Attentive Alignment Network for Multispectral Pedestrian Detection

Nuo Chen[1], Jin Xie[1*], Jing Nie[1], Jiale Cao[2], Zhuang Shao[3], Yanwei Pang[2,4]

[1]Chongqing University, [2]Tianjin University, [3]University of Warwick, [4]Shanghai Artificial Intelligence Laboratory

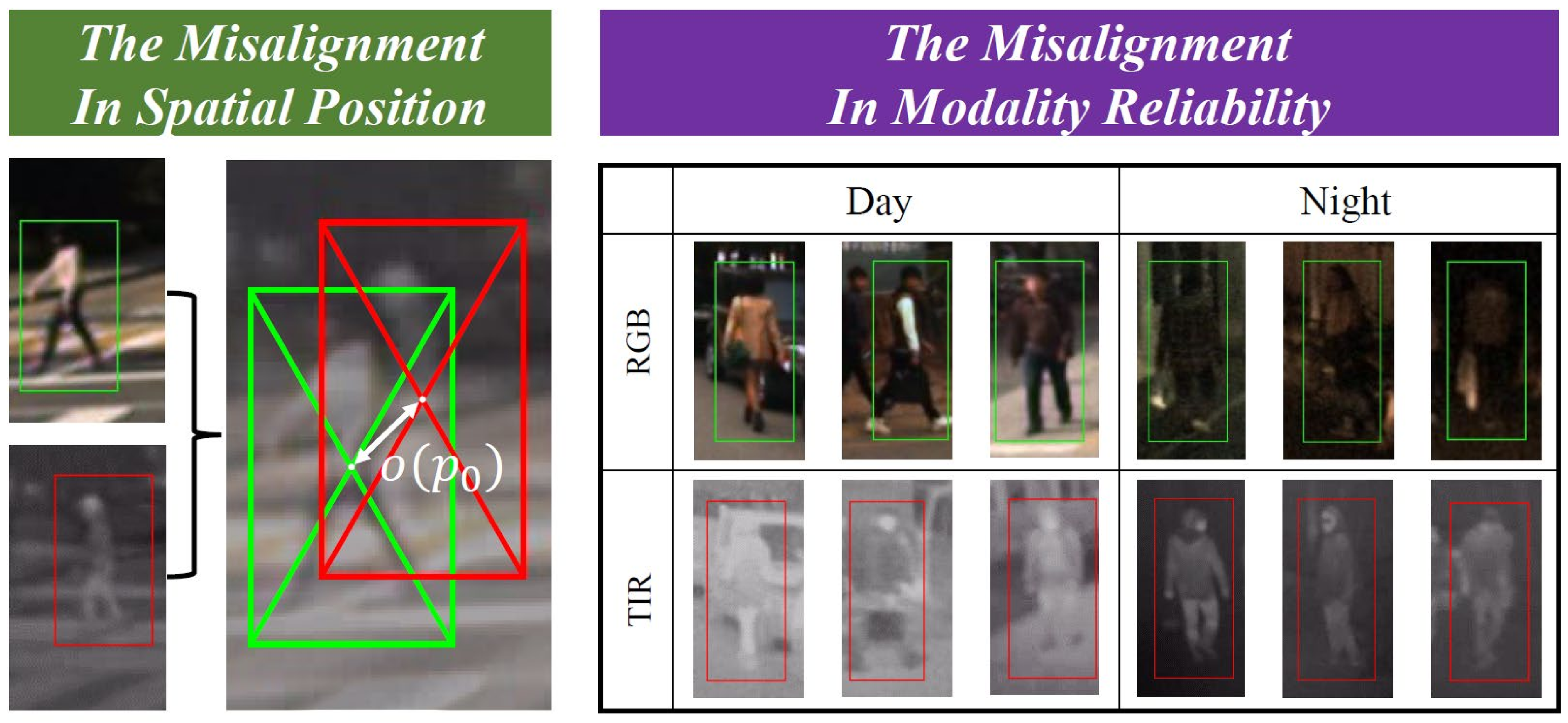
ACM multimedia

Ottawa, Canada OCT 29 – NOV 3, 2023

Summary

- Multispectral pedestrian detection is crucial for around-the-clock applications.
- The misalignment in both spatial position and modality reliability hamper its efficiency.
- Our proposed AANet addresses these misalignments, achieving state-of-the-art performance in both KAIST dataset and CVC-14 dataset.

Motivation



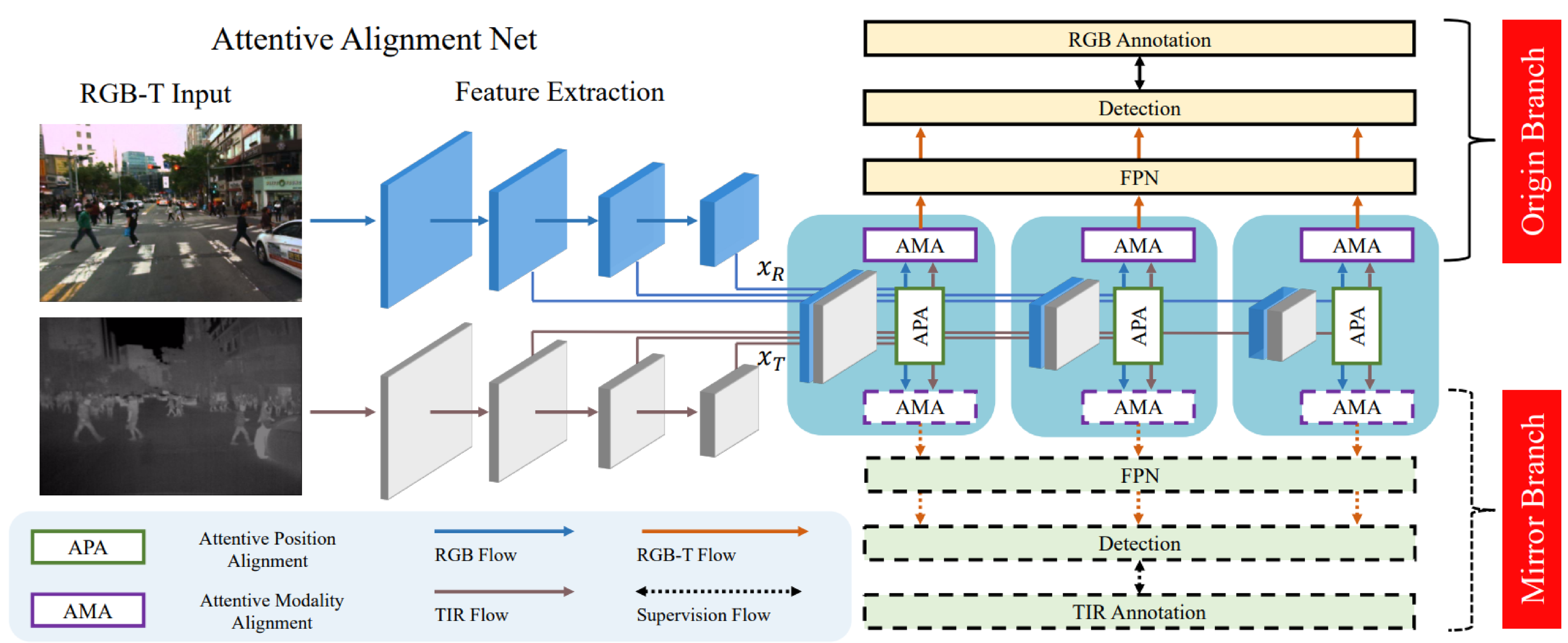
Misalignment In Spatial Position:

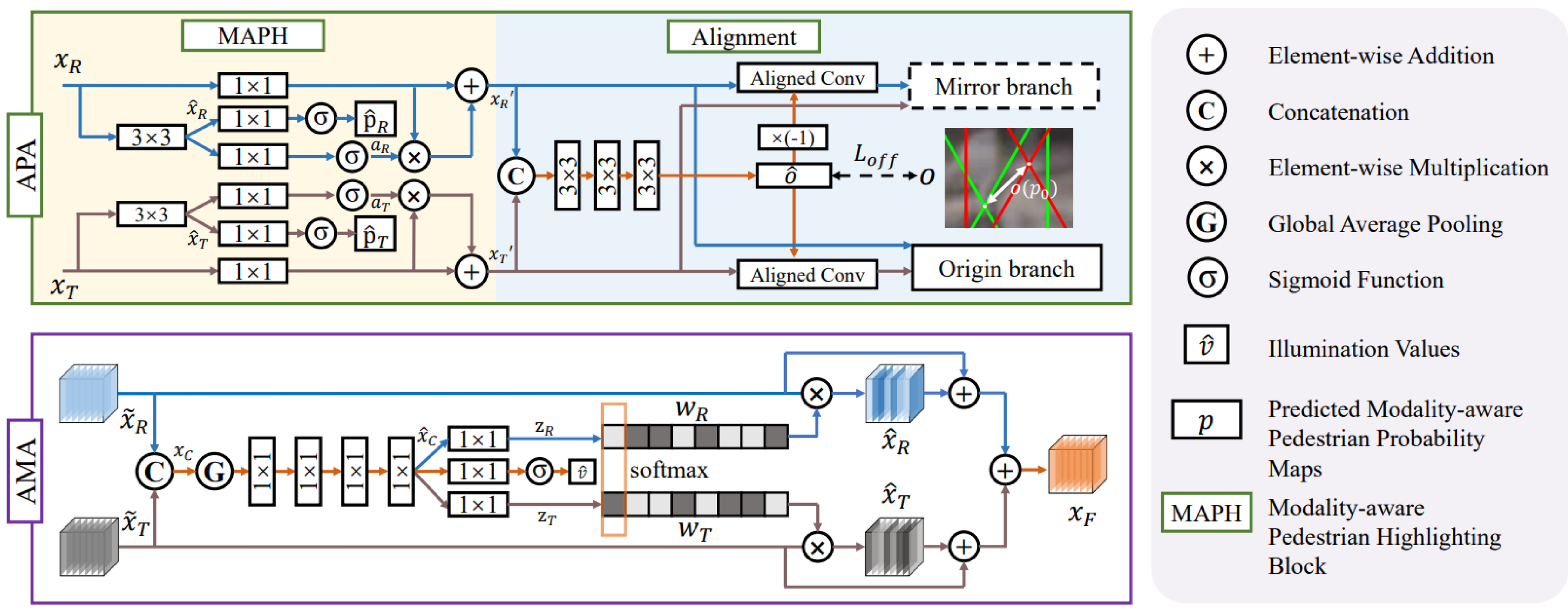
Same pedestrian has different positions in different modalities.

Misalignment In Modality Reliability:

The reliability of different modalities changes with various light conditions.

Method





Attentive Positional Alignment(APA):

Modality-aware Pedestrian Highlighting Block:

Highlighting the regions of pedestrians through predicting pixel-wise attention maps.

Aligned Convolution:

Convolution kernels are shifted by the predicted spatial offsets between different modalities in a supervised manner.

Attentive Modality Alignment(AMA):

- Propose illumination-guided attention mechanism
- Adaptively aggregating features of RGB and TIR modalities in a data-driven manner.

Mirror Training Strategy:

Introduce a mirror branch which only used in training stage, further improving the accuracy of offset prediction.

Experiment

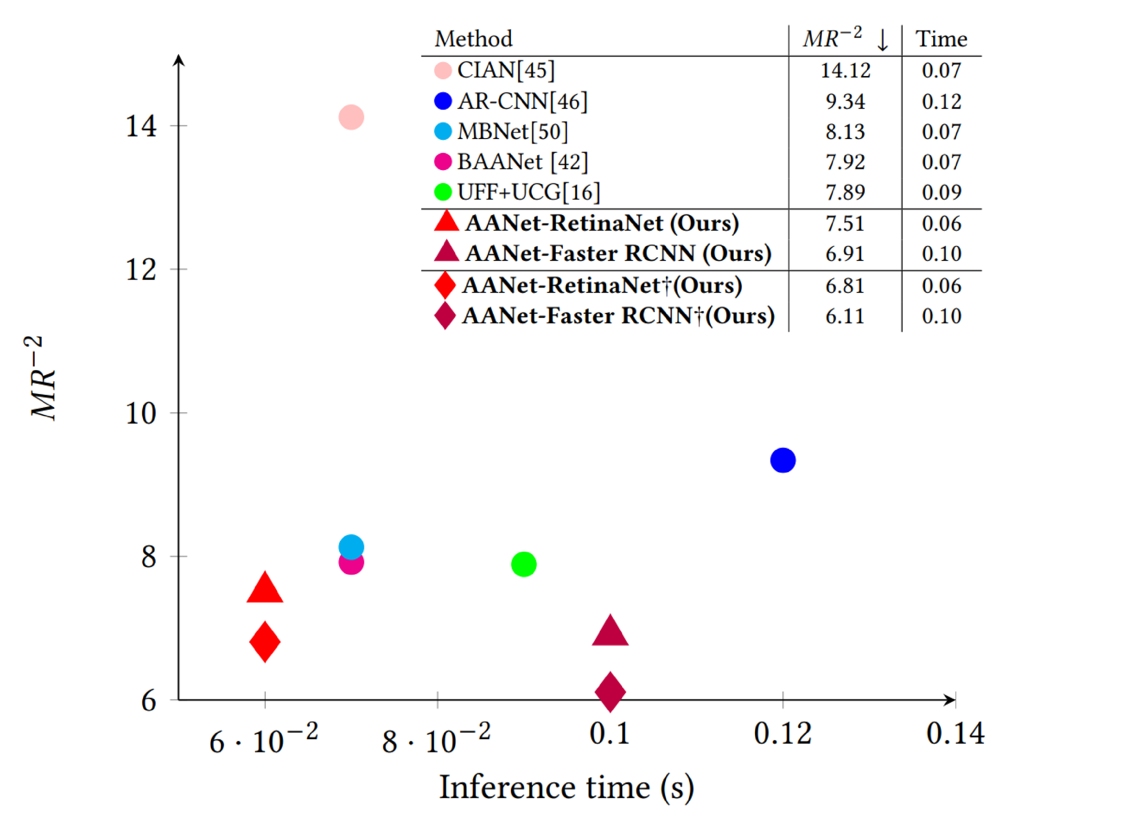
Methods	Backbone	GPU	Time	**All**	**Day**	**Night**
wo brightness distortion						
ACF [15]	-	-	2.73	47.32	42.57	56.17
Halfway Fusion [24]	VGG	Titan X	0.43	25.75	24.88	26.59
IAF-RCNN [19]	VGG	Titan X	0.21	15.73	14.55	18.26
IATDNN+IAMSS [11]	VGG	Titan X	0.25	14.95	14.67	15.72
CIAN [45]	VGG	1080Ti	0.07	14.12	14.77	11.13
MSDS-RCNN [18]	VGG	Titan X	0.22	11.34	10.53	12.94
AR-CNN [46]	VGG	1080Ti	0.12	9.34	9.94	8.38
DCRL [25]	VGG	2080Ti	0.18	9.16	9.86	8.18
MuFEm+SCoFA [5]	ResNeXt50	Tesla P6	0.10	8.07	8.16	7.51
UFF+UCG [16]	ResNet50	1080Ti	0.09	7.89	8.18	6.96
AANet-RetinaNet (ours)	ResNet50	1080Ti	**0.06**	7.51	7.74	7.39
AANet-Faster RCNN (ours)	ResNet50	1080Ti	0.10	6.91	6.66	7.31
w brightness distortion						
MBNet [50]	ResNet50	1080Ti	0.07	8.13	8.28	7.86
DCMNet-RetinaNet [38]	VGG16	Titan X	0.10	6.89	-	-
DCMNet-Faster RCNN [38]	VGG16	Titan X	0.14	6.41	-	-
AANet-RetinaNet† (ours)	ResNet50	1080Ti	**0.06**	6.81	6.72	6.59
AANet-Faster RCNN† (ours)	ResNet50	1080Ti	0.10	**6.11**	**5.94**	**6.37**

KAIST dataset

Detectors	Baseline	APA	AMA	**All**	**Day**	**Night**
RetinaNet	✓			9.62	9.39	10.06
	✓	✓		8.01	7.88	8.22
	✓	✓	✓	**7.51**	**7.74**	**7.39**
F.RCNN	✓			9.03	8.03	10.98
	✓	✓		7.37	6.81	7.80
	✓	✓	✓	**6.91**	**6.66**	**7.31**

Methods	Backbone	**All**	**Day**	**Night**
ACF [15]	-	60.10	61.30	48.20
Halfway Fusion [24]	VGG	31.99	36.29	26.29
AR-CNN [46]	VGG	22.10	24.70	18.10
MBNet [50]	ResNet50	21.10	24.70	13.50
UFF+UCG [16]	ResNet50	18.70	23.87	11.08
AANet-Faster RCNN (ours)	ResNet50	**17.88**	**22.61**	**10.86**

CVC-14 dataset



Performance comparisons