中国科学院大学
University of Chinese Academy of Sciences

Forground-Guided Vehicle Perception Framework

Kun Tian, Tong Zhou, Shiming Xiang, Chunhong Pan
{kun.tian, smxiang, chpan}@nlpr.ia.ac.cn, zhoutong18@mails.ucas.cn

Definition and Challenges

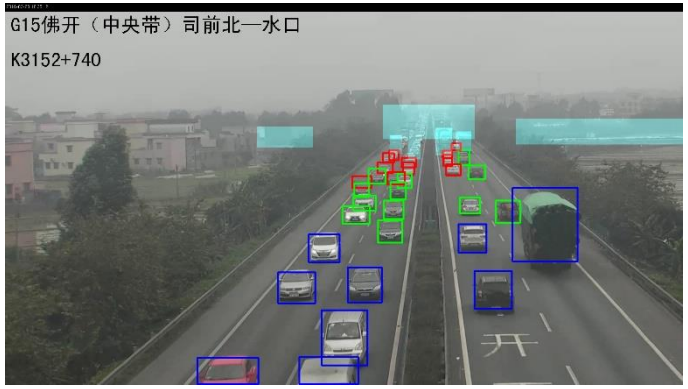
Definition

Input images

Detected Results

Challenges

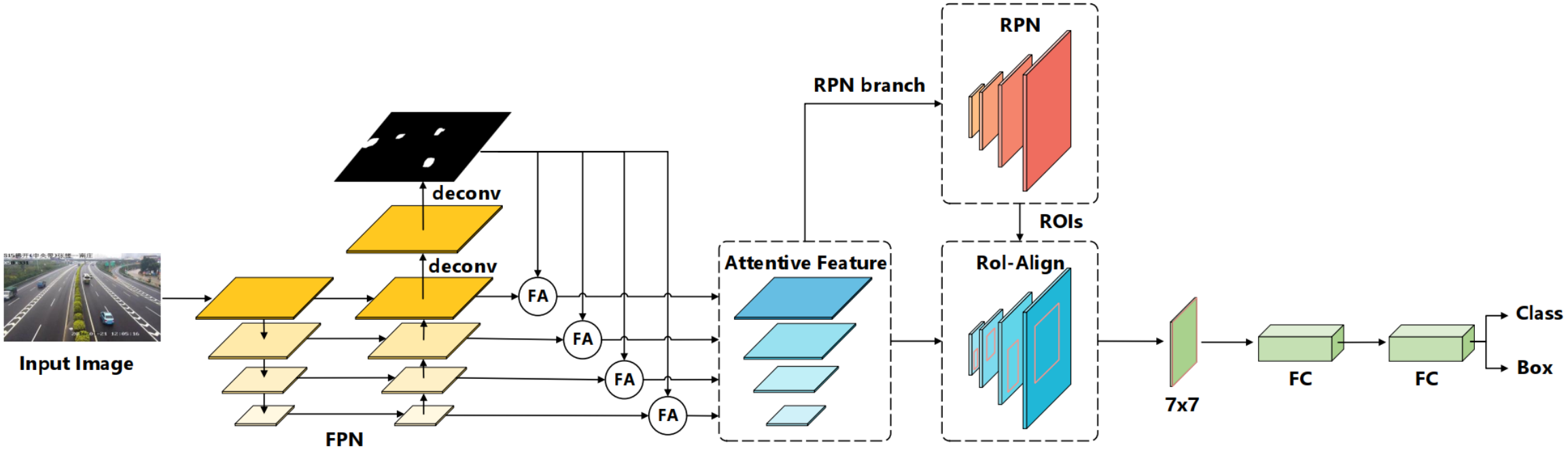
(a) False positive



(b) scale difference

University of Chinese Academy of Sciences

Our Method

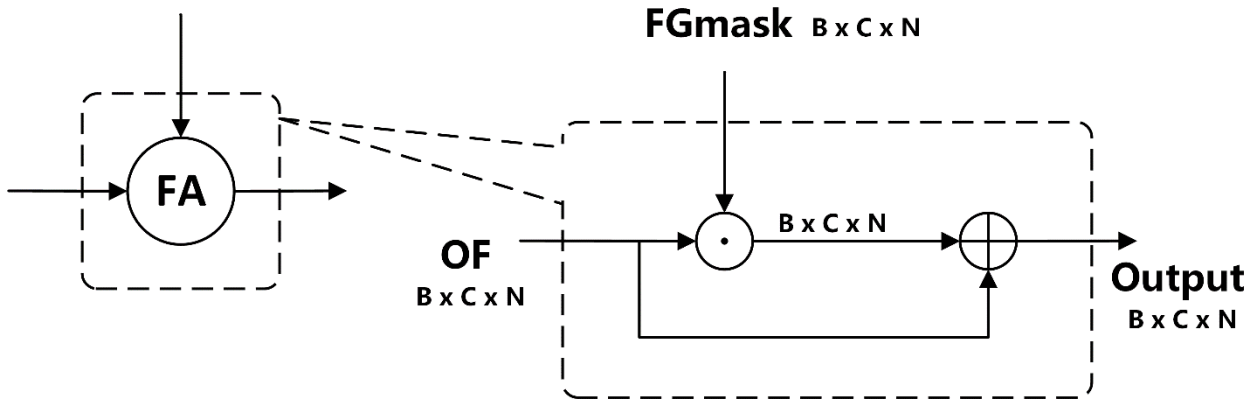


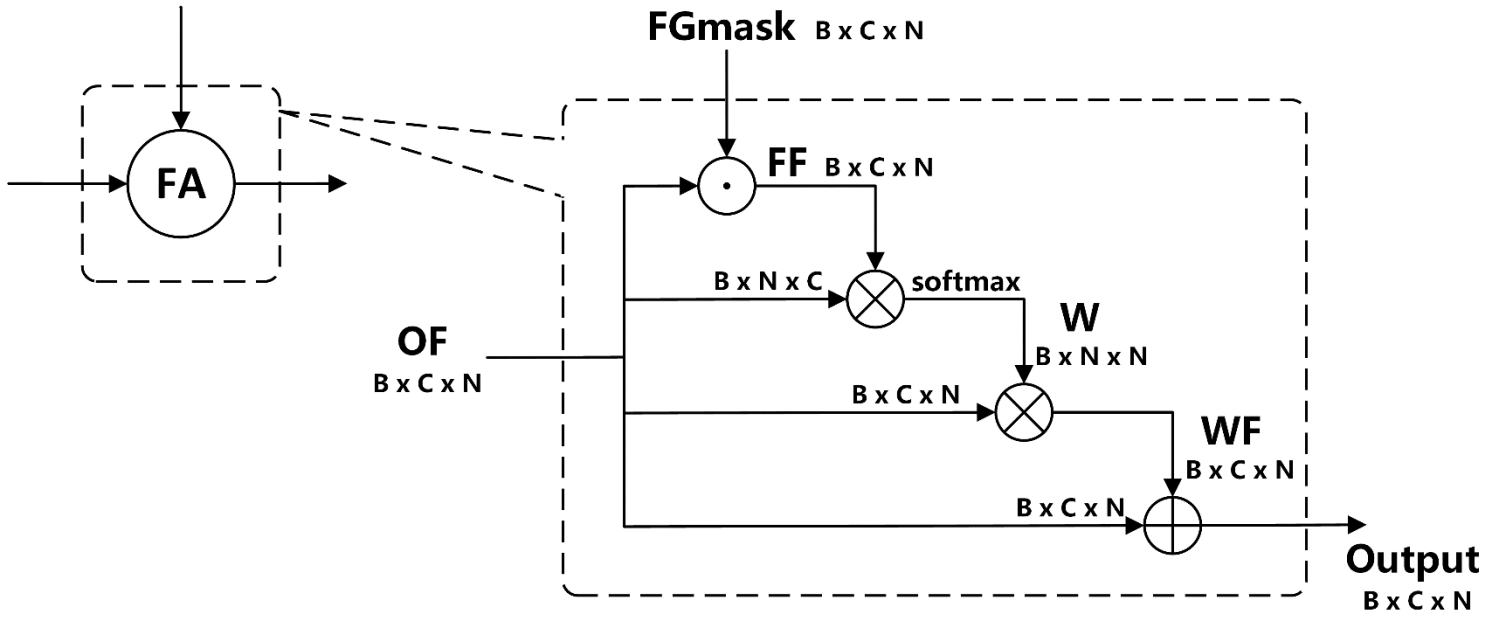
$$L = L_{seg} + L_{det} \tag{1}$$

$$L_{seg} = -\frac{1}{N}\sum_{i=1}^{N} y_i \cdot \log\left(p\left(y_i\right)\right) + \left(1 - y_i\right) \cdot \log\left(1 - p\left(y_i\right)\right) \tag{2}$$

$$L_{det} = L_{cls} + L_{reg} \tag{3}$$

Our Method





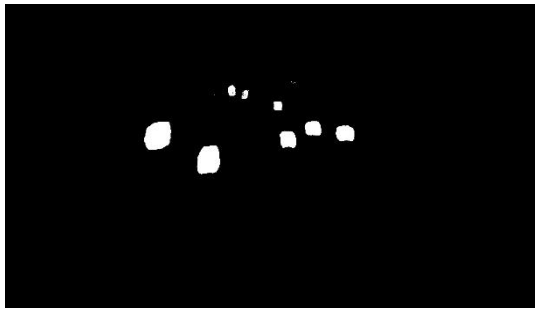
University of Chinese Academy of Sciences

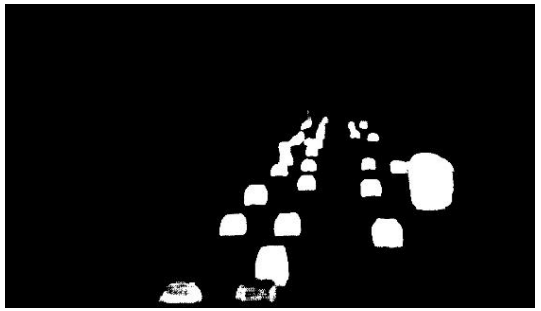
Experimental Results

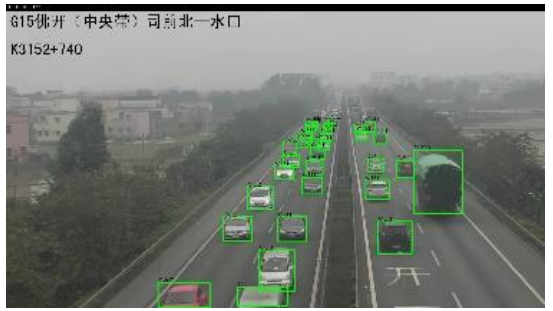
Model	Time/Image	Mean	Sparse			Crowded		
			car	bus	van	car	bus	van
YOLO	0.03	16.52	23.06	31.13	22.44	3.87	8.35	10.32
YOLOv2	0.03	43.82	59.71	65.51	58.35	17.39	21.55	40.42
RetinaNet	0.07	69.67	83.89	87.75	78.69	51.42	52.67	63.60
Faster R-CNN	0.31	46.43	60.93	66.68	60.14	26.08	24.55	40.24
MS-CNN	0.23	63.23	79.94	83.71	76.79	51.74	32.95	54.26
SINet	0.20	70.17	81.82	85.60	78.65	56.80	55.78	62.38
FPN	0.09	72.61	84.57	86.47	80.45	59.61	55.55	69.01
VPNet	0.11	74.57	84.96	88.77	81.03	59.76	59.12	73.80

Model	Time/Image	Mean	Sparse			Crowded		
			car	bus	van	car	bus	van
FPN-base	0.09	72.61	84.57	86.47	80.45	59.61	55.55	69.01
+FA(a)	0.11	74.35	85.19	87.06	82.98	59.00	59.43	72.45
+FA(b)	0.11	74.57	84.96	88.77	81.03	59.76	59.12	73.80
RetinaNet	0.07	69.67	83.89	87.75	78.69	51.42	52.67	63.60
+FA(a)	0.09	70.89	84.07	87.97	80.17	53.43	52.52	67.21
+FA(b)	0.09	70.98	84.23	87.33	80.30	52.71	55.23	66.09
Cascade R-CNN	0.13	72.70	85.07	87.50	80.04	60.05	55.16	68.42
+FA(a)	0.15	73.48	85.28	86.37	80.00	60.47	59.05	69.75
+FA(b)	0.15	73.73	85.52	87.42	82.12	60.28	57.00	70.09

Visualization of vehicle perception results.







University of Chinese Academy of Sciences

Conclusion

1.We first put forward using segmentation branch to assist detection task training, which can sense pixel position of the foreground vehicles in advance.

2.Two attention modules are designed to suppress the confidence scores of background regions, thus alleviating the impact of false alarms.

3. We verify the compatibility of our method on several classic single-stage and two-stage detection models.


中国科学院大学
University of Chinese Academy of Sciences
Thanks for listening!