Dual Attention Network for Scene Segmentation

Jun Fu Jing Liu Haijie Tian Yong Li Yongjun Bao
Zhiwei Fang Hanqing Lu

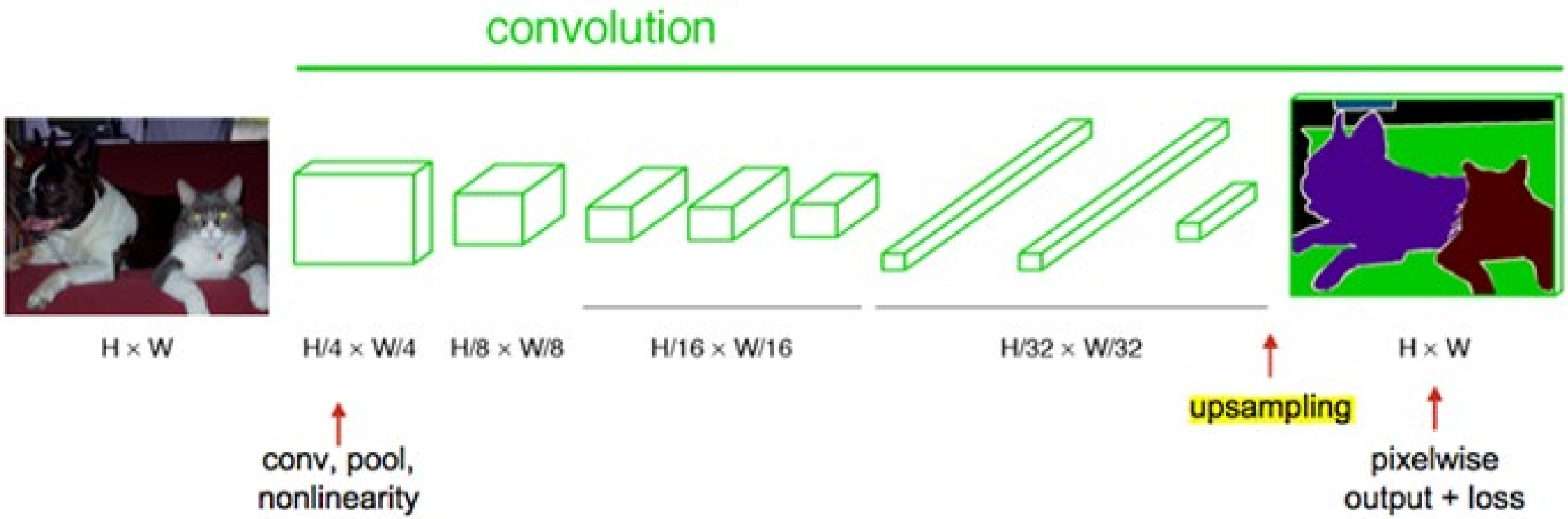
CVPR 2019

Contents

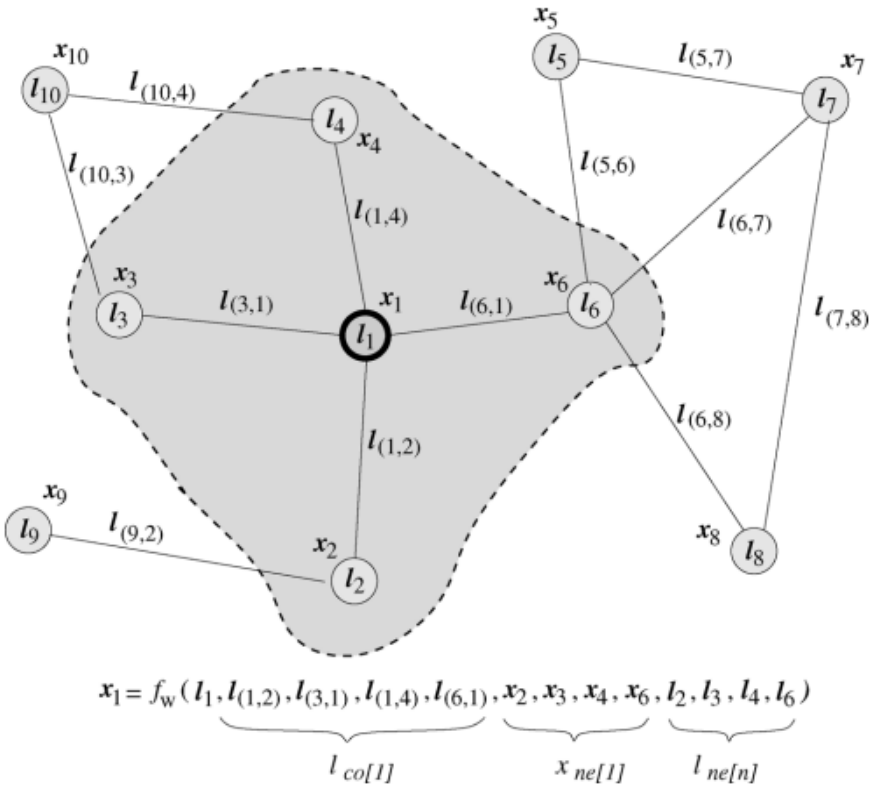
Goal

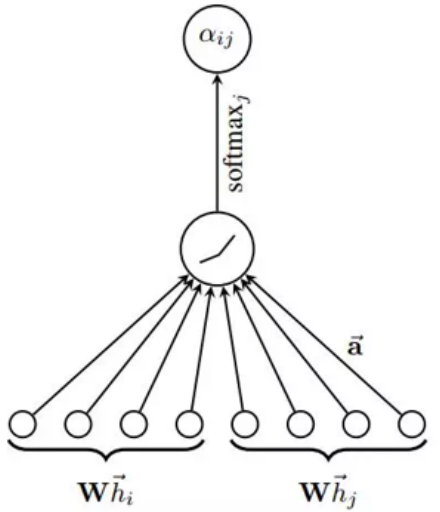
The goal of scene segmentation is to recognize each pixel including stuff, diverse objects

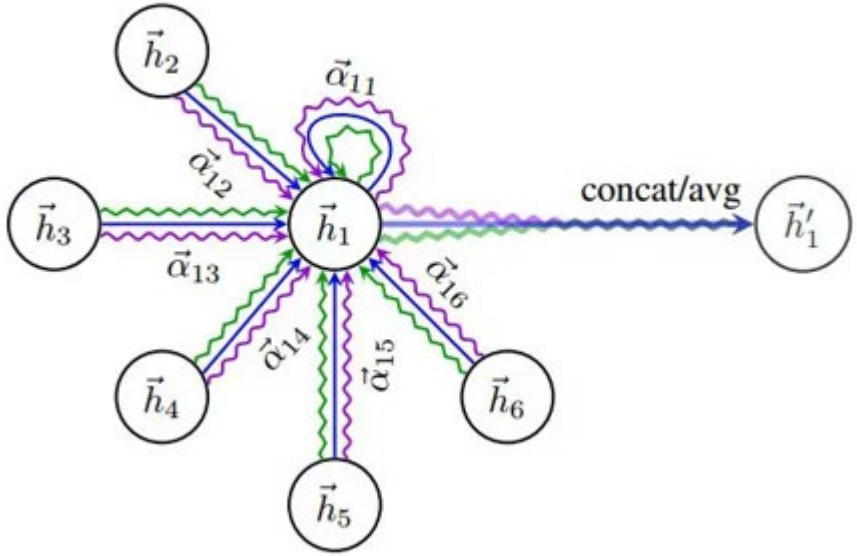
Related work： FCN(Fully Convolutional Networks)



Attention Network







Dual Attention Network

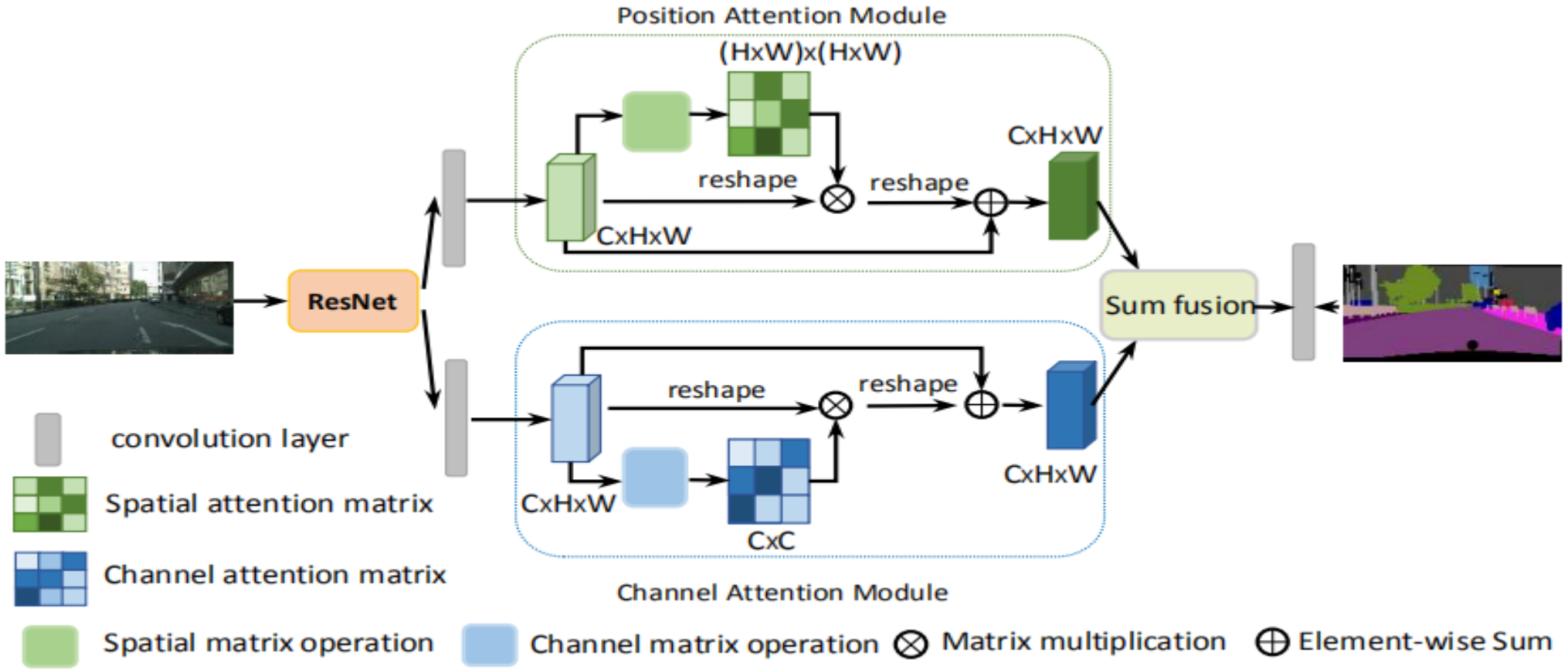
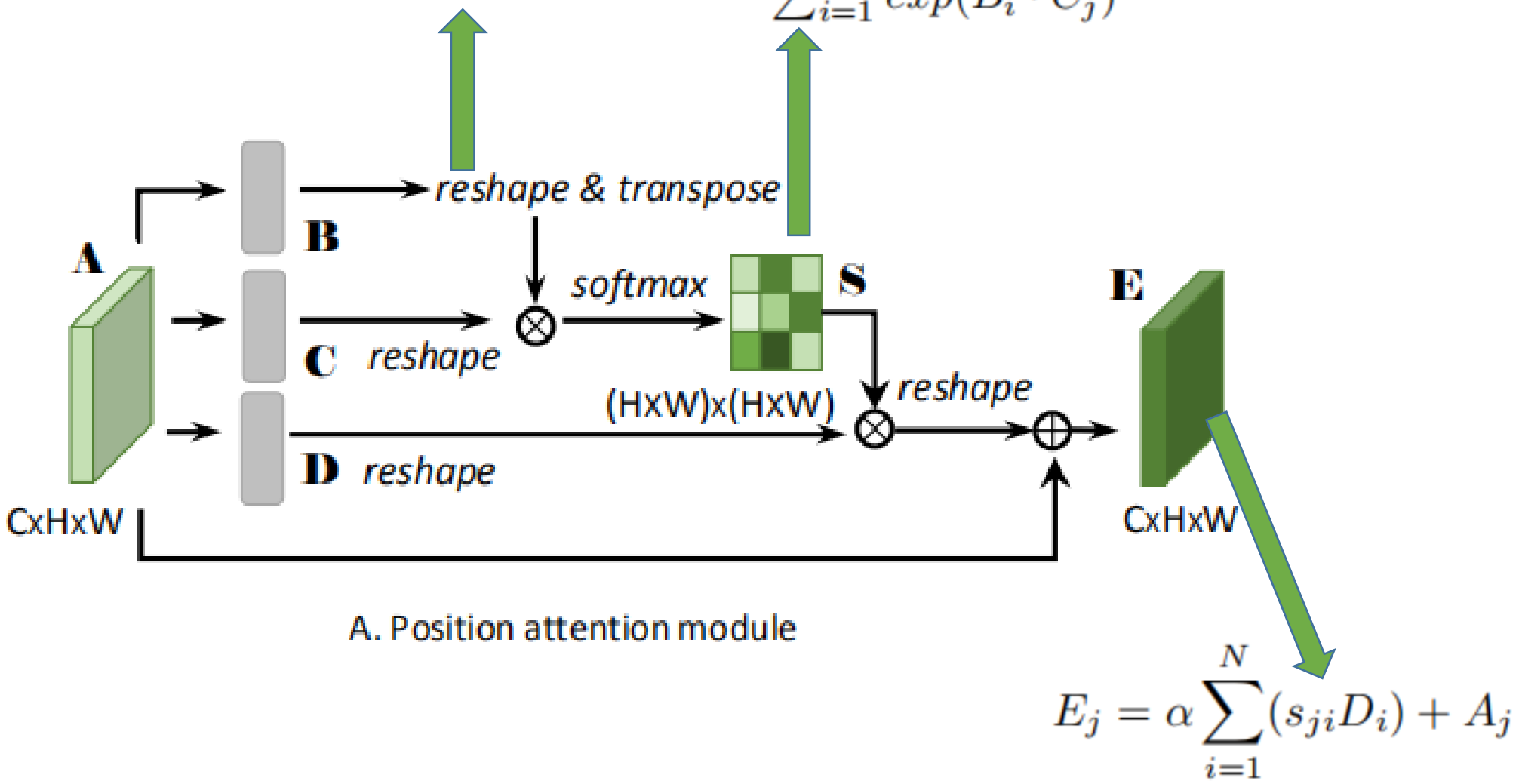


Figure 2: An overview of the Dual Attention Network. (Best viewed in color)

Position Attention Module

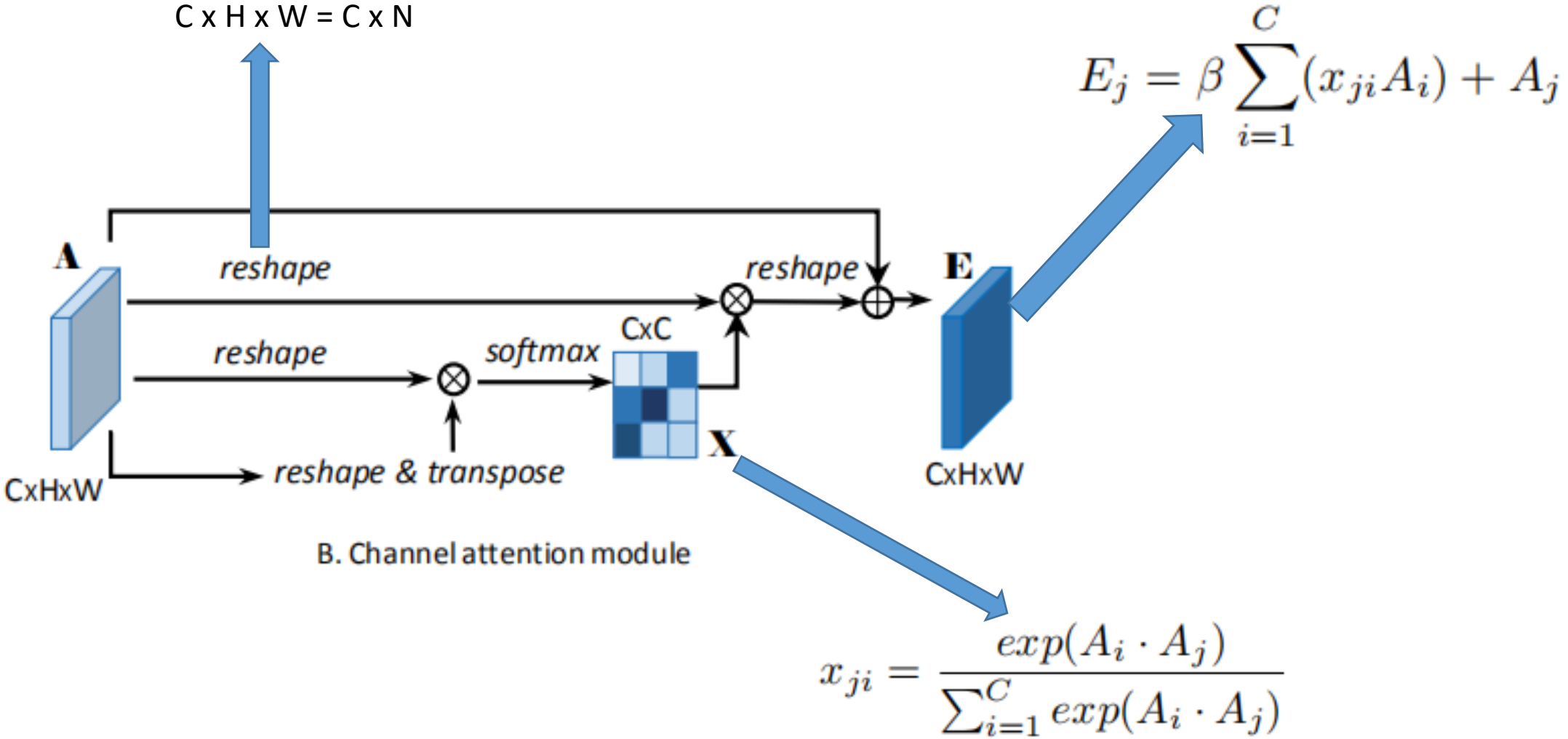
$$s_{ji} = \frac{exp(B_i \cdot C_j)}{\sum_{i=1}^{N} exp(B_i \cdot C_j)}$$



A. Position attention module

$$E_j = \alpha \sum_{i=1}^{N} (s_{ji} D_i) + A_j$$

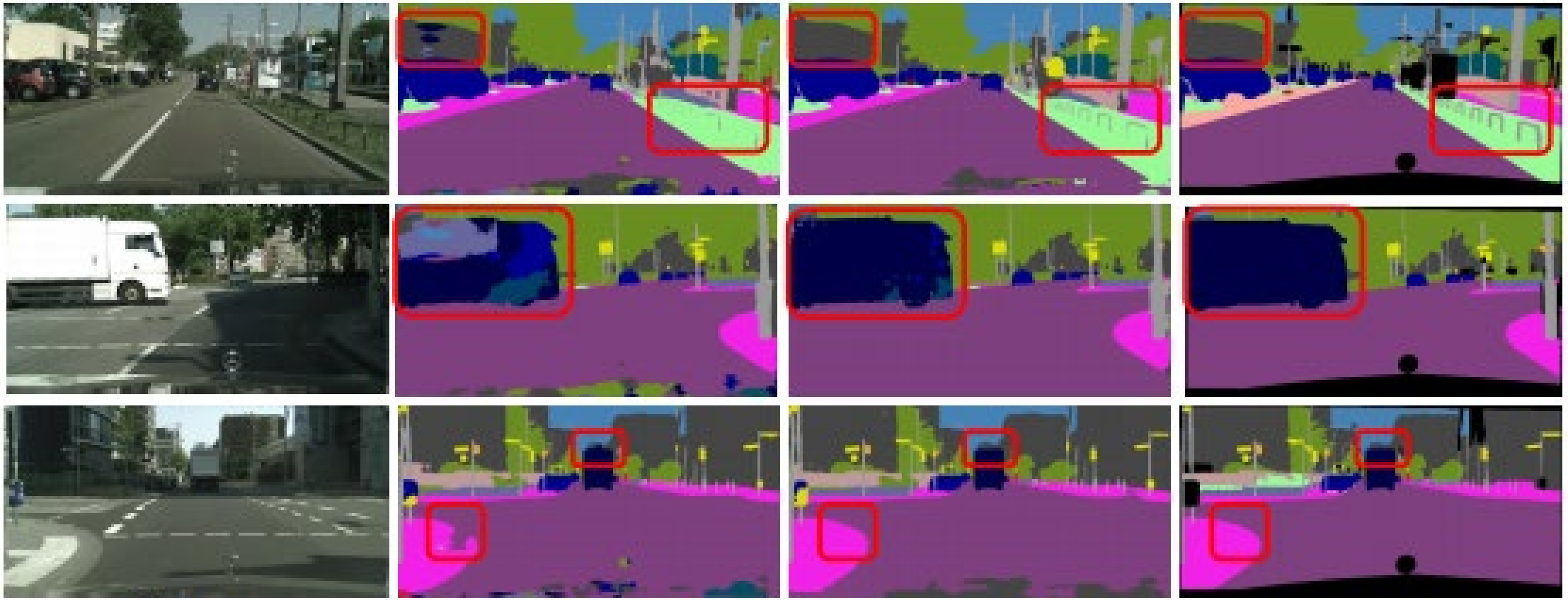
Channel Attention Module



$$E_j = \beta \sum_{i=1}^{C} (x_{ji} A_i) + A_j$$

$$x_{ji} = \frac{exp(A_i \cdot A_j)}{\sum_{i=1}^{C} exp(A_i \cdot A_j)}$$

Expriments



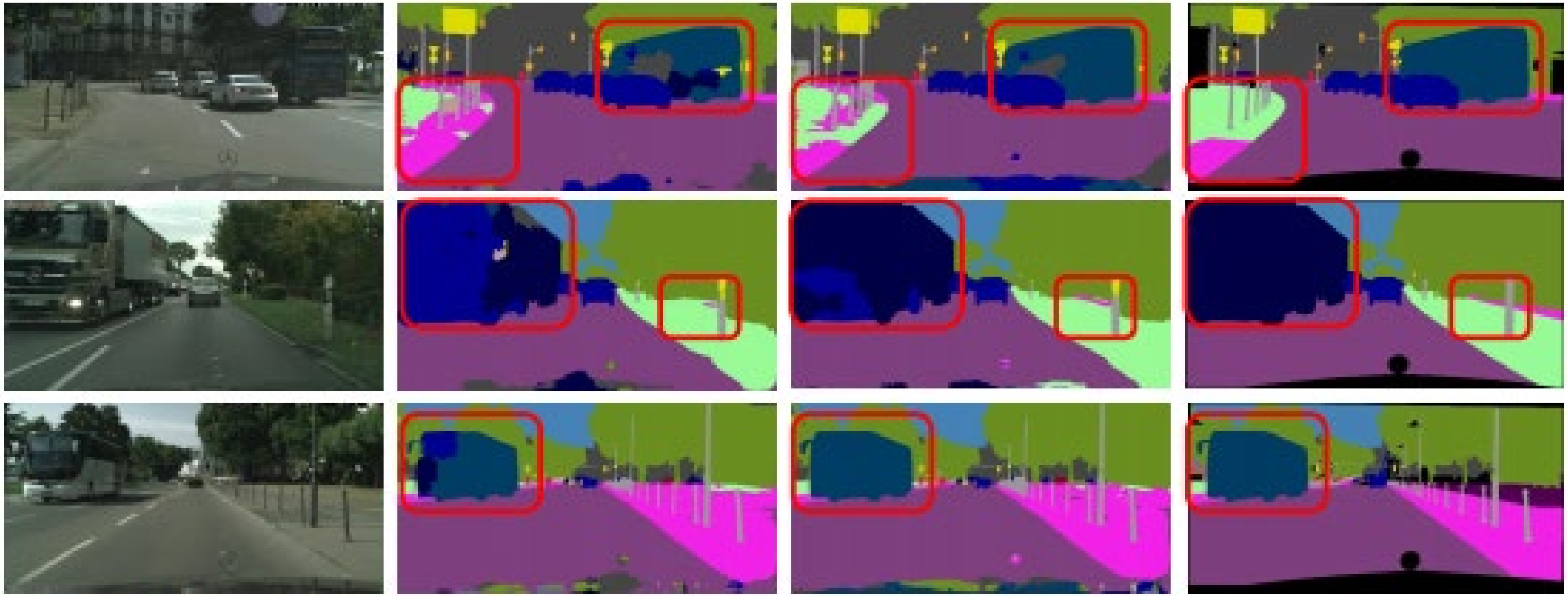


Image
Without CAM
With CAM
Groundtruth

Method	BaseNet	PAM	CAM	Mean IoU%
Dilated FCN	Res50			70.03
DANet	Res50	✓		75.74
DANet	Res50		✓	74.28
DANet	Res50	✓	✓	76.34
Dilated FCN	Res101			72.54
DANet	Res101	✓		77.03
DANet	Res101		✓	76.55
DANet	Res101	✓	✓	77.57

Table 1: Ablation study on Cityscapes val set. *PAM* represents Position Attention Module, *CAM* represents Channel Attention Module.

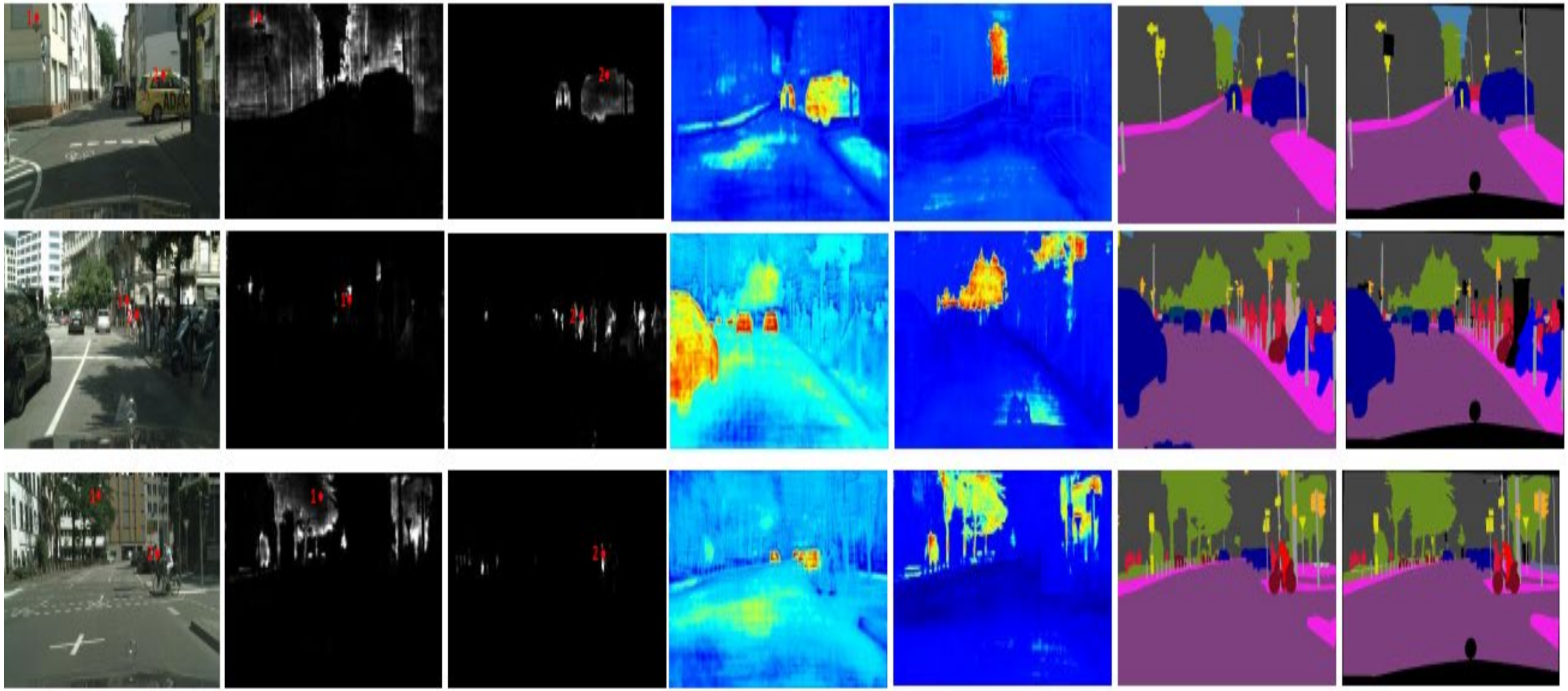


Image
Sub-attention map #1
Sub-attention map #2
Channel map #11
Channel map #4
Result
Groundtruth

Methods	Mean IoU	road	sidewalk	building	wall	fence	pole	traffic light	traffic sign	vegetation	terrain	sky	person	rider	car	truck	bus	train	motorcycle	bicycle
DeepLab-v2 [3]	70.4	97.9	81.3	90.3	48.8	47.4	49.6	57.9	67.3	91.9	69.4	94.2	79.8	59.8	93.7	56.5	67.5	57.5	57.7	68.8
RefineNet [10]	73.6	98.2	83.3	91.3	47.8	50.4	56.1	66.9	71.3	92.3	70.3	94.8	80.9	63.3	94.5	64.6	76.1	64.3	62.2	70
GCN [15]	76.9	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
DUC [22]	77.6	98.5	85.5	92.8	58.6	55.5	65	73.5	77.9	93.3	72	95.2	84.8	68.5	95.4	70.9	78.8	68.7	65.9	73.8
ResNet-38 [24]	78.4	98.5	85.7	93.1	55.5	59.1	67.1	74.8	78.7	93.7	72.6	95.5	86.6	69.2	95.7	64.5	78.8	74.1	69	76.7
PSPNet [29]	78.4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
BiSeNet [26]	78.9	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
PSANet [30]	80.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
DenseASPP [25]	80.6	**98.7**	**87.1**	93.4	**60.7**	62.7	65.6	74.6	78.5	93.6	72.5	95.4	86.2	71.9	96.0	**78.0**	**90.3**	80.7	69.7	76.8
DANet	**81.5**	98.6	86.1	**93.5**	56.1	**63.3**	**69.7**	**77.3**	**81.3**	**93.9**	**72.9**	**95.7**	**87.3**	**72.9**	**96.2**	76.8	89.4	**86.5**	**72.2**	**78.2**

Method	**Mean IoU%**
FCN-8s [13]	37.8
Piecewise [11]	43.3
DeepLab-v2 (Res101-COCO) [3]	45.7
RefineNet (Res152) [10]	47.3
PSPNet (Res101) [29]	47.8
Ding et al.(Res101) [6]	51.6
EncNet (Res101) [27]	51.7
Dilated FCN(Res50)	44.3
DANet (Res50)	50.1
DANet (Res101)	**52.6**

Table 6: Segmentation results on PASCAL Context testing set.

Method	**Mean IoU%**
FCN-8s [13]	22.7
DeepLab-v2(Res101) [3]	26.9
DAG-RNN [18]	31.2
RefineNet (Res101) [10]	33.6
Ding et al.(Res101) [6]	35.7
Dilated FCN (Res50)	31.9
DANet (Res50)	37.2
DANet (Res101)	**39.7**

Table 7: Segmentation results on COCO Stuff testing set.