Context-Aware Mobile Visual Analysis

Chen Tao

Fudan University, Shanghai, China

Jun 20, 2021

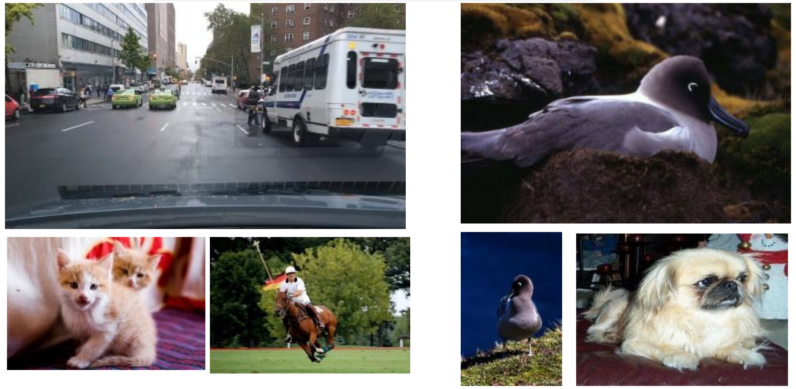
目录

一、Background

Significance and Challenges

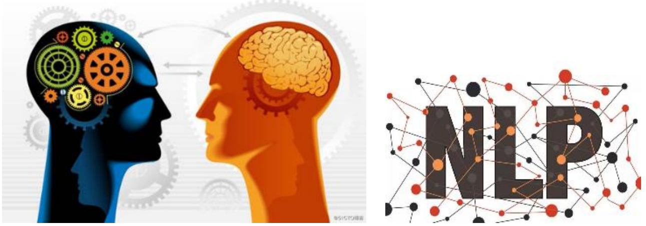
➢ Deep Learning Success

Image & Video



Speech & Audio

Text & Language



➢ Factors:

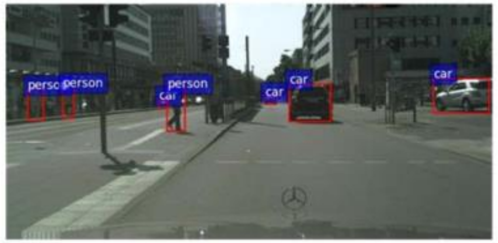
- ✔ Computation Resources： multiple GPUs for training
- ✔ Large Data： multi-class， multi-granularity， multi-scenario
- ✔ Simple Context： train/test under similar contexts with little changes

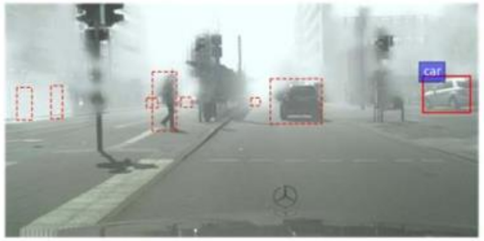
Significance and Challenges

➢Challenges for Mobile Vision：

Limited Resource

Changing Context and Scenes



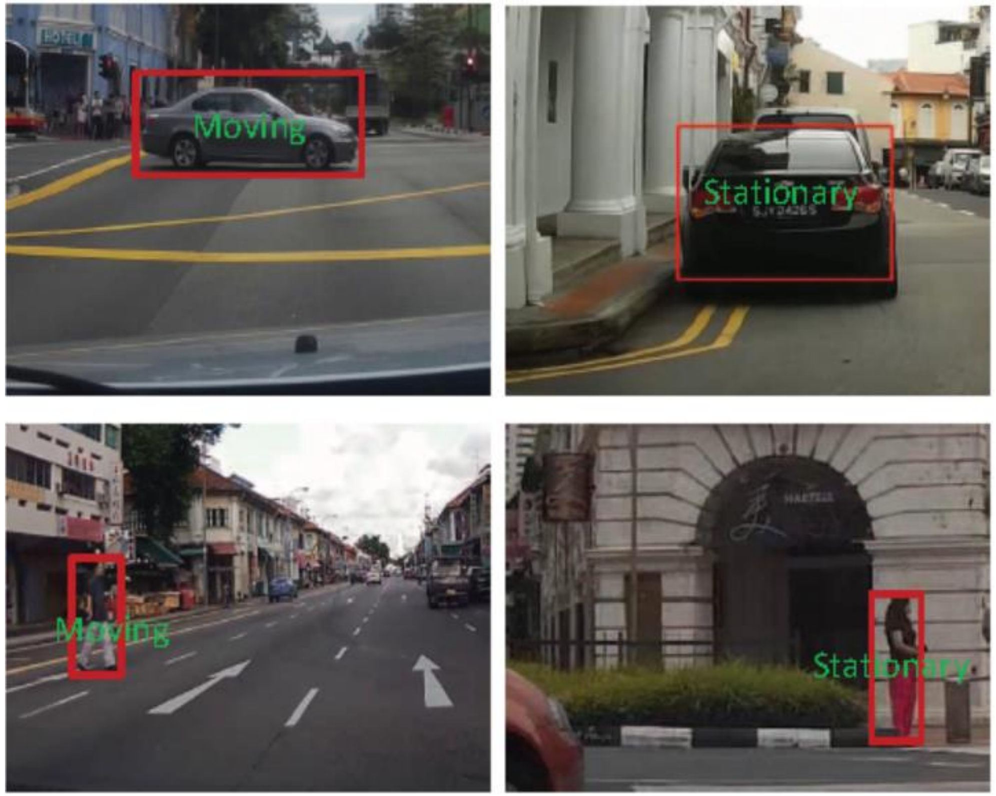


二、Methods

Context-aware Object Motion Estimation

➢Problem Definition:

- Detect an object from a moving camera, then determine its motion status: still or moving



➢Challenges:

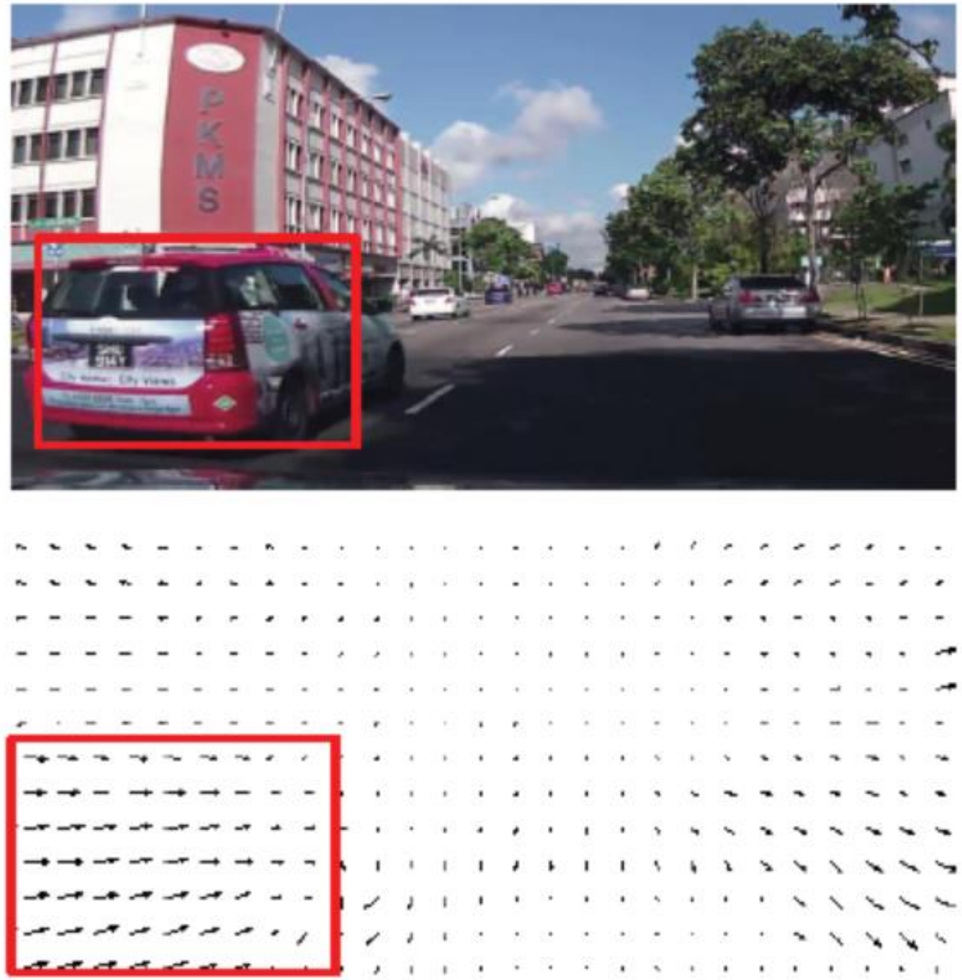
- Existing object detection works: work on a single image and cannot provide the object motion information.
- Existing motion detection techniques cannot provide either category or number of objects within the moving regions.
- All current motion detection works cannot detect stationary objects which are treated as the background.

Context-aware Object Motion Estimation

➢Our Idea: Optical Flow Deviation:

(a)

(b)



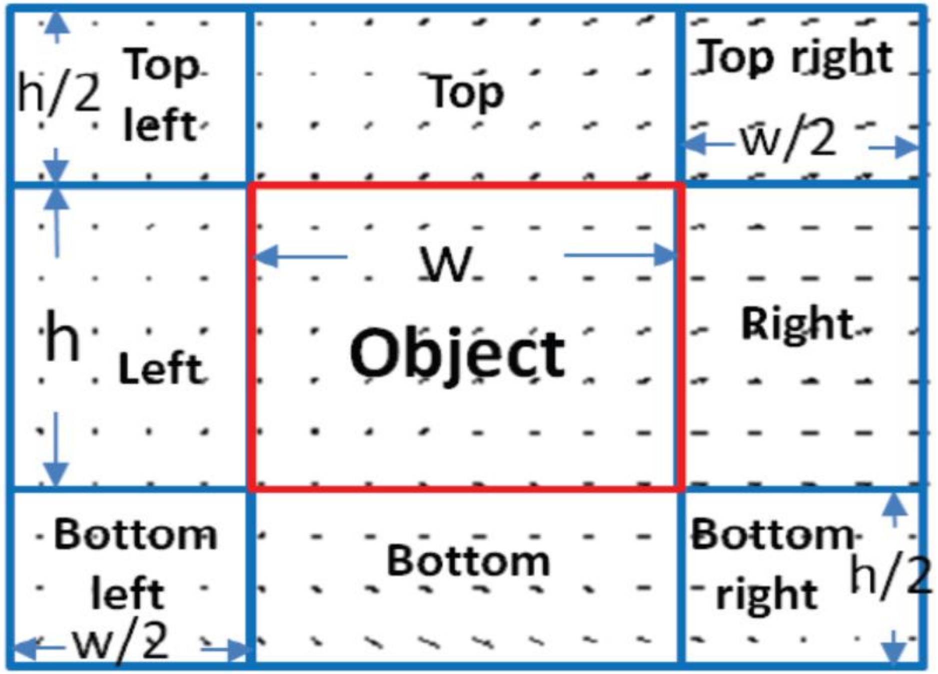
➢Context-aware Motion Descriptor:

- An object-level motion descriptor (CMD) is designed to represent the object motion behavior.

- CMD utilizes the surrounding contexts of the target object within a video frame spatially and two consecutive video frames as captured by a moving camera temporally.

➢CMD:

- Step 1: Context-Aware Histogram of Oriented Optical Flow



In each rectangle, each flow vector is binned according to its primary angle from the horizontal axis and weighted according to its magnitude.

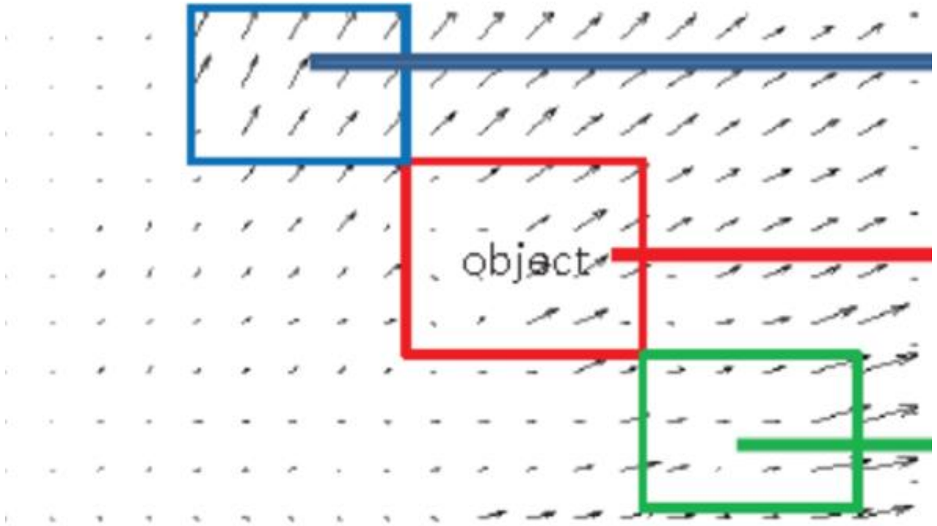
For an optical flow vector v = (x; y), its orientation is defined by

$$\theta = \begin{cases} \cos^{-1}\left(\frac{x}{(x^2+y^2)^{1/2}}\right), & \text{if } y \geq 0 \\ 2\pi - \cos^{-1}\left(\frac{x}{(x^2+y^2)^{1/2}}\right), & \text{otherwise} \end{cases}$$

➢CMD:

- Step 2: Measure the flow field inconsistency.
- A straightforward way:

$$u_b(i) = \begin{cases} 1, & \text{if } f^o(b) \geq f^i(b) \\ 0, & \text{otherwise} \end{cases}$$



Moving: 1: [11111111] or 0: [00000000]

Stationary: [00001111] or [11110000]

Problem: some stationary objects may also produce all 1:[11111111] binary vectors due to their radial direction deviations

➢CMD：

- Step 2: Measure the flow field inconsistency.

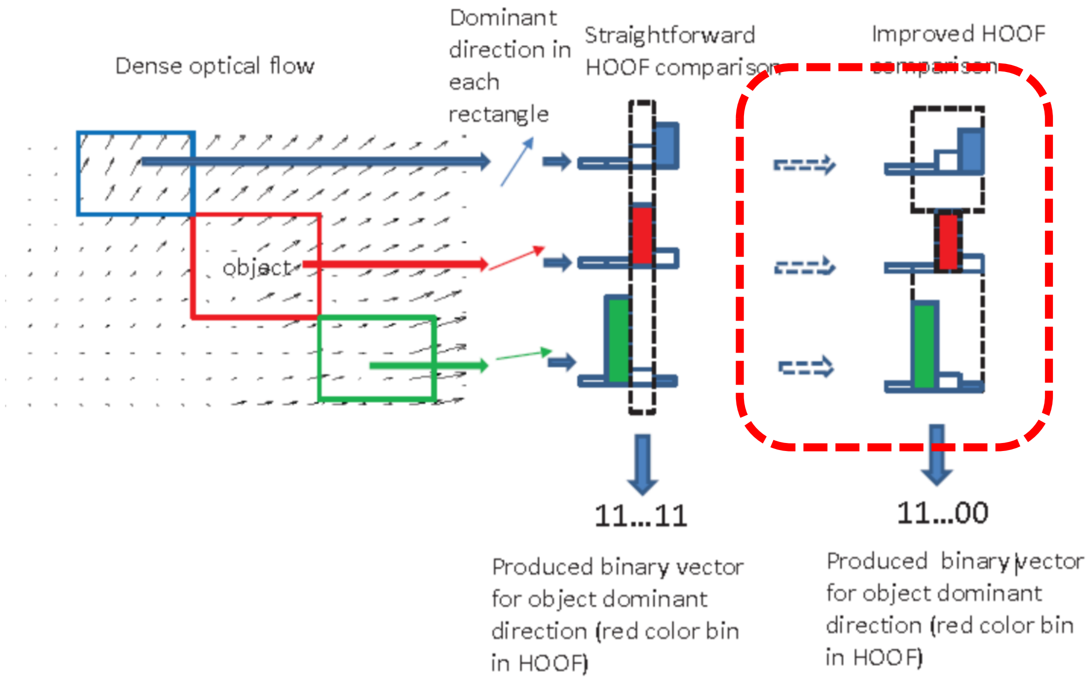
$$u_b(i) = \begin{cases} 1, & \text{if} f^o(\mathrm{b}) \geq \boxed{K\left(f^i(\mathrm{b})\right)} \\ 0, & \text{otherwise} \end{cases}$$

$$K\left(f^i(\mathrm{b})\right) = \max_{b'} f^i(\mathrm{b}')$$

where $b' = b - \delta, ..., b-1, b, ..., b+\delta$ and δ is subject to

$$\Delta\theta = 2\pi\frac{b+\delta}{B} - 2\pi\frac{b-\delta-1}{B} = 2\pi\frac{2\delta+1}{B} <= \frac{\pi}{3}$$

- Step 2: Orientation-Wise Soft Margin Operator



➢CMD:

- Step 3: CMD Construction
- Existing:

$$c(u_b + 1) = c(u_b + 1) + \boxed{1}$$

Original flow vector distribution information residing in the object HOOF is missed

- Proposed:

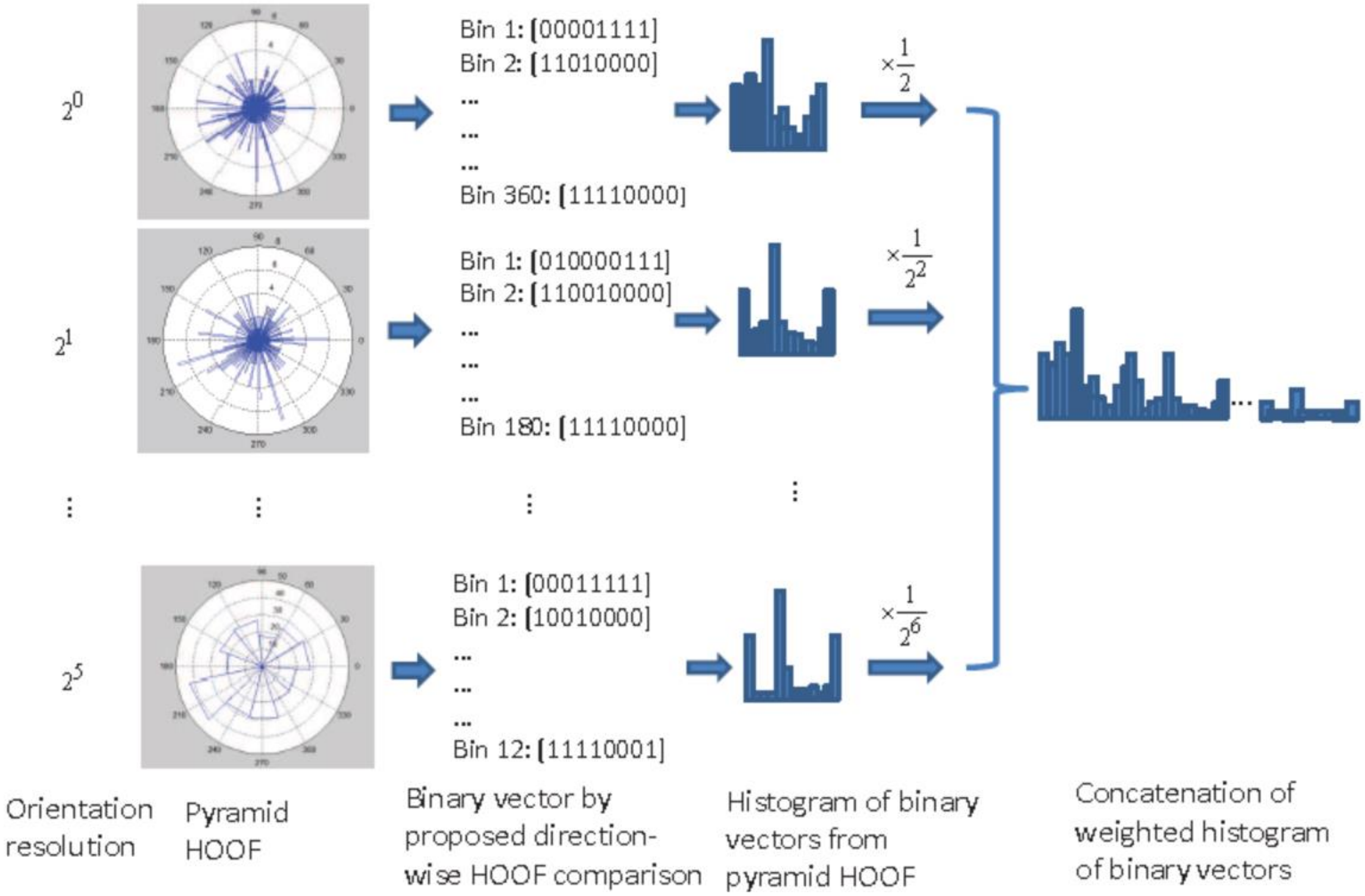
$$c(u_b + 1) = c(u_b + 1) + \boxed{f^o(b)}$$

Normalized flow speed (magnitude) value added

➢CMD：

- Step 4:

Concatenation histogram of binary vectors from pyramid HOOF.



➢Evaluation:

- Data: 23 video clips, each clip is 3 to 5 minutes with a frame rate of 30 fps.

- Resolution is 1200*900 pixels.

➢Results:

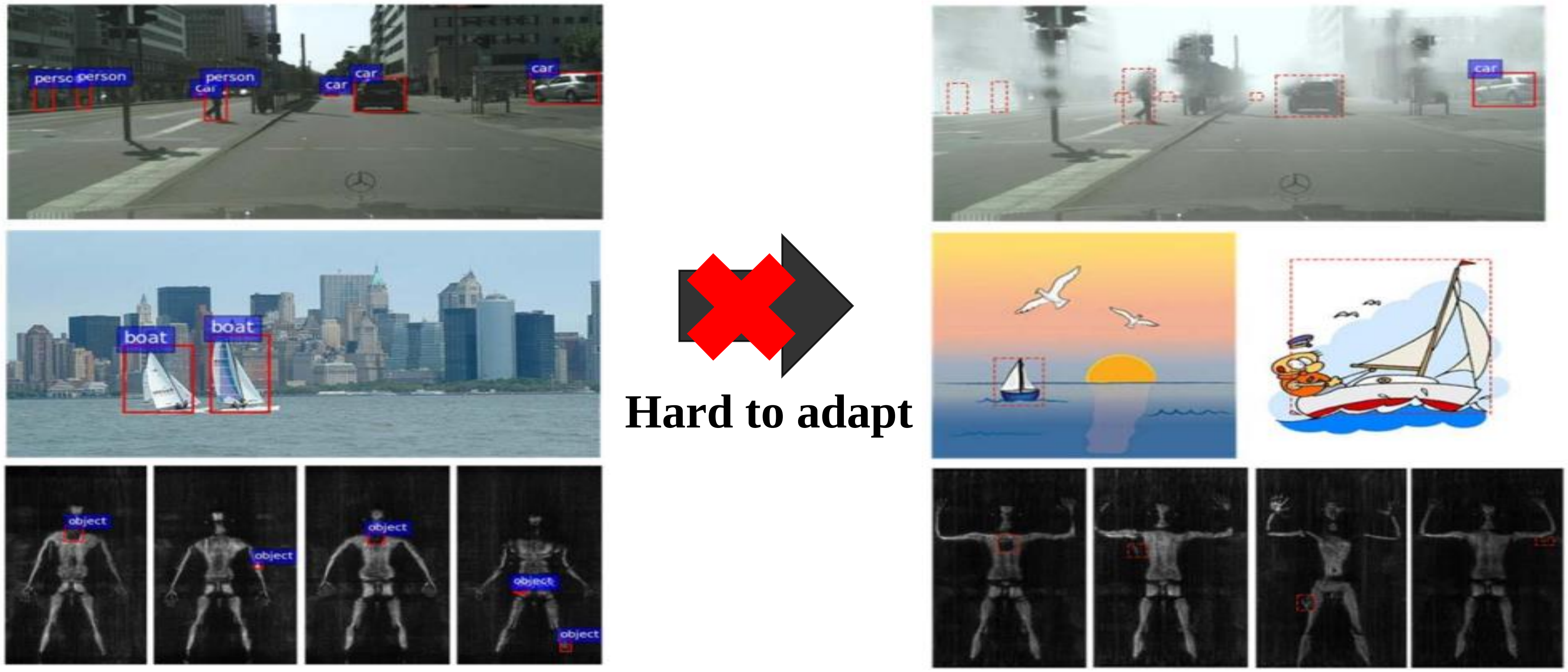
Image background	Accuracy
Little motion	0.93
Partial motion	0.90
Dense motion	0.84

Vehicle scales (pixels)	Accuracy
$[300, \infty)$	0.89
$[100, 300)$	0.92
$[30, 100)$	0.90

Camera-vehicle relative speed	Accuracy
$S_{vehicle} = 0$	0.91
$S_{vehicle} < S_{camera}$	0.87
$S_{vehicle} > S_{camera}$	0.93
$S_{vehicle} \approx S_{camera}$	0.89
$O_{vehicle}! = O_{camera}$	0.92

➢Challenges of Domain Adaptive Object Detection：

Changing Context



Context-aware Domain Adaptive Object Detection

➢Unsupervised domain adaptation for object detection：

1）Labeled Source Images

2）Unlabeled Target Images

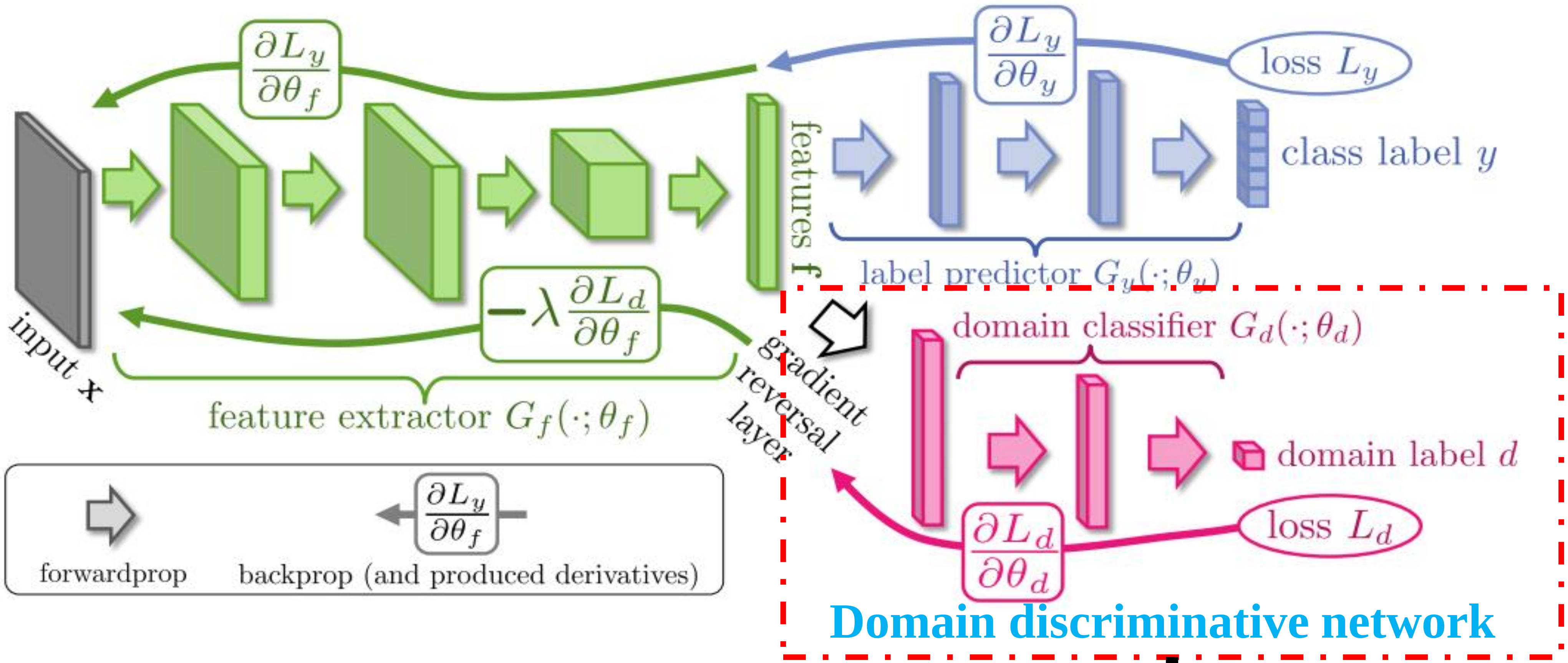
→

Predict the objects with domain shifts

➢Limitations of existing works：

- Focus on improving the domain adaptability of region-based detector family.
- Aligning the instance-level features between domains by means of RPN module.
- For region-free detector family having no RPN module, these works cannot successfully align the instance-level features between domains.
- Performing the cross-domain adaptation without considering features from different semantic levels and scales.

➢Adversarial learning for domain adaptation：

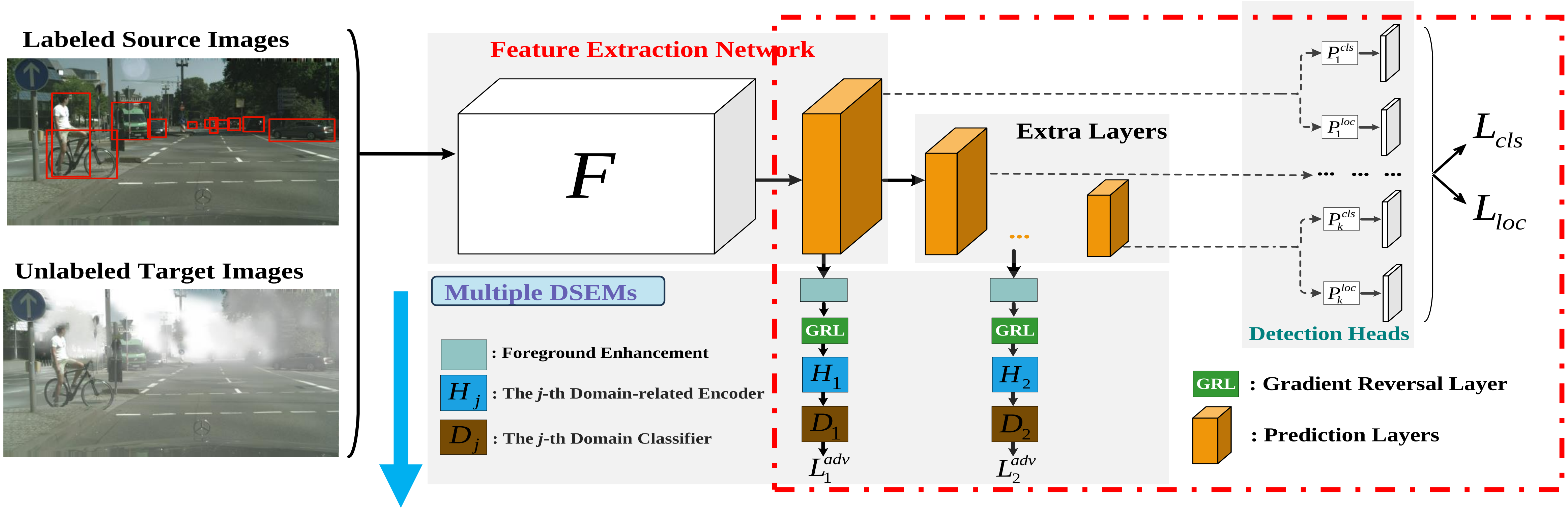


However, such a domain discriminative network cannot encode the domain-invariant information from multi-level semantics

Ganin, Y., and Lempitsky, V. Unsupervised domain adaptation by backpropagation. In ICML 2016.

Context-aware Domain Adaptive Object Detection

➢The proposed method:

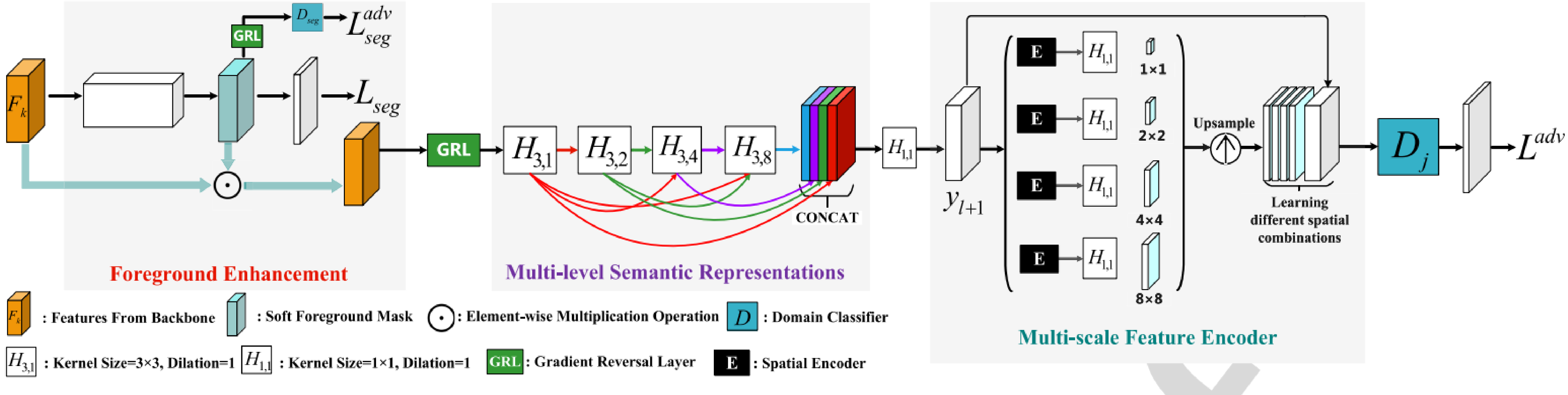


We propose a densely semantic enhancement module (DSEM), which can be easily inserted into different region-free detectors such as SSD, RefineDet, to enhancement the cross-domain detection accuracy for the target domain.

Context-aware Domain Adaptive Object Detection

➢DSEM (Act as the Domain discriminative network):

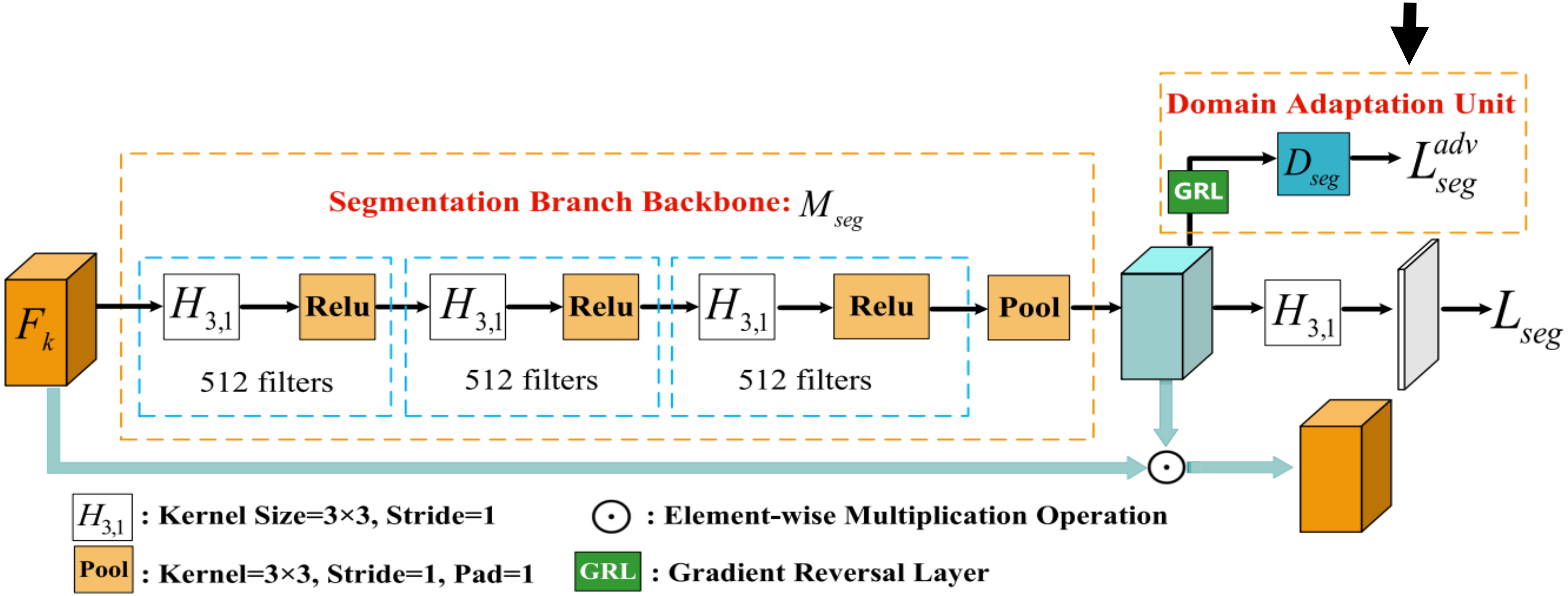
Foreground Enhancement → **Densely-connected Encoding** → **Multi-scale Feature Encoder** → **Adversarial Learning**



The domain discriminative network is endowed with the ability of encoding multi-level semantics and multi-scale features.

➢Overview of foreground enhancement:

Segmentation mask from the target domain is not available



➢Total Loss Function :

$$\min_{F,P} L_{det}(F,P)$$ ⬅ Source Domain Detection Loss

$$\max_{D_{seg}} \min_{S} L_{seg}(S) - L_{seg}^{adv}(D_{seg})$$ ⬅ Adaptive Foreground Augmentation Loss

$$\max_{H,D} \min_{F,P} L_{det}(F,P) - \lambda \sum_{j=1}^{n_j} L_j^{adv}(H_j, D_j)$$ ⬅ Adversarial Loss

1） **Source Domain Detection Loss：** Ensure detection network can learn sufficient knowledge

2） **Adaptive Foreground Augmentation Loss：** ensure foreground augmentation can be adapted

3） **Adversarial Loss：** Align the features of the source and target domains

Context-aware Domain Adaptive Object Detection

➢Experimental Results：

- **Ablation Studies**：1）The impact of **foreground enhancement** and **multi-level semantic representations** on domain adaptability；

Method	ORI	$l=0$ with FE	$l=0$ w/o FE	$l=1$ with FE	$l=3$ with FE	$l=3$ w/o FE	$l=4$ with FE
SSD	27.6	33.6	31.0	35.8	**38.1**	36.6	37.1
RDet	22.8	27.7	25.7	29.3	**34.7**	33.8	34.0

Features from Different Semantic Levels

Foreground Enhancement

- **Ablation Studies**：2）The impact of **multi-scale feature encoder** on domain adaptability；

Multi-scale Feature Encoder

Spatial Pooling Modules with Different Sizes

Method	mAP
SSD	27.6
Proposed SSD w/o P	38.0
Proposed SSD with P_124	39.1
Proposed SSD with P_1248	**40.1**
RDet	22.8
Proposed RDet with P_1248	**36.7**
Proposed RDet with P_124	35.2
Proposed RDet w/o P	34.8

Context-aware Domain Adaptive Object Detection

➢Experimental Results：

➢ Natural Images to Anime Images：

Verify Different Region-free Detectors

	Method	G	L	DN-8	DN-32	P-A	aero	bcy.	bird	boat	bott.	bus	car	cat	chair	cow	table	dog	hrs	bike	prsn	plnt	sleep	sofa	train	tv	mAP
Region-free	SSD [13]						20.9	56.2	20.3	16.4	9.5	38.1	33.9	10.9	37.6	22.9	22.6	10.6	22.6	48.9	43.3	35.2	7.3	30.2	36.2	27.5	27.6
	L-SSD		√				20.8	58.2	19.2	17.4	11.9	47.1	38.6	11.0	35.4	22.9	22.9	16.2	23.5	50.3	45.6	36.6	9.5	33.3	39.0	31.1	29.5
	G-SSD	√					19.6	55.3	21.5	19.8	8.1	45.9	32.1	6.9	37.1	22.7	26.1	10.6	24.8	59.5	45.6	34.5	11.8	34.5	41.2	32.1	29.5
	G-L-SSD	√	√				20.7	62.2	21.6	22.7	20.4	44.8	34.8	9.4	38.8	24.7	26.5	10.6	22.7	64.1	48.6	36.1	10.2	32.4	43.9	34.6	31.5
	LW-SSD	√	√				19.8	48.4	27.4	26.2	25.1	62.2	40.0	7.3	38.9	38.4	25.1	7.7	14.1	65.6	53.4	41.5	14.1	34.4	49.6	48.5	34.3
				√			20.5	58.9	29.1	27.1	27.3	54.5	39.1	11.3	40.9	42.5	30.2	12.9	29.3	75.2	56.9	45.3	16.1	39.6	57.3	48.9	38.1
	Proposed SSD			√	√		23.8	63.8	27.3	27.9	31.2	60.5	41.2	16.7	45.2	47.7	38.6	16.3	27.4	77.6	58.2	49.2	17.1	31.5	50.9	48.2	40.1
				√	√	√	25.9	62.3	30.1	34.2	27.0	77.4	47.2	12.2	45.9	48.8	40.1	11.8	28.0	75.5	62.8	43.4	23.7	37.7	61.9	48.4	**42.2**
	RDet [15]						20.0	41.5	21.7	17.5	25.8	46.2	24.0	10.9	34.7	12.5	24.5	16.2	17.9	48.8	32.3	39.8	3.0	20.3	35.0	26.6	26.0
				√			20.3	55.0	25.1	17.5	50.7	52.5	27.3	17.5	36.4	20.5	19.4	17.9	23.0	65.3	43.6	48.1	12.6	23.6	44.4	41.6	33.1
	Proposed RDet			√	√		24.1	58.3	29.1	26.2	46.4	61.0	39.5	17.5	44.3	44.9	25.9	16.8	28.1	65.1	60.9	49.4	18.9	30.8	56.3	50.6	39.8
				√	√	√	28.9	71.7	31.8	19.3	44.7	69.3	44.6	24.1	40.4	39.8	22.5	22.2	30.5	92.4	63.8	51.0	16.1	32.1	67.0	56.4	**43.5**
	SSD* [13]						23.1	59.8	23.1	15.0	18.0	58.5	40.8	15.1	41.5	40.1	33.4	20.1	29.8	58.9	49.7	25.3	18.4	30.5	41.8	43.8	34.3
	Proposed SSD*			√	√		27.9	62.1	29.7	28.9	38.6	81.5	50.7	14.9	49.5	56.1	40.2	15.6	38.7	73.4	60.5	39.5	21.5	41.3	63.1	51.7	44.3
	RDet* [15]						26.0	55.9	28.0	25.4	34.4	52.3	45.1	16.4	52.8	25.9	26.8	19.1	40.7	50.3	46.1	41.3	16.1	29.6	47.3	32.6	35.6
	Proposed RDet*			√	√		30.0	60.3	39.1	30.6	55.4	69.2	55.6	27.5	51.3	52.1	37.7	26.7	43.3	77.0	72.0	59.0	26.5	43.1	64.9	56.1	48.9
	WST+BSR [37]						28.0	64.5	23.9	19.0	21.9	64.3	43.5	16.4	42.2	25.9	30.5	7.9	25.5	67.6	54.5	36.4	10.3	31.2	57.4	43.5	35.7
Region-based	Faster [10]♠						15.7	31.9	22.4	8.2	38.8	59.4	17.8	6.6	37.0	5.7	12.7	7.2	17.4	49.0	36.0	32.1	11.2	2.9	29.8	28.4	23.5
	Faster [10]♣						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
	DA-Faster [43]♣	√					15.8	33.9	22.5	14.8	24.9	48.7	27.9	12.5	32.7	35.5	21.3	17.9	17.4	55.0	48.5	34.8	11.4	21.3	47.1	37.7	29.1
	G-L-Faster [42]♠	√	√				16.0	53.2	27.5	21.6	32.0	48.4	32.4	12.2	32.5	27.3	12.3	13.1	24.3	62.4	55.5	41.2	21.0	13.2	37.8	46.1	31.5
	G-L-Faster [42]♣	√	√				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
	ICR-CCR [39]♣						28.7	55.3	31.8	26.0	40.1	63.6	36.6	9.4	38.7	49.3	17.6	14.1	33.3	74.3	61.3	46.3	22.3	24.3	49.1	44.3	38.3
	DD+MRL [36]♠	√	√			√	25.8	63.2	24.5	42.4	47.9	43.1	37.5	9.1	47.0	46.7	26.8	24.9	48.1	78.7	63.0	45.0	21.3	36.1	52.3	53.4	41.8

G:Global Alignment
L:Local Alignment

Common DA modules offer limited improvements to region-free detectors

➢Experimental Results:

➢ Natural Images to Ink Painting Images:

TABLE III
RESULTS ON ADAPTATION FROM PASCAL VOC TO COMIC. THE EVALUATION OF TARGET DOMAIN AND SOURCE DOMAIN IS ON THE TEST SET OF COMIC AND TEST SET OF PASCAL VOC 2007, RESPECTIVELY. THE DEFINITION OF DN-8, DN-32, AND P-A FOLLOWS TABLE I.

Method	DN -8	DN -32	P-A	Target Domain							Source Domain
				bicycle	bird	car	cat	dog	prsn	mAP	mAP
SSD [14]				21.7	12.8	**34.4**	11.0	14.6	44.4	23.1	81.4
SSD+DSEMs (ours)	✓			39.7	15.2	22.6	14.9	25.9	50.3	28.1	81.1
	✓	✓		49.6	18.2	26.6	**28.8**	30.8	46.3	33.4	80.1
	✓	✓	✓	**57.8**	**22.2**	32.2	28.5	**32.9**	**56.8**	**38.4**	79.5
ADDA [32]				39.5	9.8	17.2	12.7	20.4	43.3	23.8	\
DD+MRL [45]				\	\	\	\	\	\	**34.5**	\
WST+BSR [46]				**50.6**	**13.6**	**31.0**	7.5	16.4	41.4	26.8	\
DT [47]				43.6	13.6	30.2	**16.0**	**26.9**	**48.3**	29.8	\

Common DA modules offer limited improvements to region-free detectors

Context-aware Domain Adaptive Object Detection

➢Experimental Results:

➢ Sunny Images to Foggy Images:

TABLE V

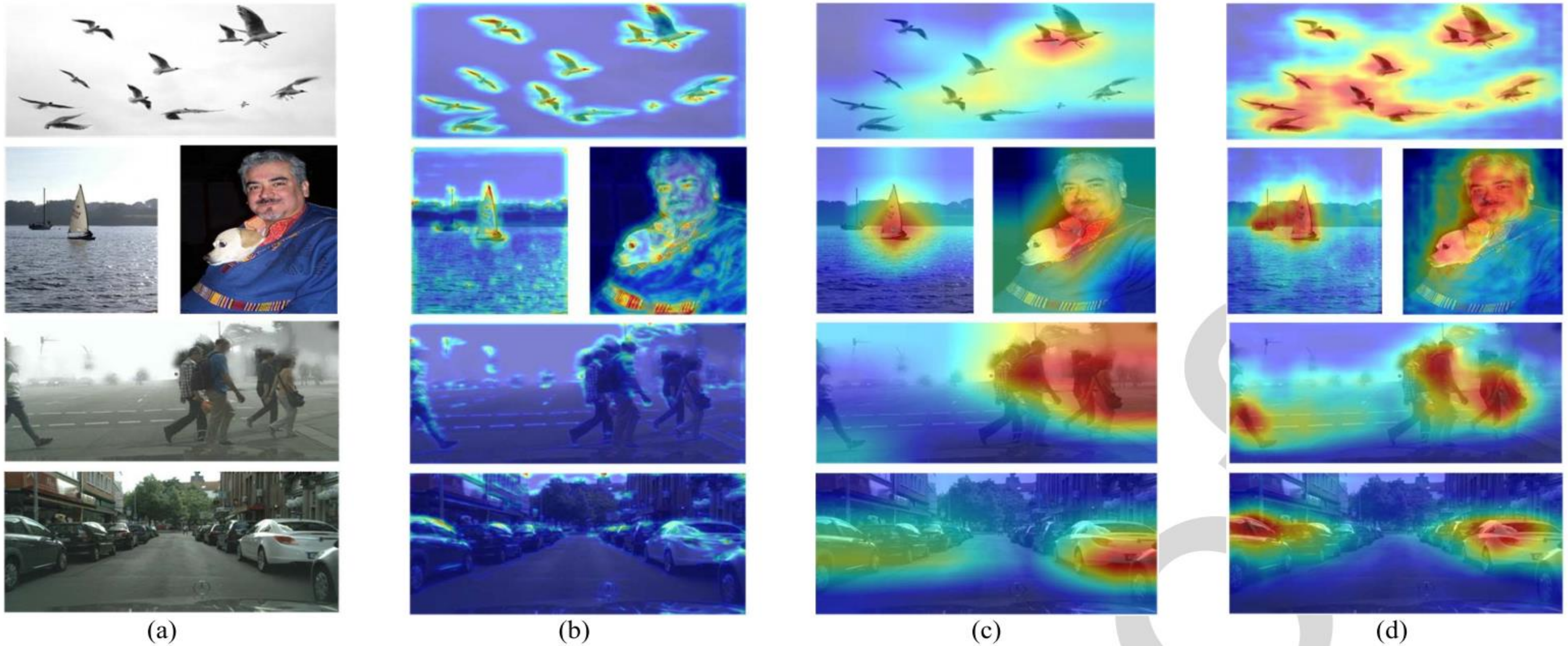
RESULTS ON THE VALIDATION SET OF FOGGYCITYSCAPES. THE DEFINITION OF G, L, LW, DN-8 AND DN-16 FOLLOWS TABLE I. ORACLE REFERS TO TRAINING THE DETECTOR ON THE LABELED TARGET IMAGES.

Method		G	L	DN-8	DN-16	bus	bicycle	car	bike	prsn	rider	train	truck	mAP
Region-free	RDet [16]					24.3	28.9	38.0	21.8	26.1	28.5	8.3	6.6	22.8
	L-RDet		√			32.0	33.3	44.6	26.8	30.8	34.3	20.0	12.7	29.3
	G-RDet	√				31.9	32.1	44.3	27.0	30.3	33.9	20.2	16.2	29.5
	G-L-RDet	√	√			33.5	33.6	46.7	30.1	32.4	34.7	25.3	15.5	31.5
	LW-RDet	√	√			33.8	32.1	45.2	26.7	31.4	35.3	25.8	14.3	30.6
	RDet+DSEMs (ours)			√		40.7	35.3	55.1	27.4	34.8	**38.5**	26.5	19.0	34.7
				√	√	**40.8**	**35.3**	**56.2**	**31.2**	**35.9**	38.1	**34.9**	**21.2**	**36.7**
	Oracle					41.9	38.7	63.3	33.3	39.9	42.8	31.8	27.3	39.8
Region-based	Faster [11]					22.3	26.5	34.3	15.3	24.1	33.1	3.0	4.1	20.3
	DA-Faster [52]	√				25.0	31.0	40.5	22.1	**35.3**	20.2	20.0	27.1	27.6
	G-L-Faster [51]	√	√			36.2	**35.3**	43.5	30.0	29.9	42.3	32.6	24.5	34.3
	MAF [49]					39.9	33.9	43.9	29.2	28.2	39.5	33.3	23.8	34.0
	ICR-CCR [48]					**45.1**	34.6	**49.2**	**30.3**	32.9	**43.8**	**36.4**	**27.2**	**37.4**
	DD+MRL [45]					38.4	32.2	44.3	28.4	30.8	40.5	34.5	27.2	34.6
	Oracle					51.9	37.8	53.0	36.8	36.2	47.7	41.0	34.7	42.4

G:Global Alignment
L:Local Alignment

Common DA modules offer limited improvements to region-free detectors

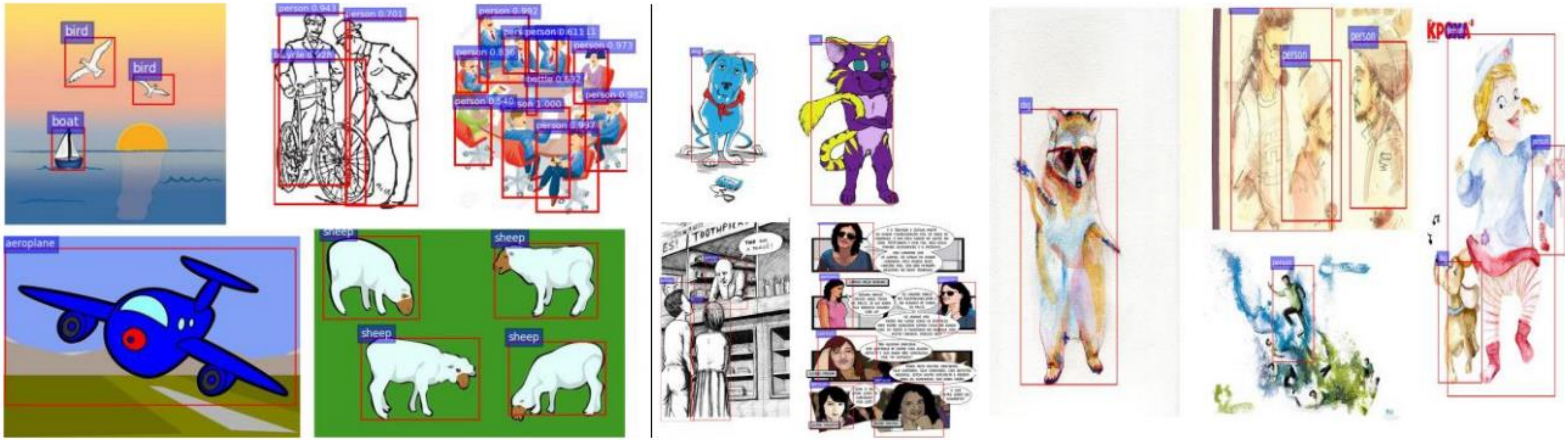
➢Experimental Results：



Original Images	local alignment can only capture detailed features	global alignment ignores multiple instances	**DSEM** can capture multi-scale instance information

Context-aware Domain Adaptive Object Detection

➢Experimental Results：



(a) (b) (c)

Art Painting Dataset

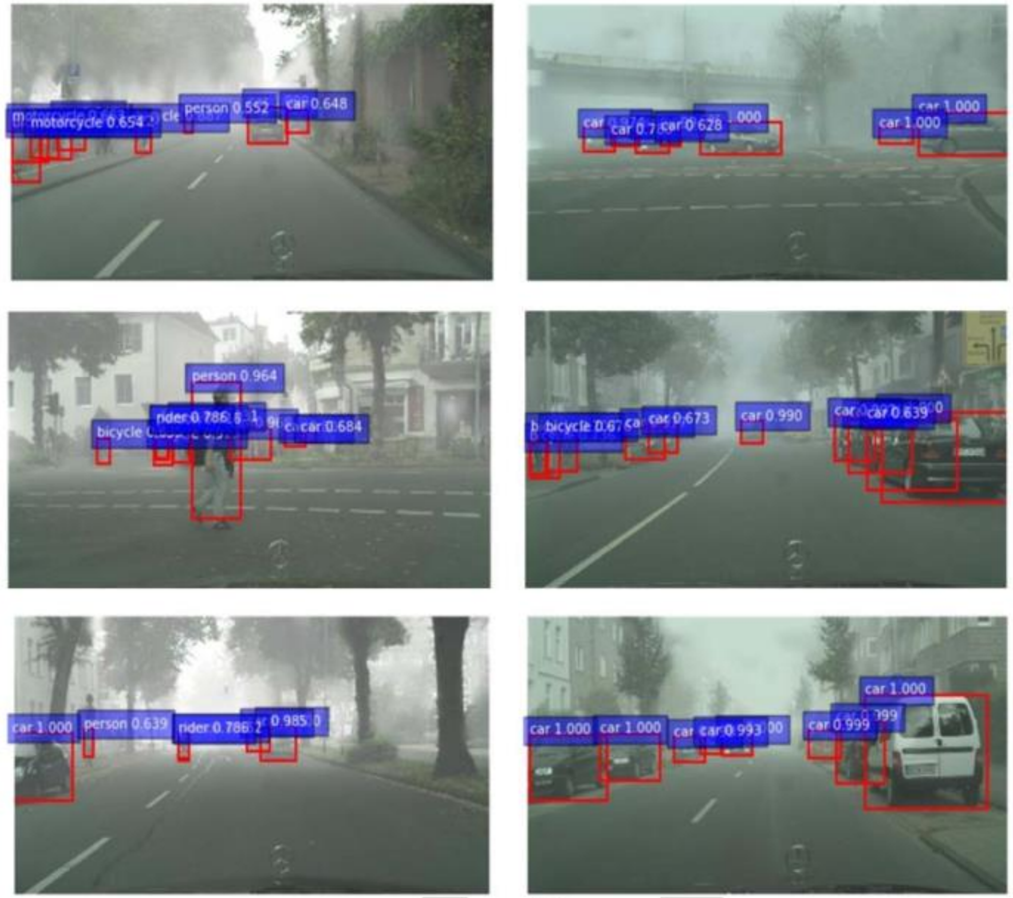
Cartoon Dataset

Ink Painting Dataset

Context-aware Domain Adaptive Object Detection

➢Experimental Results：

(a)



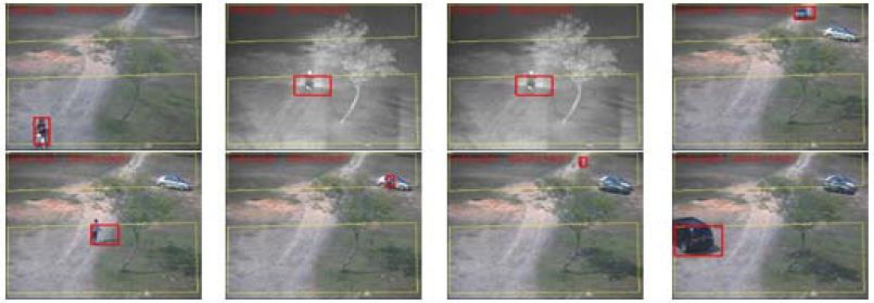
(b)

Without DSEM Adaptation

With DSEM Adaptation

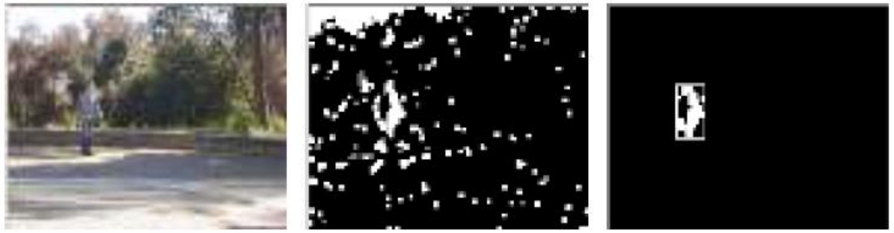
- **Static PID：Detect abnormal persons from static scenes captured by a fixed camera**

[1]Chen C H, Chen T Y, Lin Y C, et al. Moving-Object Intrusion Detection Based on Retinex-Enhanced Method



[2]Liang K M, Hon H W, Khairunnisa M J, et al. Real time intrusion detection system for outdoor environment

[3]Wang J. Research and implementation of intrusion detection algorithm in video surveillance

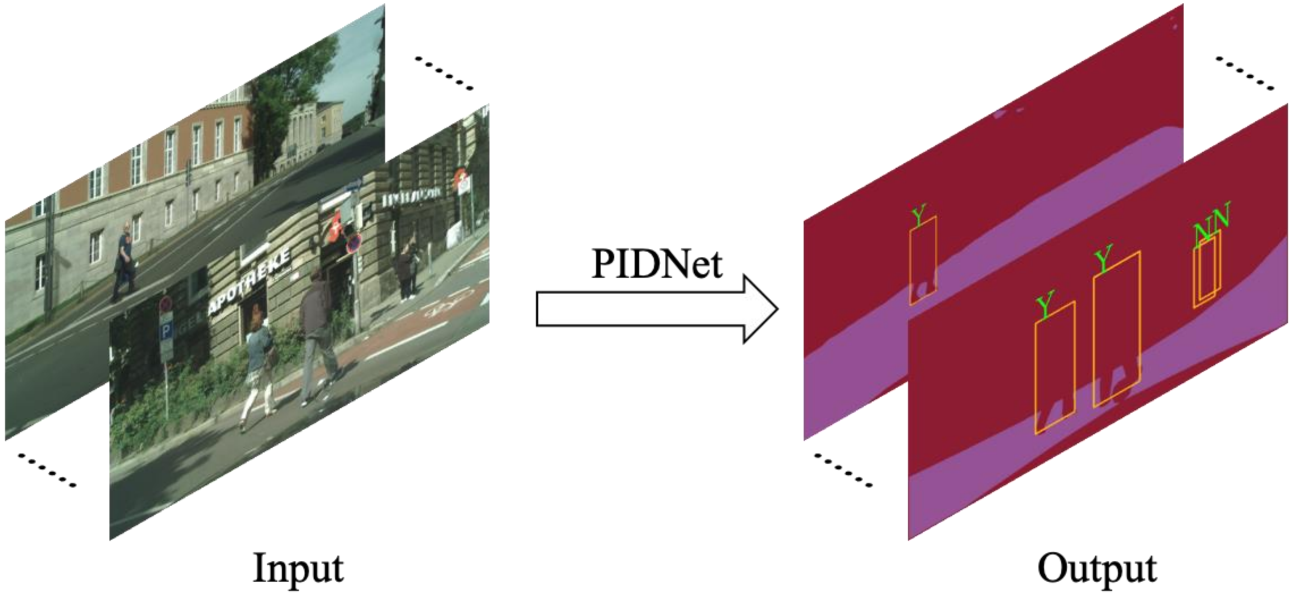


[4]Zhang M, Jin J S, Wang M, et al. Pedestrian intrusion detection based on improved GMM and SVM

Context-aware Dynamic Pedestrian Intrusion Detection

➢Dynamic PID:

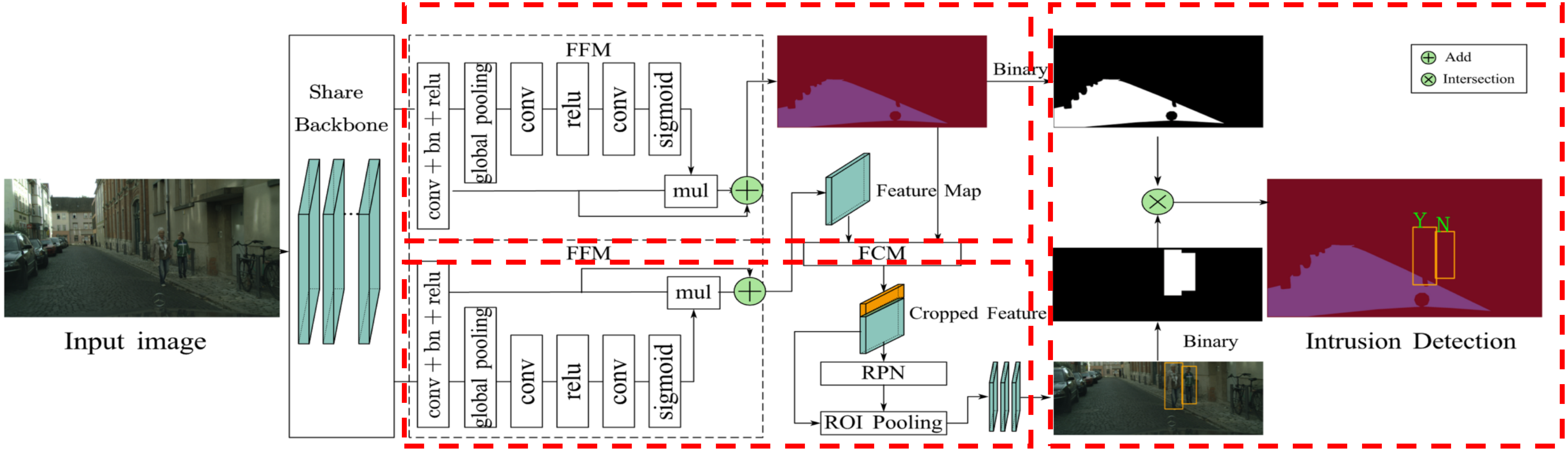
➢ Our Method:



AoI Segmentation + Pedestrian Detection

➢ Overview:

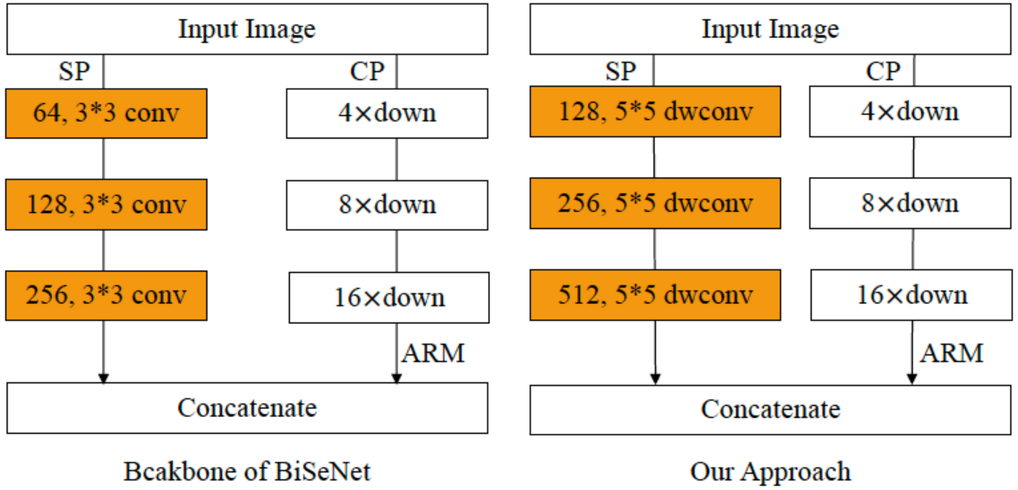
PIDNET



PIDNet mainly consists of a segmentation network in the upper branch and an object detection network in the lower branch

Context-aware Dynamic Pedestrian Intrusion Detection

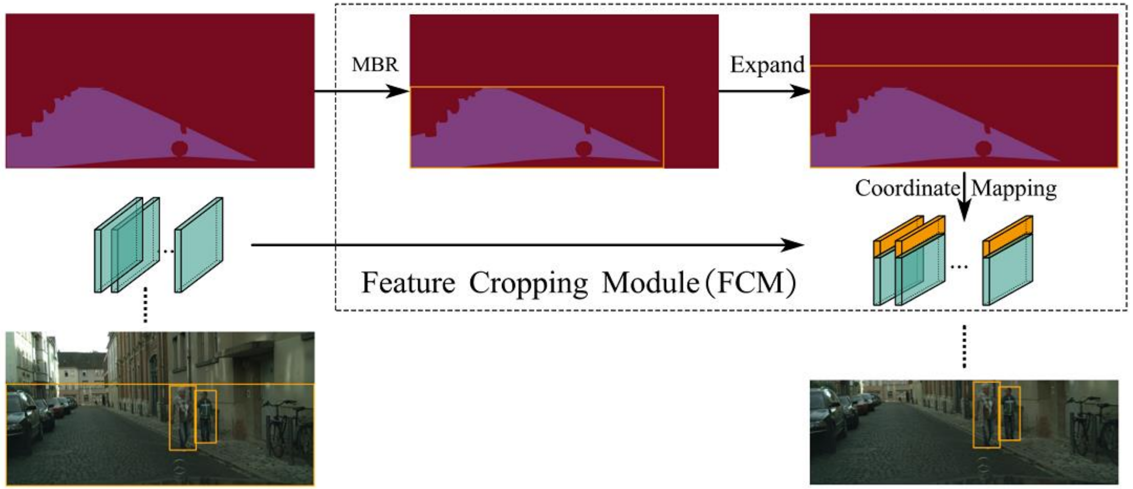
➢ Feature Sharing Design：



High level: context information

Low level: spatial information

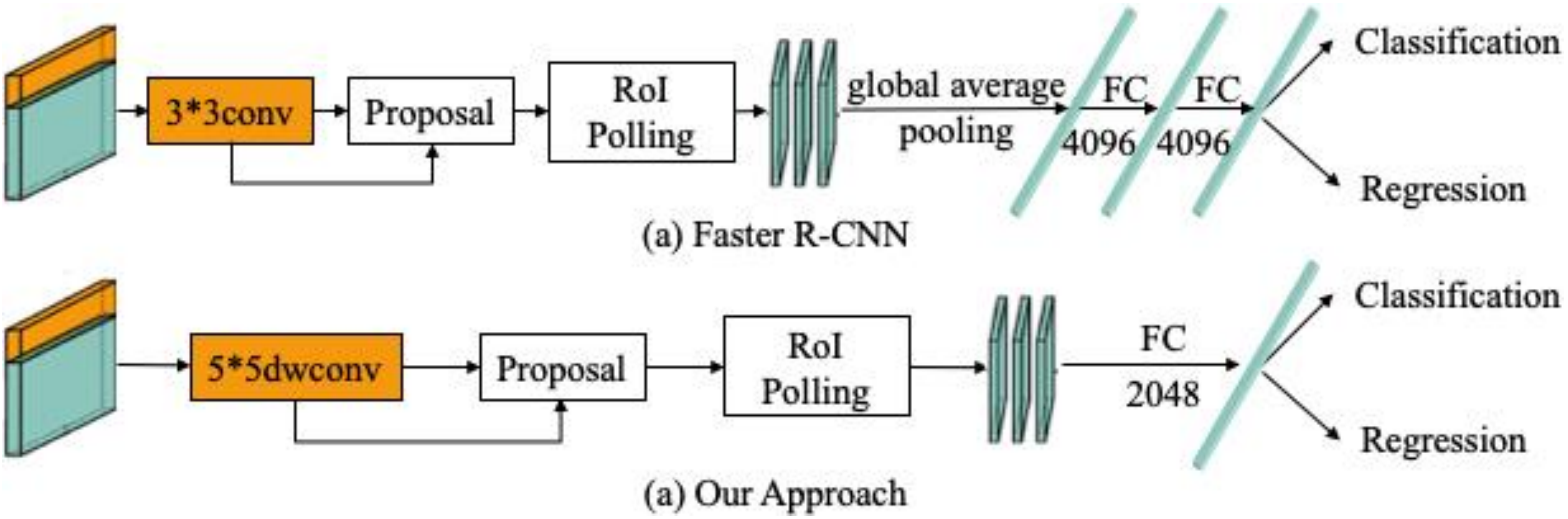
➢ Feature Cropping Design：



$$\begin{cases} Y'_{max} = \alpha(Y_{max} - Y_{min}) + Y_{min} \\ X'_{max} = \alpha(X_{max} - X_{min}) + X_{min} \end{cases}, \qquad (1)$$

$$\begin{cases} y_{max} = \lfloor Y'_{max}/s \rfloor + 1, x_{max} = \lfloor X'_{max}/s \rfloor + 1 \\ y_{min} = \lceil Y_{min}/s \rceil - 1, x_{min} = \lceil X_{min}/s \rceil - 1 \end{cases}, \qquad (2)$$

➢Feature Compression Design：



➢ Dataset Comparison

Cityintrusion Dataset

Image　　Cityscape　　Cityperson　　Ours

➢ Dataset：

Cityintrusion Dataset

Categories	Train	Val	Total
Cites	18	3	21
Images	2303	398	2701
Intrusion Cases	3829	770	4599
No-Intrusion Cases	12691	2393	15084
Cases per image	7.2	7.9	7.3
Intrusion Rate(%)	23.2	24.3	23.3

$$PID\_AP = \frac{1}{N} \sum_{r \in \{0,0.1,0.2,\ldots 1\}} max\ pre(c,p) | re(c,p) \geqslant r \quad (3)$$

$$tp = tp + 1, if (IoU > 0.5) \cap (c > c_t) \cap (p > p_t) \quad (4)$$

➢ Experiments:

Results-Table

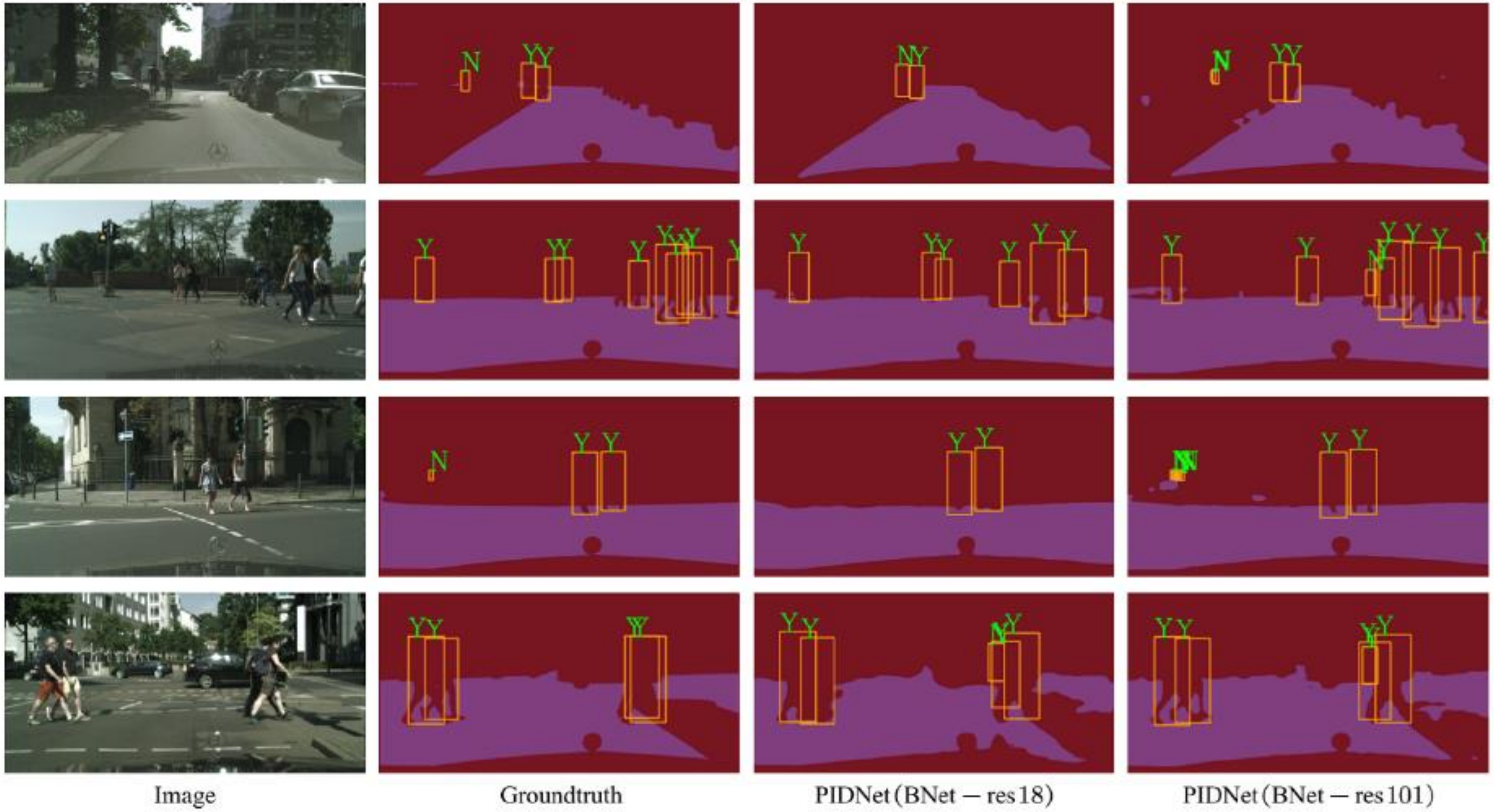
Model	Backbone	PID mAP	PID Acc	Speed(fps)	Params(M)
(a) PSPNet+Faster R-CNN	res101+vgg16	29.8	57.4	0.09	202.8
(b) ICNet+Faster R-CNN	resnet50+vgg16	34.5	61.1	0.15	184.6
(c) BISeNet+Faster R-CNN	resnet 18+vgg16	36.7	63.1	0.18	150.7
(d) PIDNet	BNet-res18	36.7	63.3	9.6	105.8
(e) PIDNet	BNet-res101	**49.2**	67.1	**5.4**	**138.7**

The rows a, b and c in the table represent results obtained using two existing networks, and rows d and e indicate results obtained using our network.

PID mAP, PID Acc are the evaluation metrics of dynamic PID.

➢ Experiments:

Results-Images



Context-aware Dynamic Pedestrian Intrusion Detection

➢ Experiments:

Ablation Study

Backbone	Seg IoU	Det AP	PID Acc	Params(M)
SP+CP	97.8	69.6	59.8	12.5
SP+CP+5*5	98.0	71.5	61.1	14.1
SP+CP+5*5Dw	98.0	71.5	61.1	11.7
SP+CP+5*5Dw+Add channel	98.0	72.8	63.3	12.2

Ablation study on the shared backbone

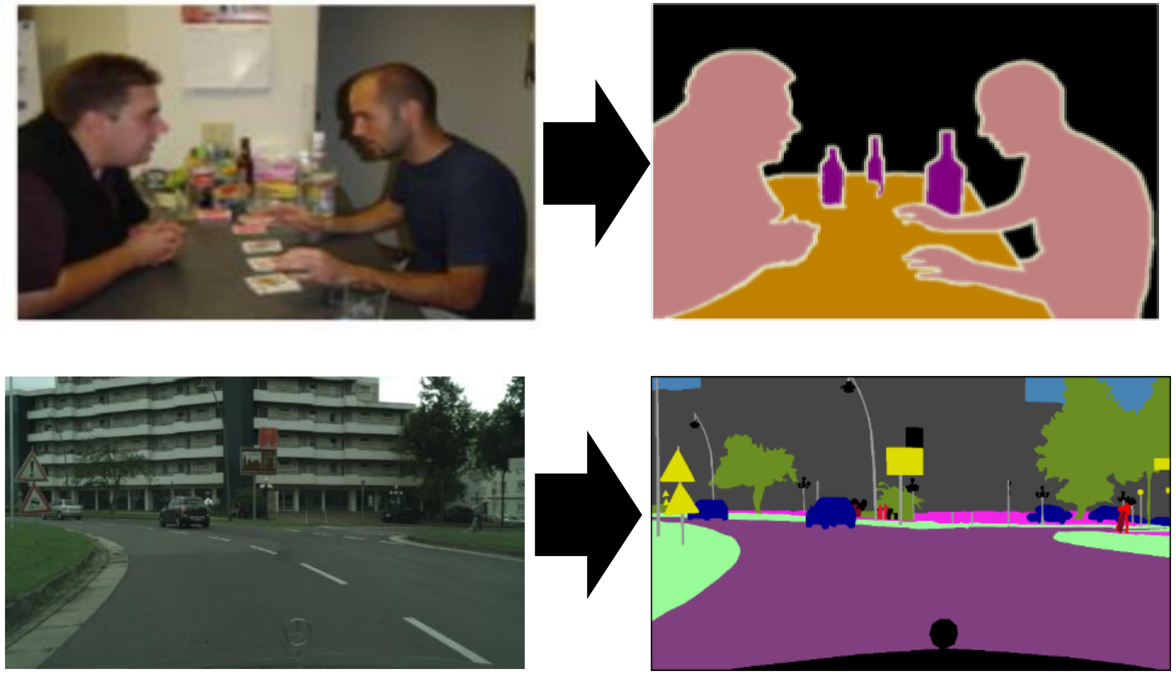
PID Net	FC	FC-Extension	M-RPN	S-RCNN	PID Acc	Speed(fps)
✔					61.5	3.6
✔	✔				60.7	6.4
		✔			61.4	6.1
✔			✔		63.3	2.9
✔				✔	61.3	7.4
✔		✔	✔	✔	63.3	9.6

Ablation studies on feature cropping module and network compression

Context-aware Rapid Semantic Segmentation

➢Semantic segmentation:

- ➢ **Task definition:** The image semantic segmentation task is dividing the pixels of an image into two or more sets, each set represents a specific semantic.



Context-aware Rapid Semantic Segmentation

- **An increasing desire to design lightweight semantic segmentation neural network：**

- **Lightweight networks focus on four indicators:**
  - Inference speed
  - Number of parameters
  - FLOPs
  - Accuracy

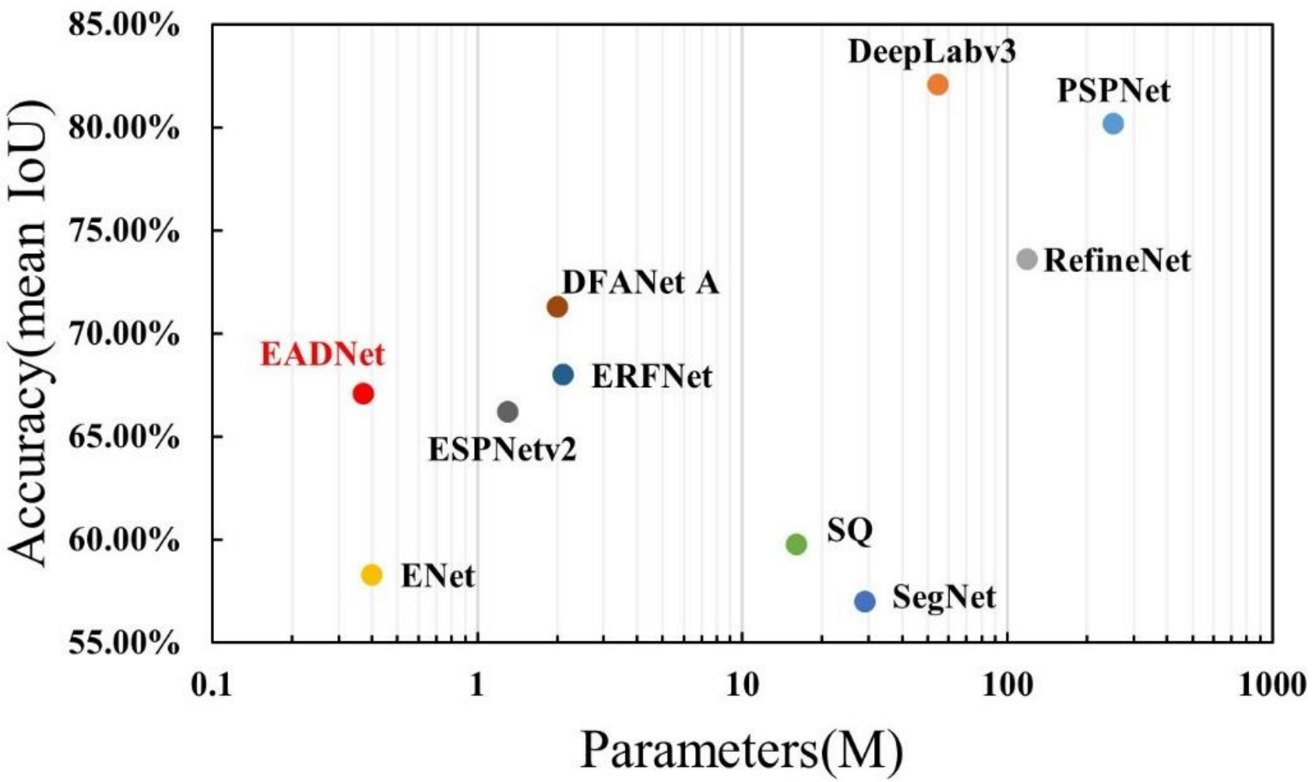
Context-aware Rapid Semantic Segmentation

➢The high-accuracy networks:

- ➢ Have a larger number of parameters
- ➢ Causing heavier computational cost
- ➢ Difficult to meet the real-time requirement on edge devices.

➢Lightweight networks:

- ➢ Sacrificing the prediction accuracies

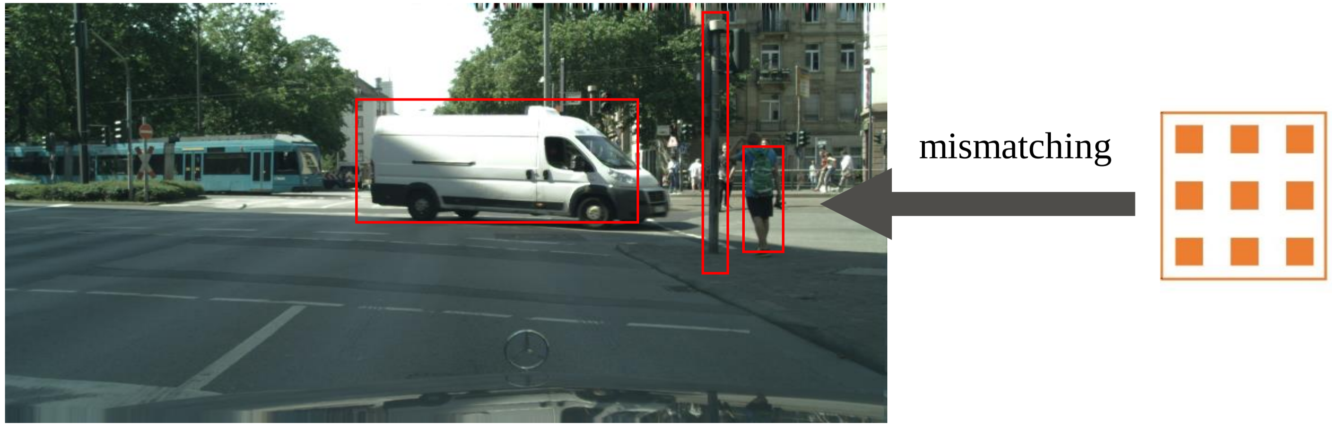


Our EADNet is on the left top of the figure, and achieves the best trade-off of parameters and accuracy.

➢Motivation and Objective：

- **Problem One：** There are lots of irregular rectangular objects with different scales in urban street images, traditional square receptive field of network can not effectively matching these objects.

- **Problem Two：** Lacking the convolution block which can extract context multi-scale multi-shape with less parameters and lower computation cost.

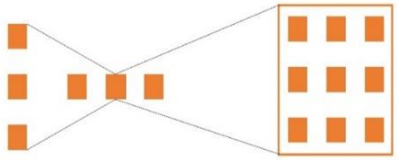


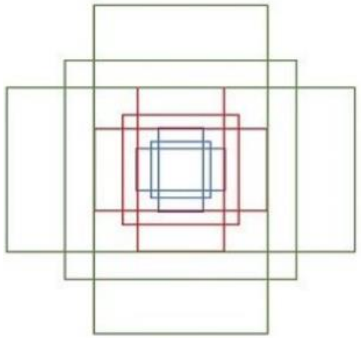
Context-aware Rapid Semantic Segmentation

➢The proposed method:

- The receptive field of 1*3 and 3*1 asymmetric convolution group with the same dilated rate equal to 3*3 convolution.

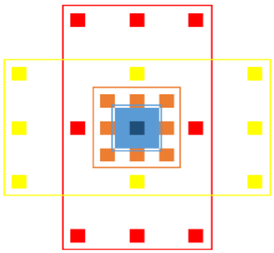
- The anchor boxes in faster R-CNN can capture the features of objects with different sizes and shapes.

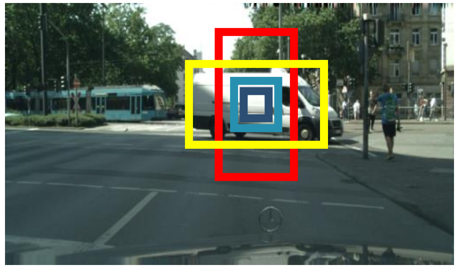




asymmetric dilated convolution + anchor boxes in faster R-CNN

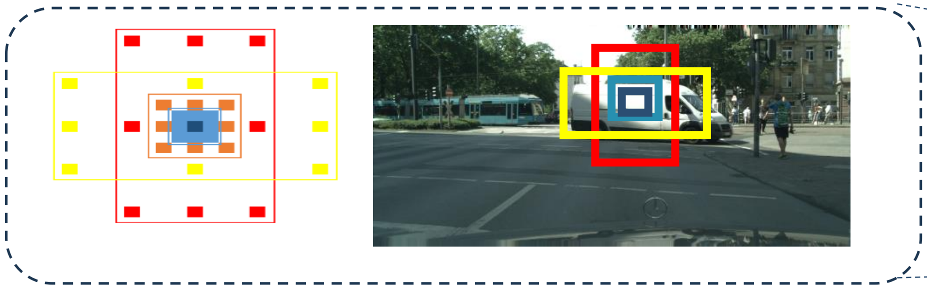
Different dilation rate

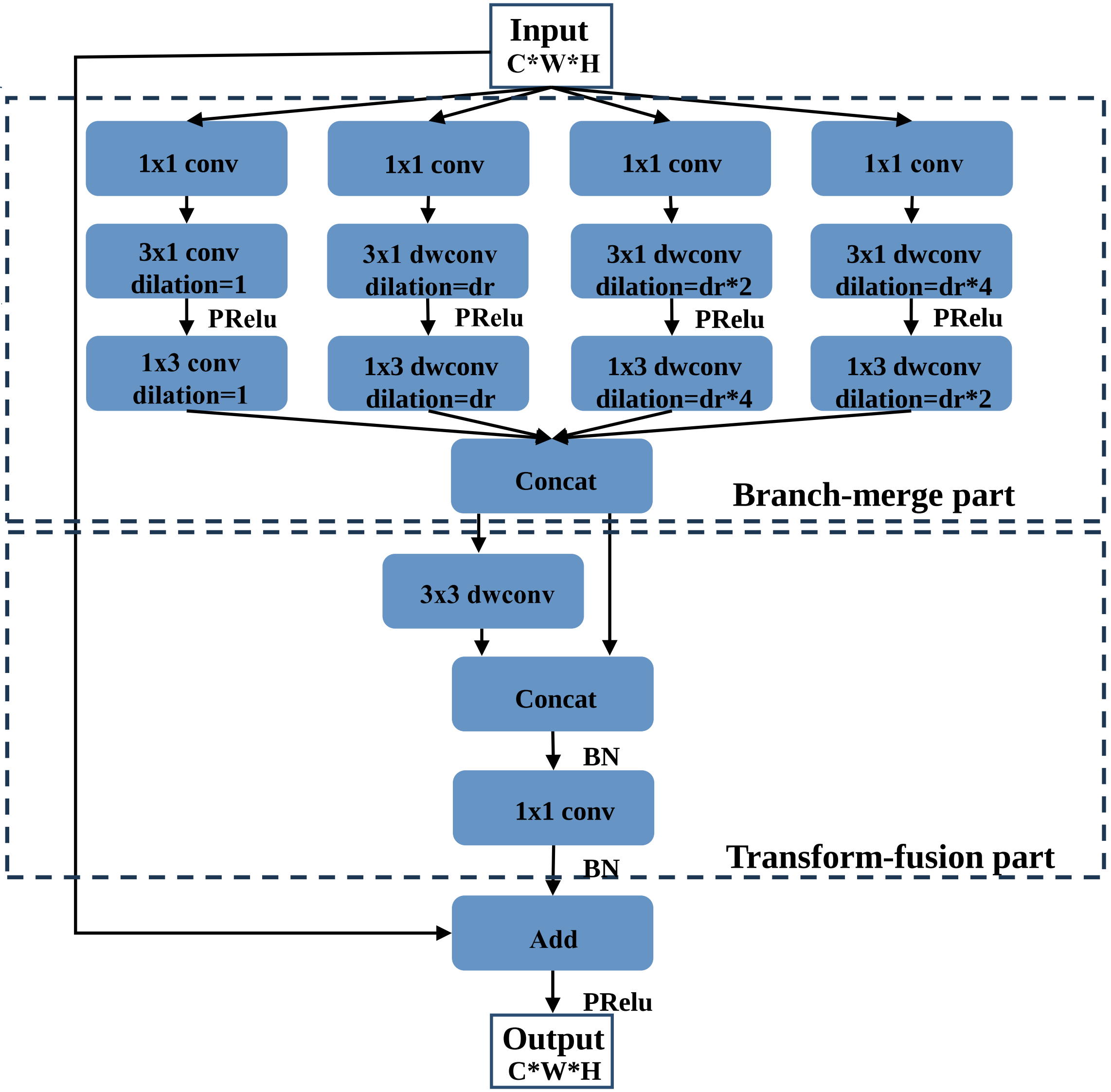




Multi-scale multi-shape receptive field

Context-aware Rapid Semantic Segmentation

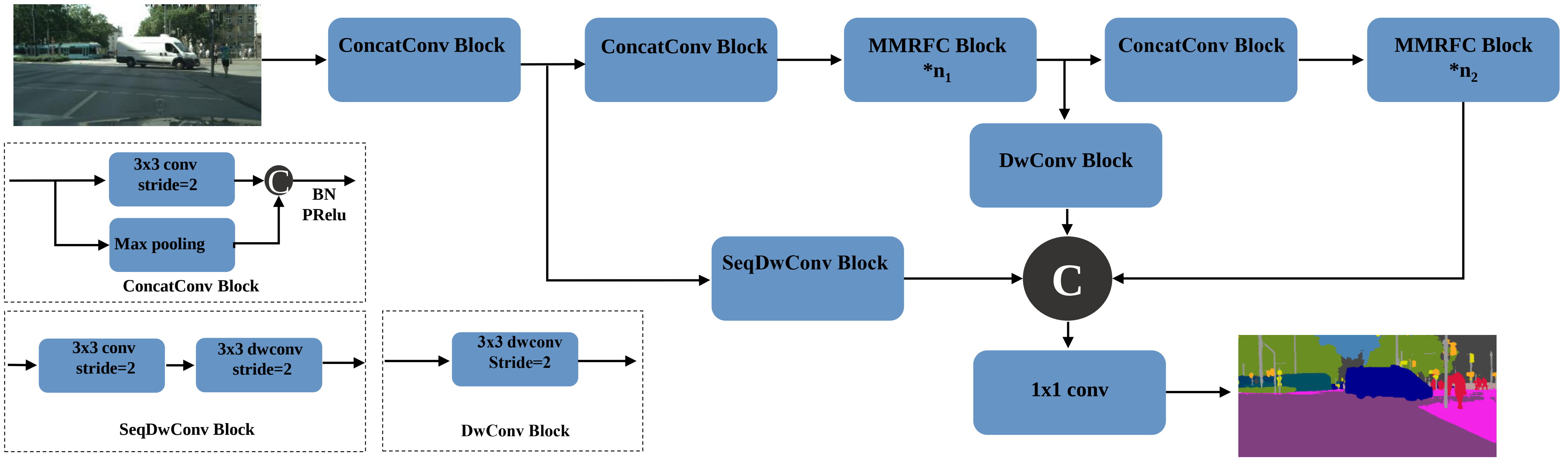




- **MMRFC block:**
  - Bottleneck + Highpass way structure
  - Depthwise conv and dilated conv in each branch
  - Each branch captures multi-scale multi-shape receptive field
  - Transform-fusion does feature mapping and restoring

➢ EADNet:

- Three special designed down sampling blocks in different layers.
- Skip connections to combine detailed information and semantic information.



Context-aware Rapid Semantic Segmentation

➢ Experimental results:

- ➢ EADNet is the smallest semantic segmentation network amongst state-of-the-art networks.
- ➢ EADNet achieves a competitive performance in CamVid and Cityscapes dataset.
- ➢ EADNet has fast inference speed, less FLOPs and strong feature extraction ability.

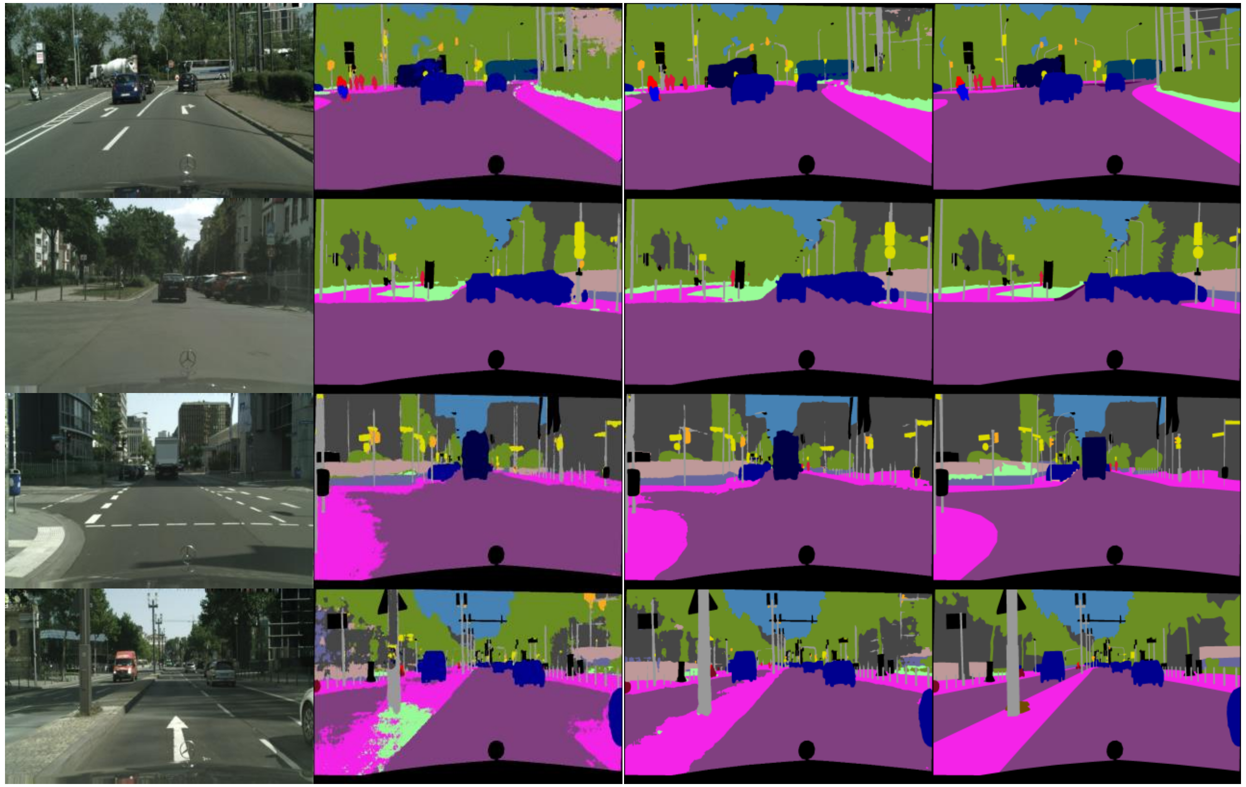
FPS, FLOPs, parameter size and mIoU comparison on Cityscapes test set

method	Pre-train	Input size	Inference time(ms)	FPS	FLOPs(G)	Parameters(M)	mIoU(%)
SegNet	ImageNet	1024*2048	152.68	6.55	1310	29	57.0
SQ	ImageNet	1024*2048	88.19	11.34	501	16	59.8
ERFNet	_	1024*2048	43.71	22.88	103	2.1	68.0
ENet	_	1024*2048	36.9	27.11	22	0.37	58.3
DFANet A	ImageNet	1024*2048	26.91	37.16	28	2.0	**71.3**
ESPNetv2	_	1024*2048	24.58	40.69	23.5	1.3	66.2
Ours	_	1024*2048	**23.98**	**41.7**	**18**	**0.35**	67.1

Context-aware Rapid Semantic Segmentation

Experimental results:

Visualization result on Cityscapes test set



(a) Image (b) ENet (c) EADNet (d) GT

Experiment result on CamVid test set

method	Input size	FLOPs(G)	mIoU(%)
SegNet	960*720	427.34	46.4
DFANet	960*720	9.03	64.7
Ours	960*720	**5.99**	**68.3**

三、Conclusion and Future Work

Conclusion and Future Work

Limited Computation Resource

Memory consumption

Inference speed

Bandwidth

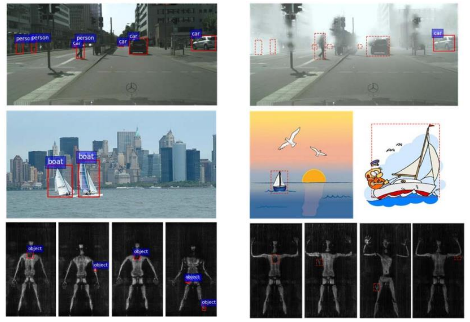
Parameters

FLOPs

Communication

Dynamic mobile device deployment

Limited Data with Changing Context



Contexts in the image

Contexts in the device

Few-shot/weakly-supervised learning

Published Context-aware MVA Works

[1] B. Zhang, **T. Chen**, X. F. Wu, L. M. Zhang, J. Fan, "Densely Semantic Enhancement for Domain Adaptive Region-free Detectors," in press, ***IEEE Transactions on Circuits and Systems for Video Technology (T-CSVT)***, 2021.

[2] Yang, Q., Chen, T., Fan, J., Lu, Y., Zuo, C., and Chi, Q. EADNet: Efficient Asymmetric Dilated Network For Semantic Segmentation. In ***IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP***, 2021.

[3] **T. Chen,** S. Lu, J.Fan*, "SS-HCNN: Semi-Supervised Hierarchical Convolutional Neural Network for Image Classification,"***IEEE Transactions on Image Processing (T-IP),*** 28(5):2389-2398, 2019.

[4] **T. Chen**, S. Lu, J. Fan, " S-CNN: Subcategory-aware convolutional networks for object detection ," ***IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI)***, 40(10):2522-2528, 2018.

[5] J. Sun, J. Chen, **T. Chen**, J. Fan, S. He, "PIDNet: An Efficient Network for Dynamic Pedestrian Intrusion Detection," ***ACM Conference on Multimedia (ACM'MM),*** Seattle, Washington, USA, 2020.

[6] R. Wu, G. Zhang, S. Lu, **T. Chen**, "Cascade EF-GAN: Progressive Facial Expression Editing with Local Focuses," ***IEEE Internatonal Conference on Computer Vision and Pattern Recognition (CVPR, Oral),*** Seattle, Washington, 2020.

[7] T. Chen, S. Lu, "Object-level motion detection from moving optic cameras," ***IEEE Transactions on Circuits and Systems for Video Technology (T-CSVT),*** 27(11):2333-2343, 2017.

[8] T. Chen, S. Lu, "Robust vehicle detection and viewpoint estimation with soft discriminative mixture model," ***IEEE Transactions on Circuits and Systems for Video Technology (T-CSVT),*** vol.27, no. 2, pp. 394-403, Feb 2017.

Thanks for Listening!