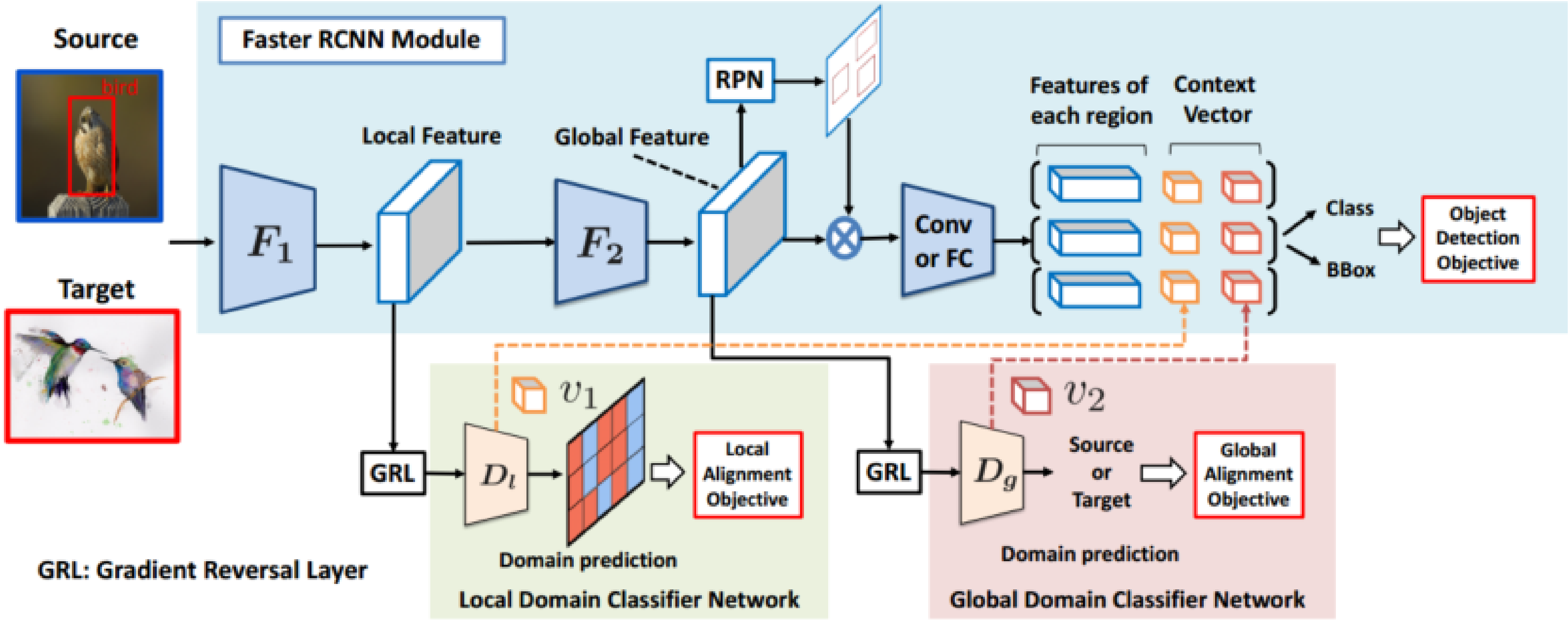
- **Domain adaptive Faster R-CNN**



[1] Chen Y, Li W, Sakaridis C, et al. Domain adaptive faster r-cnn for object detection in the wild. CVPR 2018

Adversarial Feature Learning

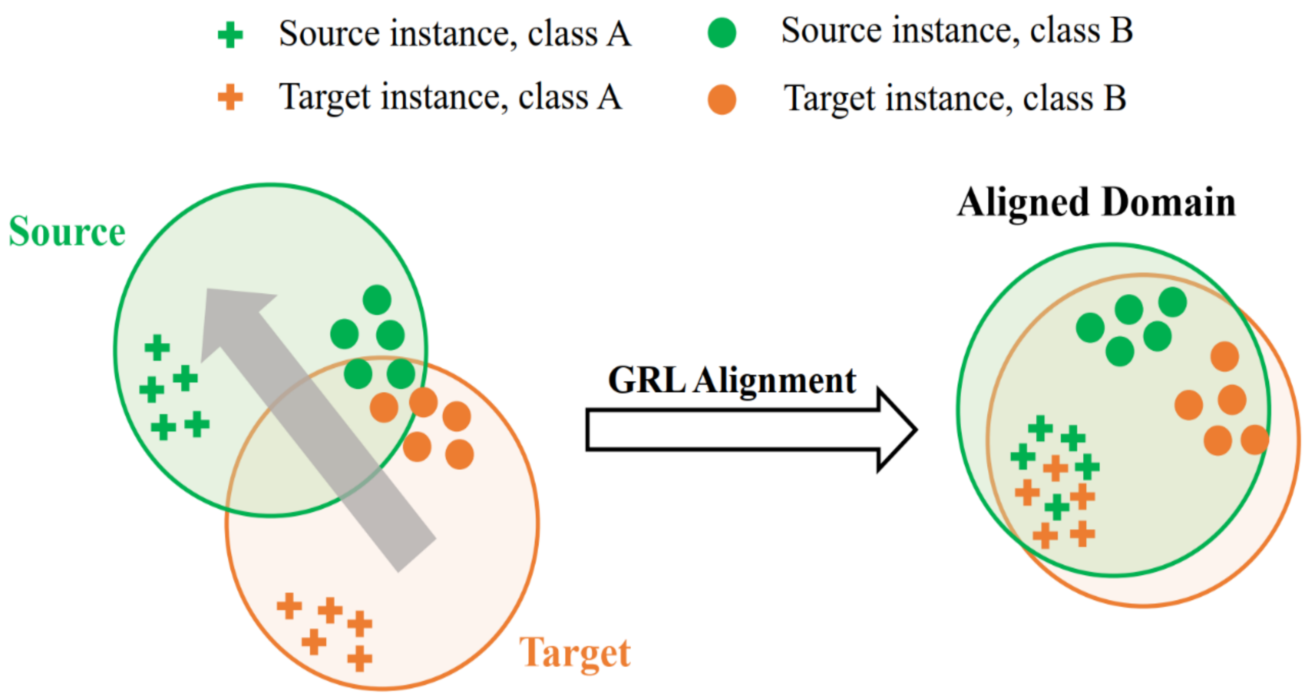
● Strong-weak distribution alignment

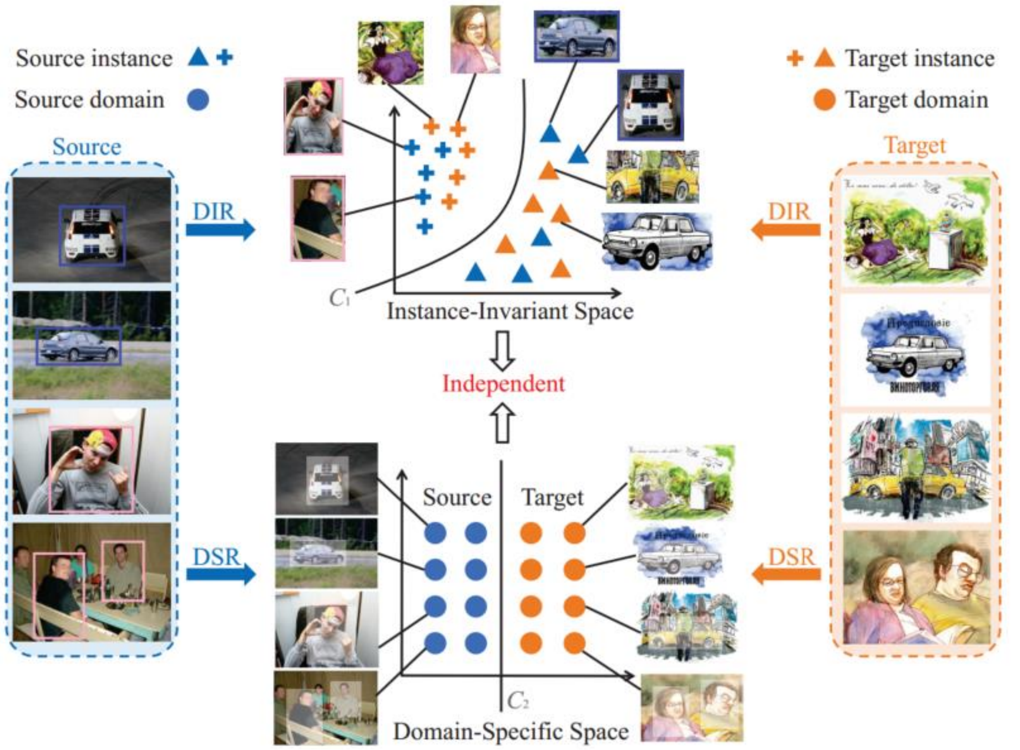


[1] Saito K, Ushiku Y, Harada T, et al. Strong-weak distribution alignment for adaptive object detection. CVPR 2019

Motivation

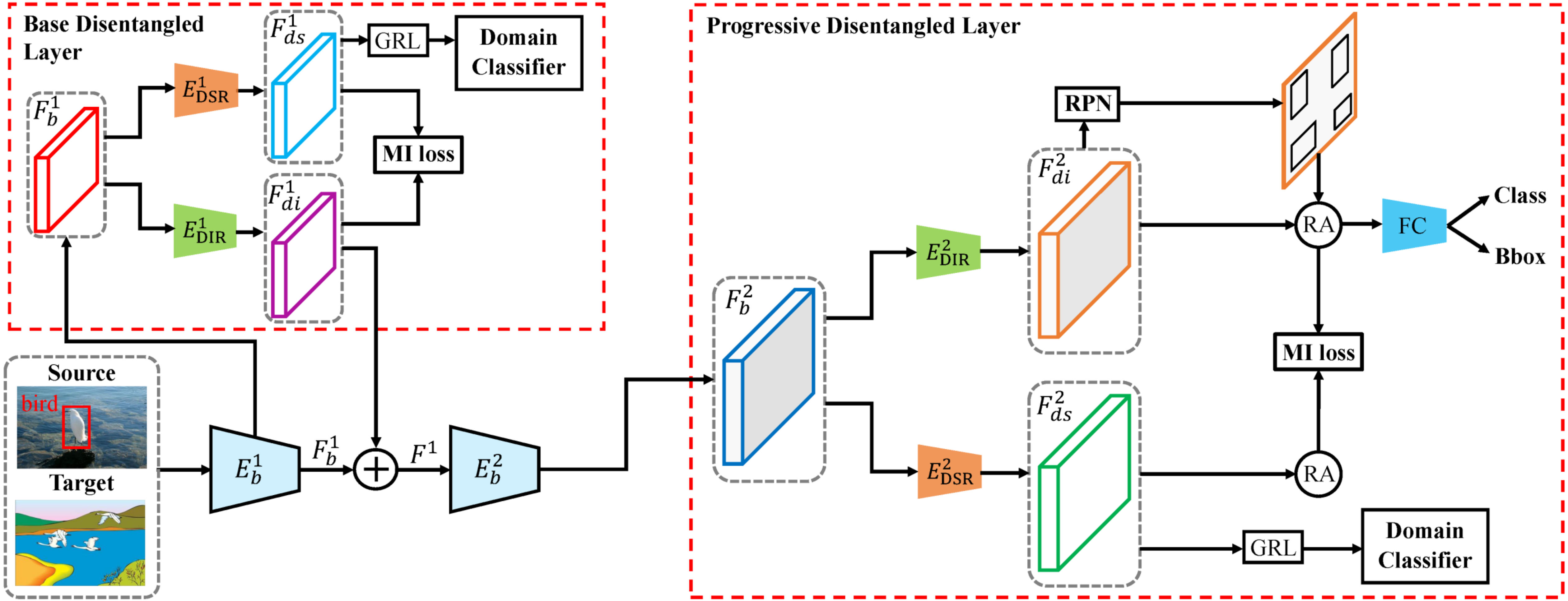
- **The domain-specific information existing in the aligned features may affect the performance**
- **Domain-invariant features play a key role in transferring detection ability**





Aming Wu, *et al*. Instance-Invariant Domain Adaptive Object Detection via Progressive Disentanglement. TPAMI 2021

Progressive Disentanglement



Progressive Disentanglement

- **The goal of the base disentangled layer is to enhance the domain-invariant information in a middle-layer feature map**

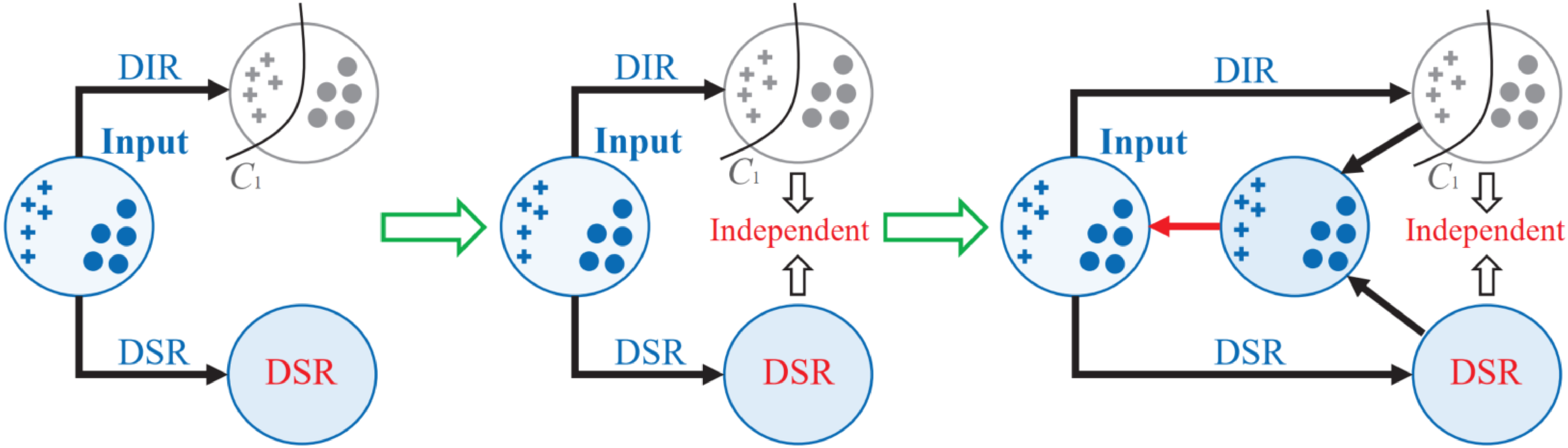
$$F_{di}^{1} = E_{\mathrm{DIR}}^{1}\left(F_{b}^{1}\right), \qquad F_{ds}^{1} = E_{\mathrm{DSR}}^{1}\left(F_{b}^{1}\right), \qquad F^{1} = F_{di}^{1} + F_{b}^{1}$$

- **The progressive disentangled layer aims to extract instance-level domain-invariant features**

$$F_{b}^{2} = E_{b}^{2}\left(F^{1}\right), \qquad F_{di}^{2} = E_{\mathrm{DIR}}^{2}\left(F_{b}^{2}\right), \qquad F_{ds}^{2} = E_{\mathrm{DSR}}^{2}\left(F_{b}^{2}\right)$$

Three-Stage Optimization

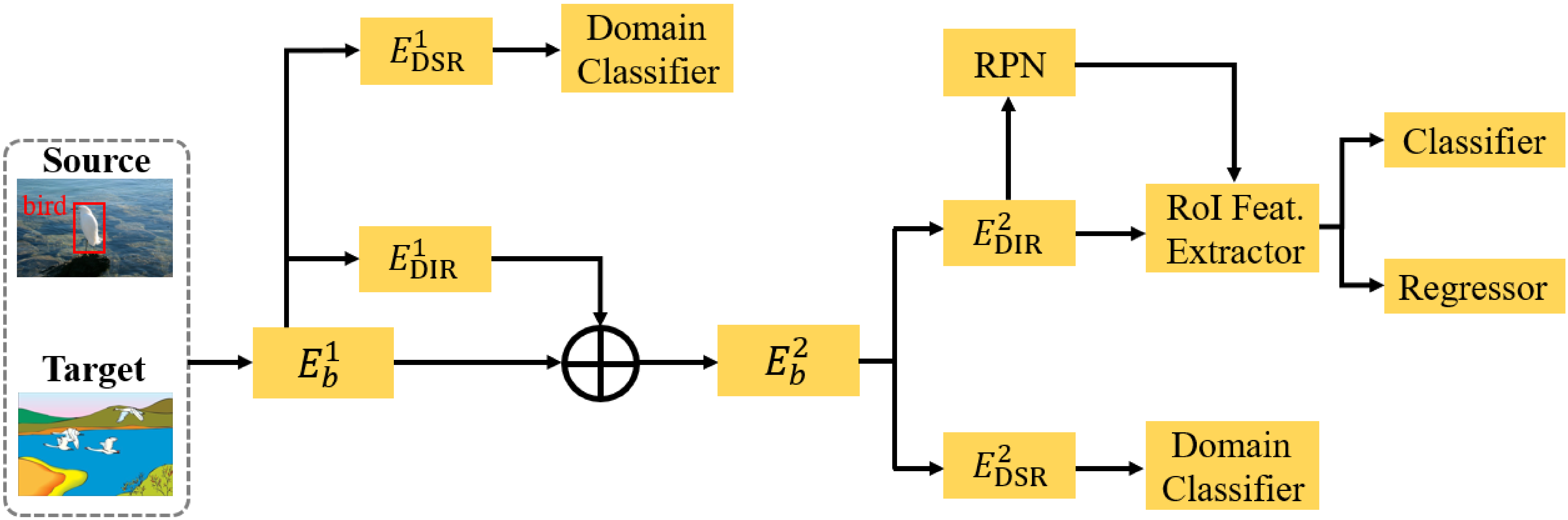
- The goal of disentangled learning is to **uncover a set of independent factors that give rise to the current observation**
- A detached optimization is devised to **breaks the disentangled process into three sequential sub-processes**



(a) Feature Decomposition (b) Feature Separation (c) Feature Reconstruction

Three-Stage Optimization

Stage I: Feature Decomposition

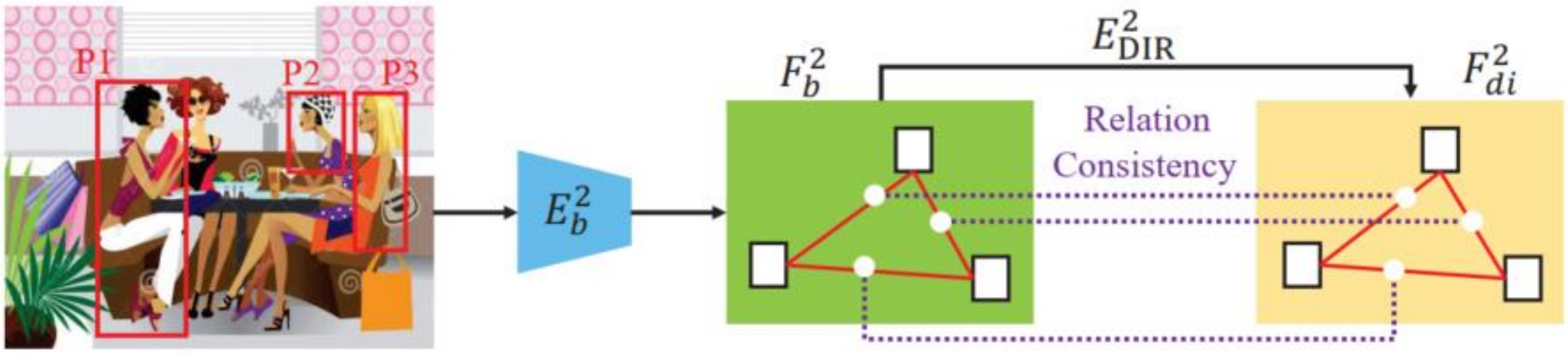


Three-Stage Optimization

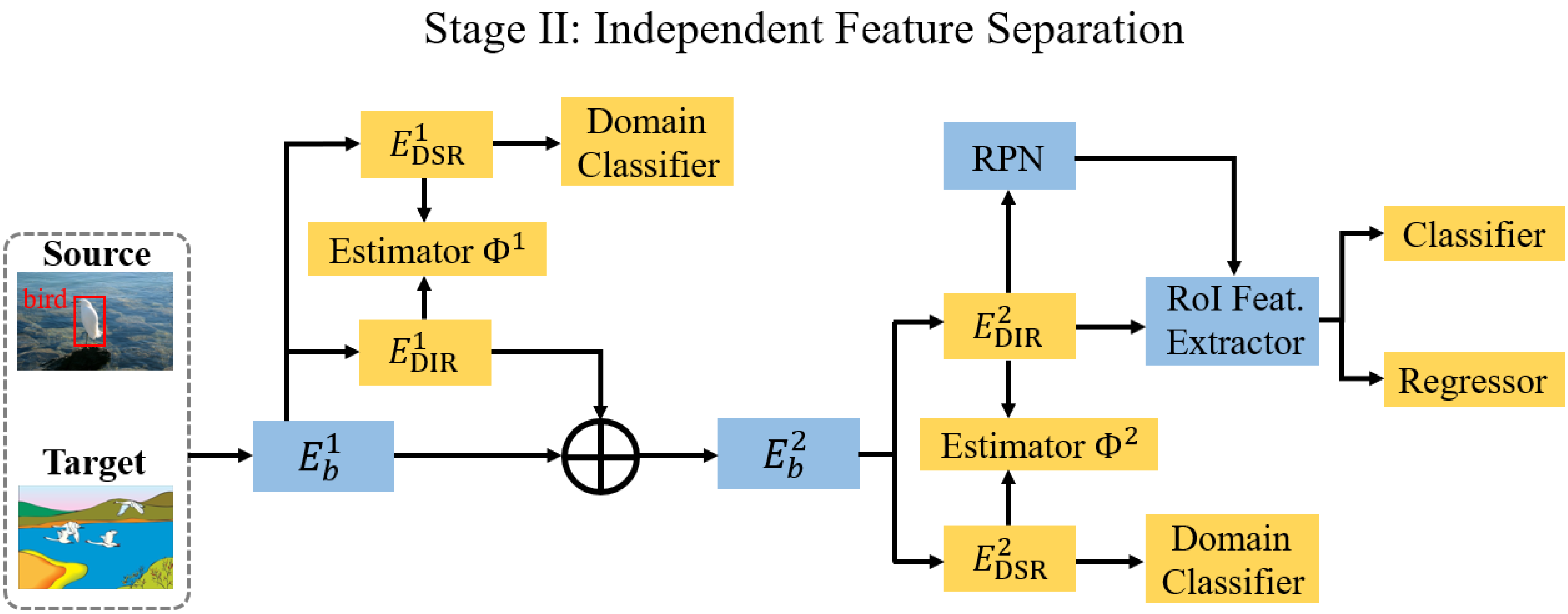
- **Mutual Information Minimization**

$$I(x,z) = \int \log \frac{d\mathbb{P}xz}{d\mathbb{P}x \otimes d\mathbb{P}z} d\mathbb{P}xz$$

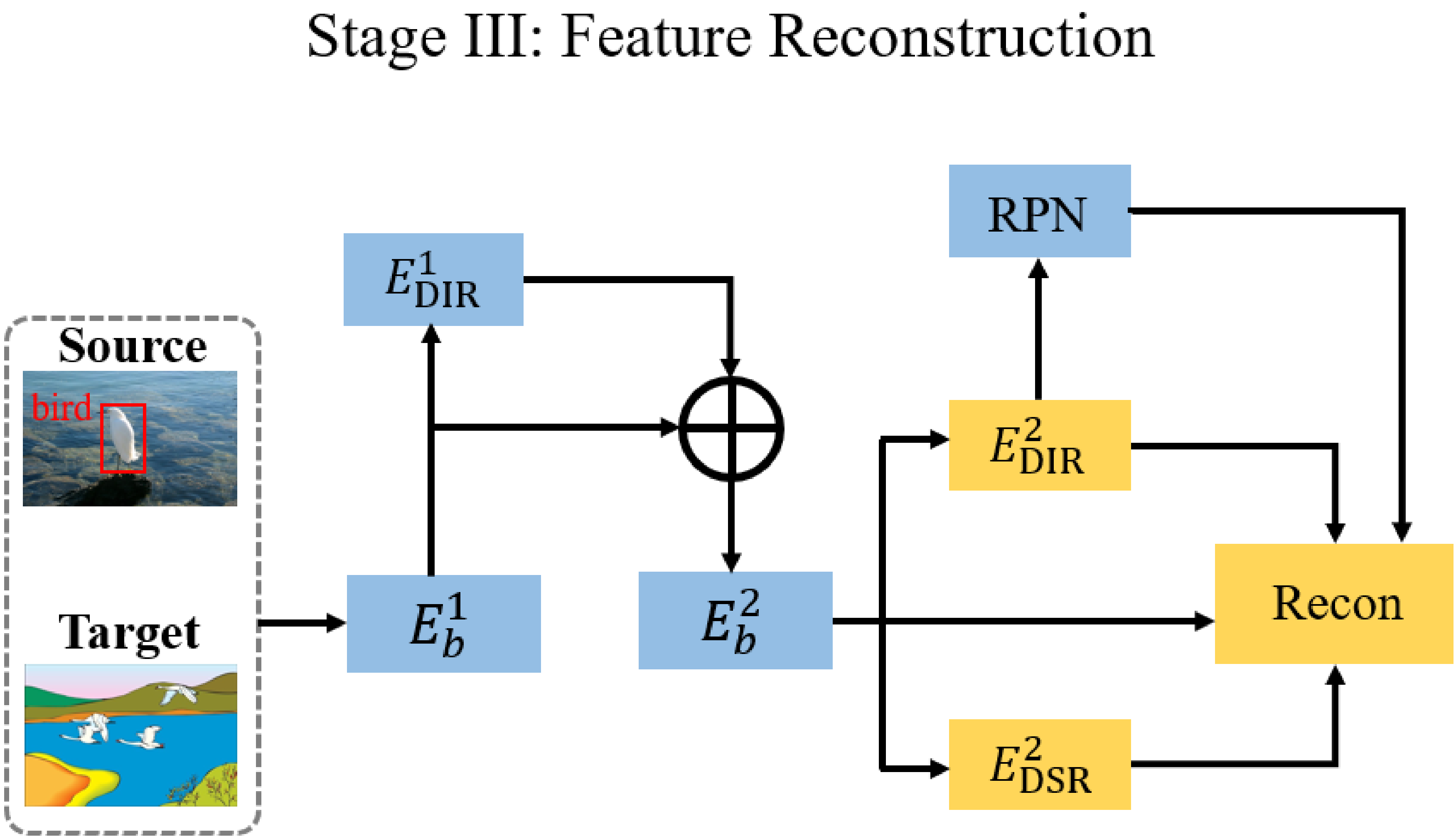
- **Relation-consistency Loss**



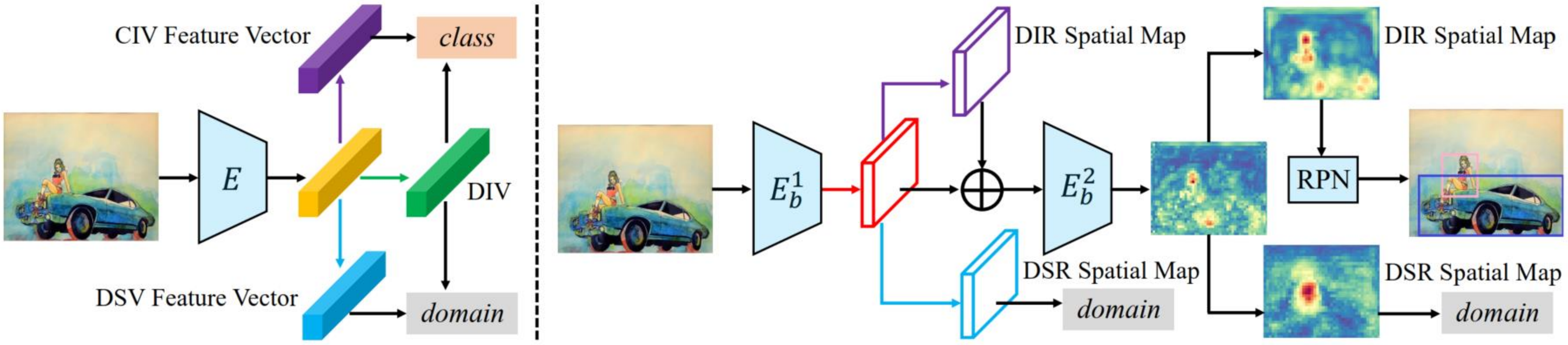
Three-Stage Optimization



Three-Stage Optimization



Comparison of Different Disentangled Methods



(a) Disentanglement on High-level Feature Vectors

(b) Disentanglement on Spatial Feature Maps

[1] Peng X, Huang Z, Sun X, et al. Domain agnostic learning with disentangled representations. ICML 2019

Experiments

Performance Analysis

Table 1: Cityscapes → FoggyCityscapes

Method	backbone	person	rider	car	truck	bus	train	motorcycle	bicycle	mAP
Source Only	VGG16	24.7	31.9	33.1	11.0	26.4	9.2	18.0	27.9	22.8
DAF [9]	VGG16	25.0	31.0	40.5	22.1	35.3	20.2	20.0	27.1	27.6
DT [30]	VGG16	25.4	39.3	42.4	24.9	40.4	23.1	25.9	30.4	31.5
SC-DA(Type3) [49]	VGG16	**33.5**	38.0	48.5	26.5	39.0	23.3	28.0	33.6	33.8
DMRL [14]	VGG16	30.8	40.5	44.3	27.2	38.4	**34.5**	28.4	32.2	34.6
MLDA [52]	VGG16	33.2	**44.2**	44.8	**28.2**	41.8	28.7	**30.5**	36.5	36.0
FSDA [53]	VGG16	29.1	39.7	42.9	20.8	37.4	24.1	26.5	29.9	31.3
MAF [11]	VGG16	28.2	39.5	43.9	23.8	39.9	33.3	29.2	33.9	34.0
SW (B) [10]	VGG16	29.9	42.3	43.5	24.5	36.2	32.6	30.0	35.3	34.3
Ours	VGG16	33.12	43.41	**49.63**	21.98	**45.75**	32.04	29.59	**37.08**	**36.57**
Ours	ResNet101	32.82	44.37	49.57	33.02	46.10	37.97	29.90	35.26	38.63

Experiments

Performance Analysis

Table 2: Daytime → Nighttime

Method	bus	bike	car	motor	person	rider	truck	mAP
Source Only	21.1	19.6	**45.8**	12.9	22.4	22.3	31.6	25.1
DAF [9]	24.2	**21.7**	45.3	7.2	23.3	25.5	32.1	25.6
SW (B) [10]	26.0	18.3	44.1	10.4	23.0	**25.9**	33.4	25.9
Ours	**33.1**	19.5	44.2	**13.2**	**24.7**	24.9	**35.5**	**27.9**

Table 3: VOC → Watercolor

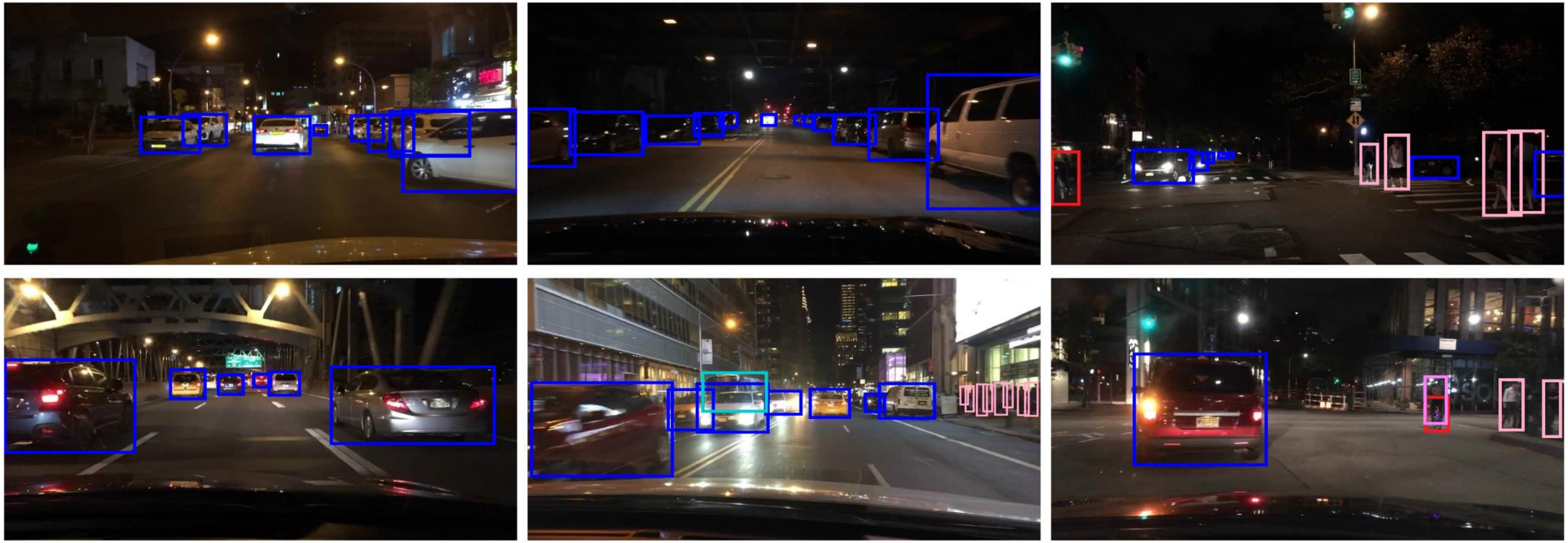
Method	bike	bird	car	cat	dog	person	mAP
Source Only	68.8	46.8	37.2	32.7	21.3	60.7	44.6
BDC-Faster [10]	68.6	48.3	47.2	26.5	21.7	60.5	45.5
DAF [9]	75.2	40.6	48.0	31.5	20.6	60.0	46.0
SW (B) [10]	82.3	**55.9**	46.5	32.7	**35.5**	**66.7**	53.3
Ours	**95.8**	54.3	**48.3**	**42.4**	35.1	65.8	**56.9**

Table 4: VOC → Clipart

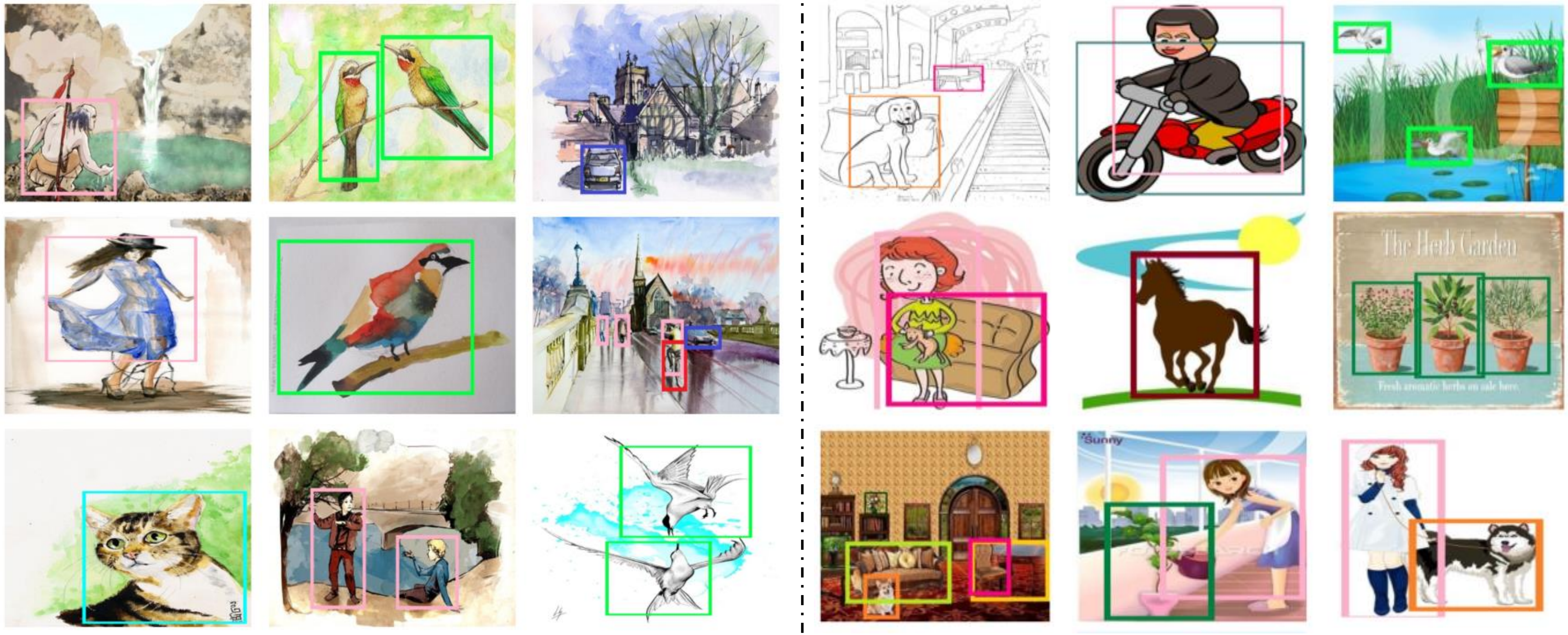
Method	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	sheep	sofa	train	tv	mAP
Source Only	35.6	52.5	24.3	23.0	20.0	43.9	32.8	10.7	30.6	11.7	13.8	6.0	36.8	45.9	48.7	41.9	**16.5**	7.3	22.9	32.0	27.8
BDC-Faster [10]	20.2	46.4	20.4	19.3	18.7	41.3	26.5	6.4	33.2	11.7	26.0	1.7	36.6	41.5	37.7	44.5	10.6	20.4	33.3	15.5	25.6
DAF [9]	15.0	34.6	12.4	11.9	19.8	21.1	23.2	3.1	22.1	26.3	10.0	10.0	19.6	39.4	34.6	29.3	1.0	17.1	19.7	24.8	19.8
SW (B) [10]	26.2	48.5	32.6	**33.7**	38.5	54.3	37.1	**18.6**	34.8	**58.3**	17.0	12.5	33.8	65.5	61.6	**52.0**	9.3	24.9	54.1	**49.1**	38.1
Ours	**41.5**	**52.7**	**34.5**	28.1	**43.7**	**58.5**	**41.8**	15.3	**40.1**	54.4	**26.7**	**28.5**	**37.7**	**75.4**	**63.7**	48.7	**16.5**	**30.8**	**54.5**	48.7	**42.1**

Experiments

Experiments



Experiments



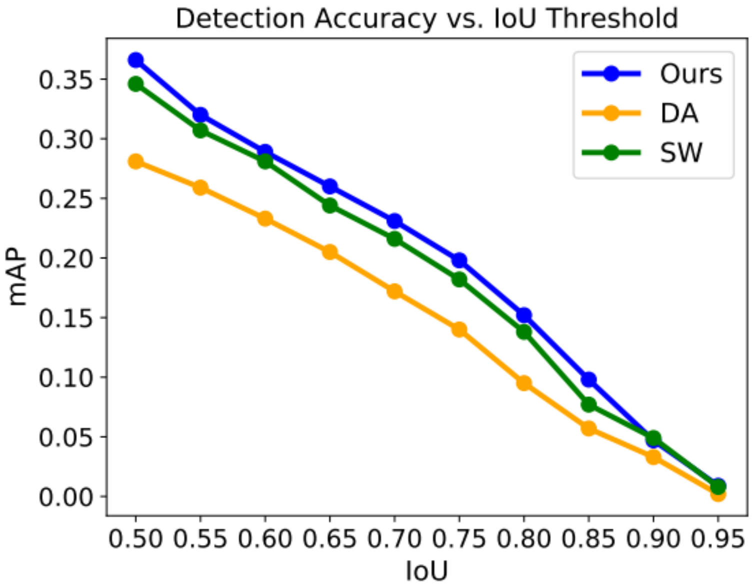
Experiments

- **Ablation Analysis**

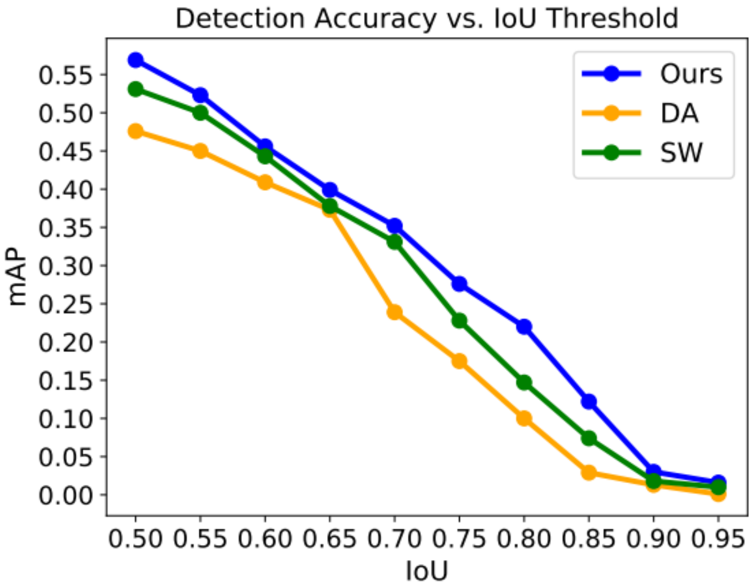
Method	OW	1st	2nd	3rd	RC	C → F	V → W
BP layer	✓				✓	34.1%	52.9%
BP layer		✓				33.6%	53.5%
BP layer			✓		✓	35.3%	55.3%
BP layer				✓		35.5%	55.2%
BP layer				✓	✓	**36.6%**	**56.9%**
SD layer	✓				✓	31.9%	50.3%
SD layer		✓				32.1%	51.6%
SD layer			✓		✓	33.6%	53.1%
SD layer				✓		33.5%	53.5%
SD layer				✓	✓	**34.1%**	**54.6%**

Experiments

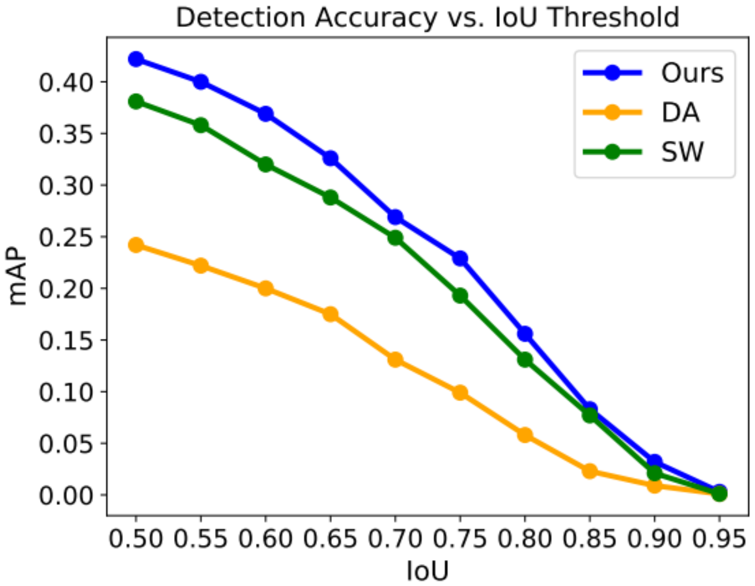
Ablation Analysis



(a) from Cityscapes to FoggyCityscapes



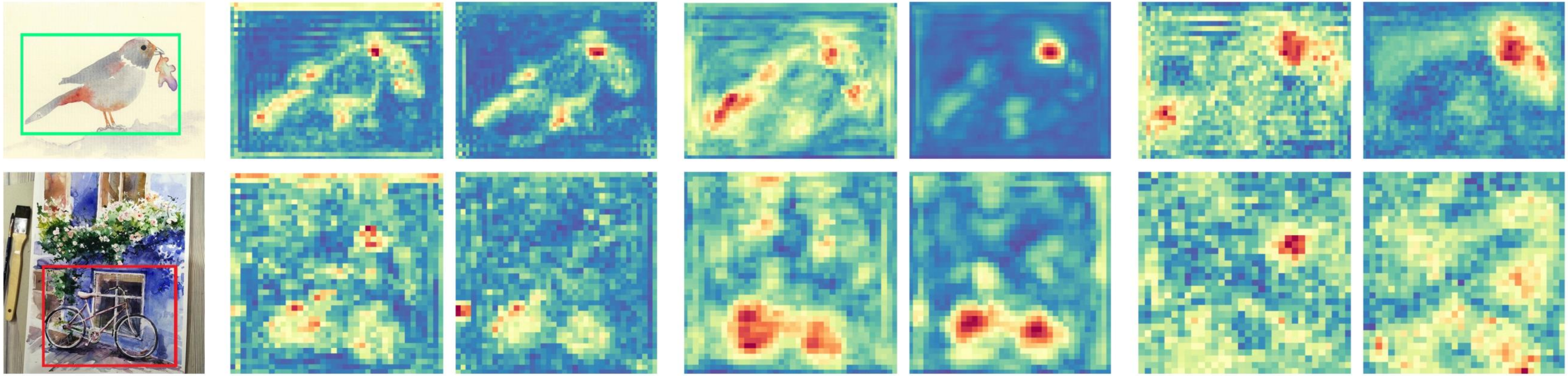
(b) from PASCAL VOC to Watercolor



(c) from PASCAL VOC to Clipart

Experiments

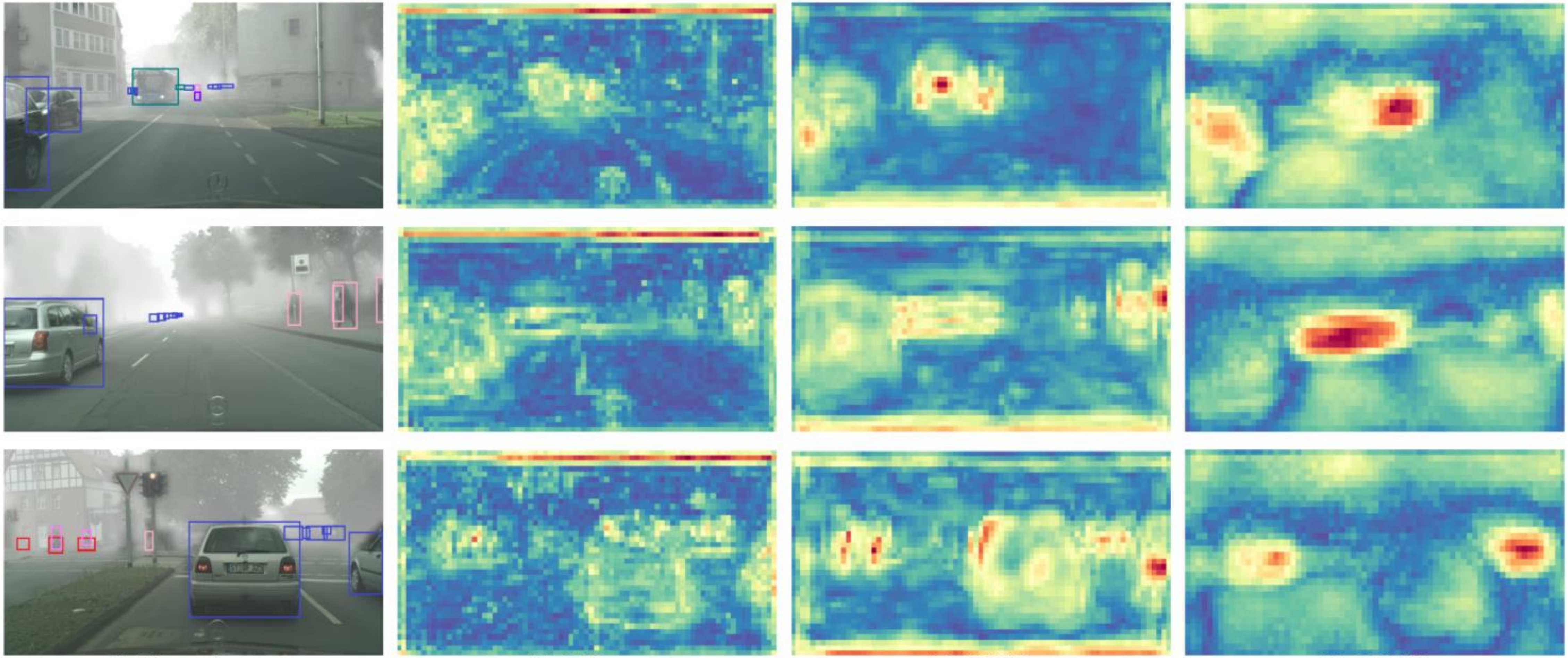
- **Visualization Analysis**



(a) GT (b) SD-Base (c) BP-Base (d) SD-DIR (e) BP-DIR (f) SD-DSR (g) BP-DSR

Experiments

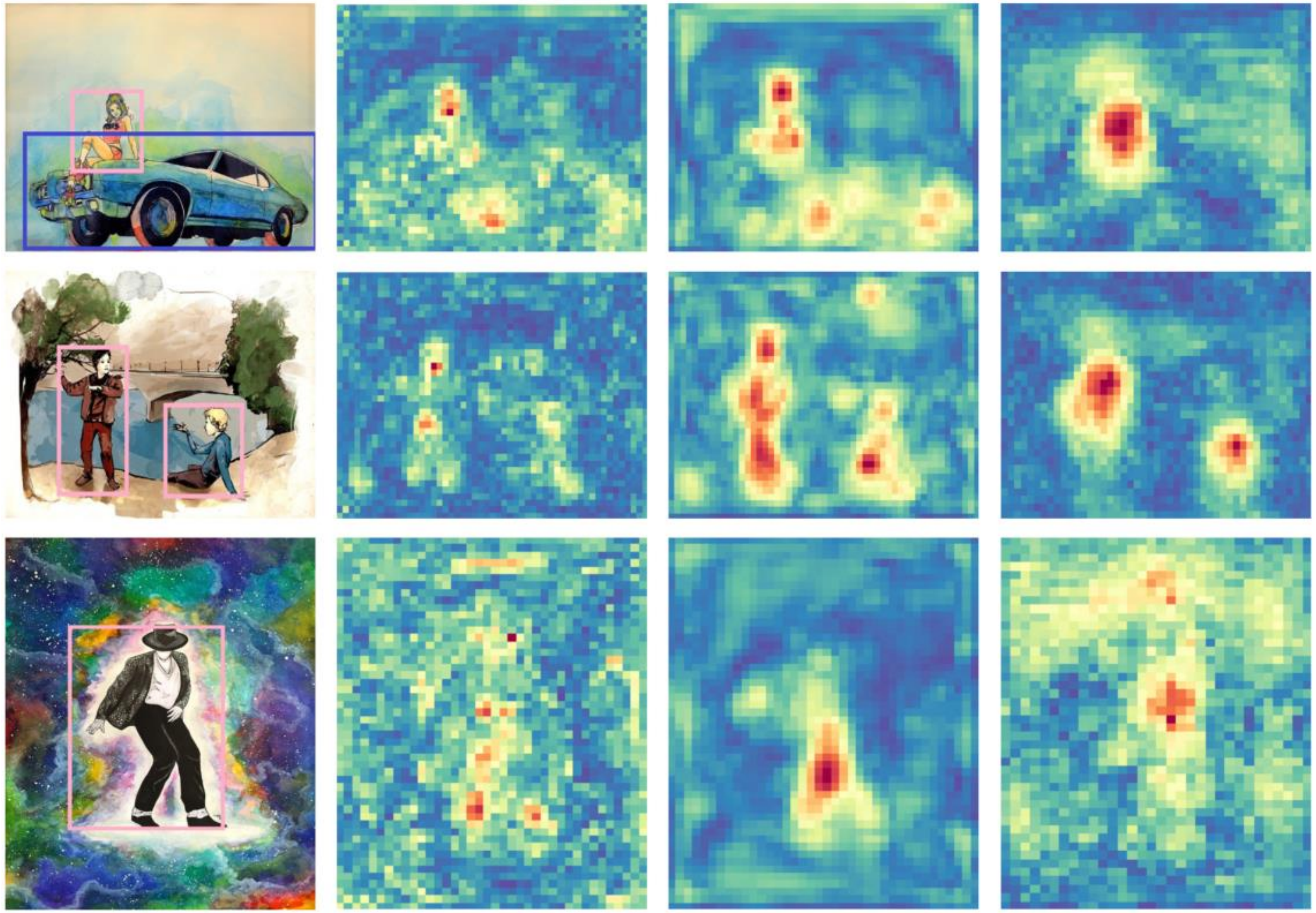
Visualization Analysis



(a) Foggy image (b) P-Base (c) P-DIR (d) P-DSR

Experiments

- **Visualization Analysis**



Thanks for your attention!