

Multi-Direction Convolution for Semantic Segmentation

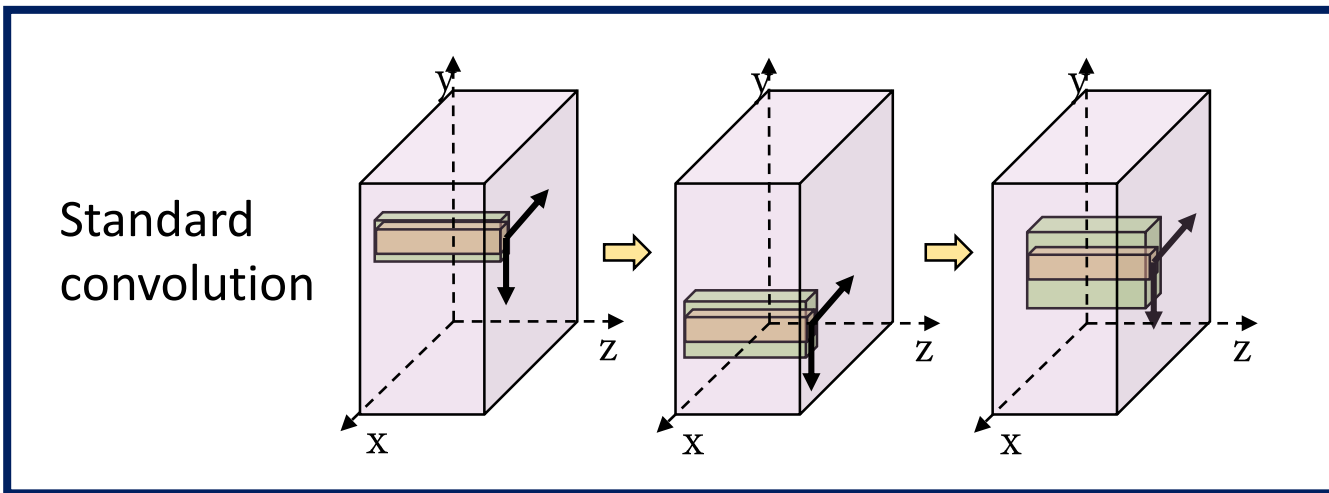
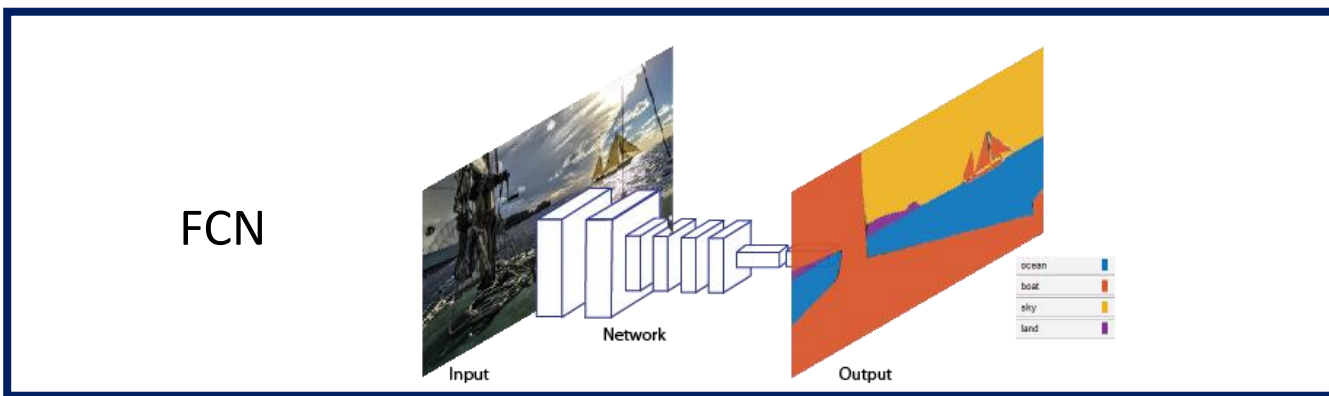
Dehui Li¹, Zhiguo Cao¹, Ke Xian^{1*}, Xinyuan Qi¹, Chao Zhang¹ and Hao Lu²

¹ Huazhong University of Science and Technology

² The University of Adelaide

➤ Introduction

- Standard convolution is lack of context information



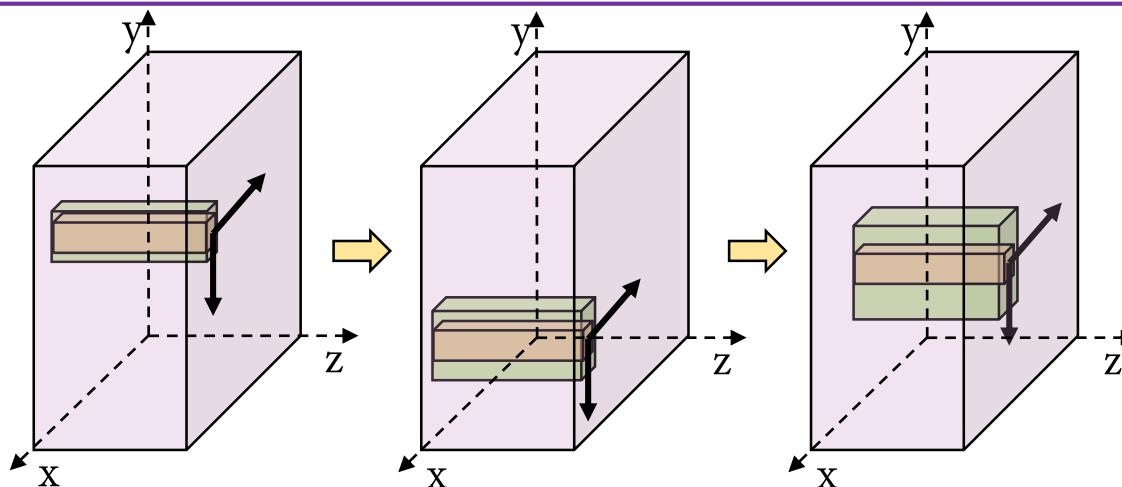
➤ Introduction

Existing context encode method

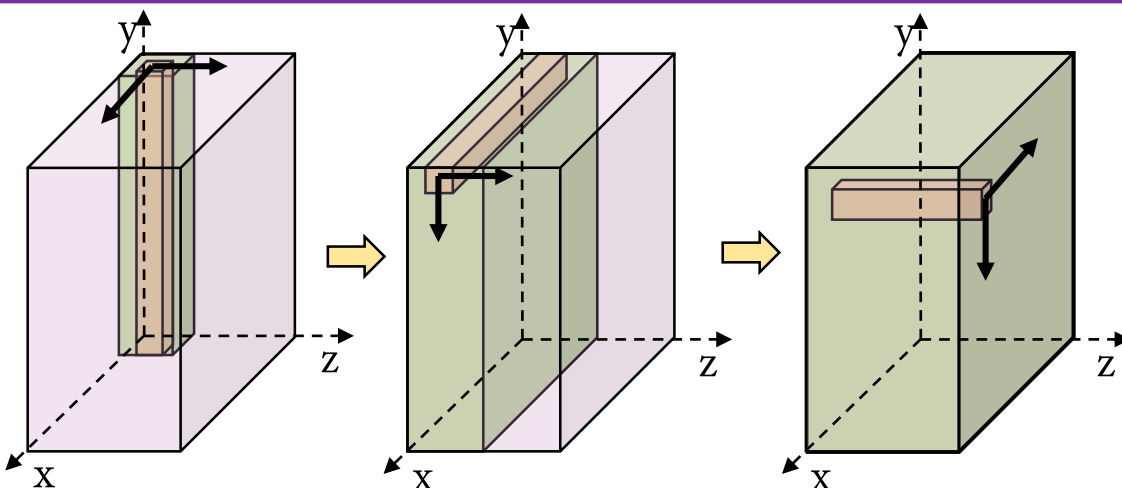
- dilated convolution
 - large kernel
- } Still local operation
-
- Global pooling
 - PSP
- } unadaptable
-
- Self-attention → inefficient

➤ Method

Standard
convolution

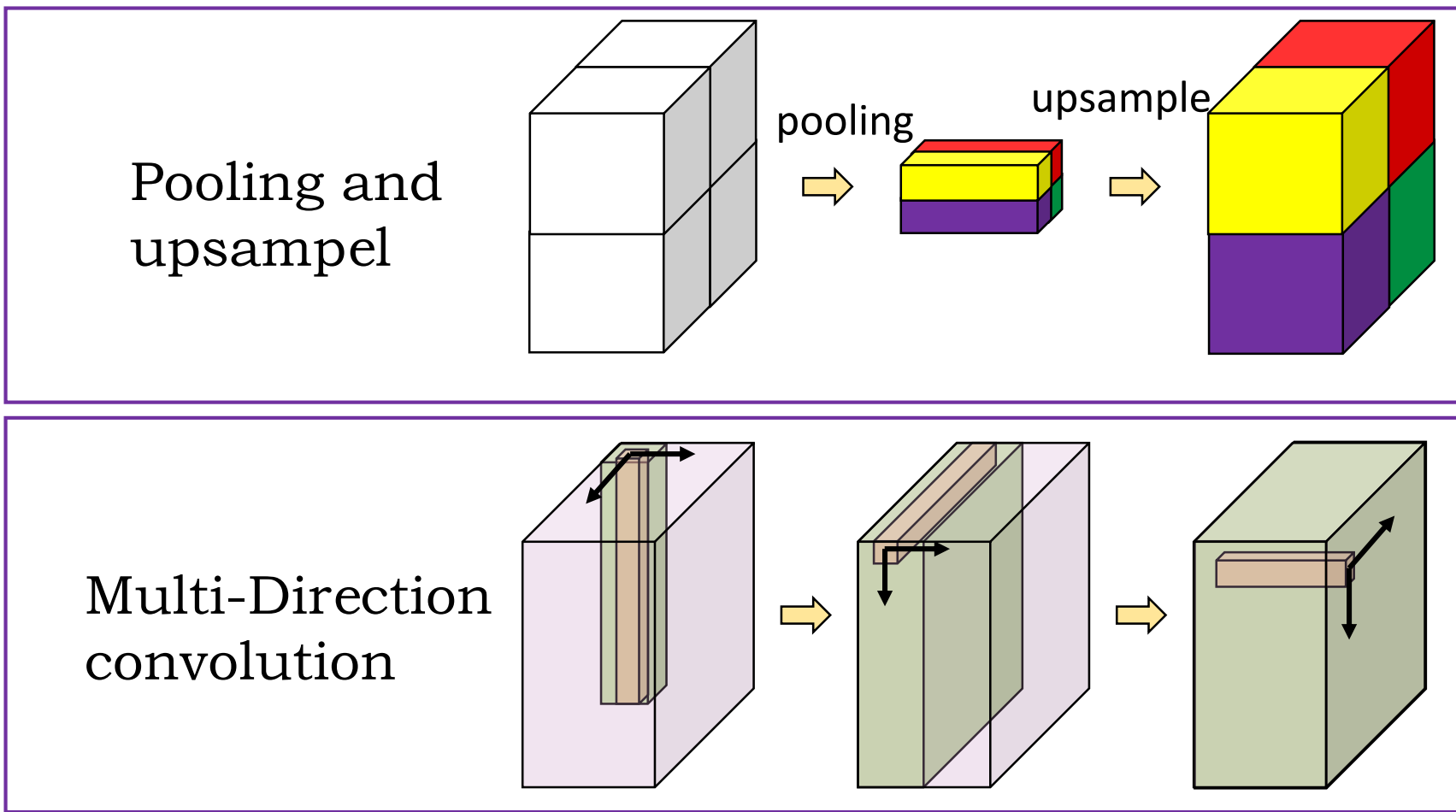


Multi-Direction
convolution



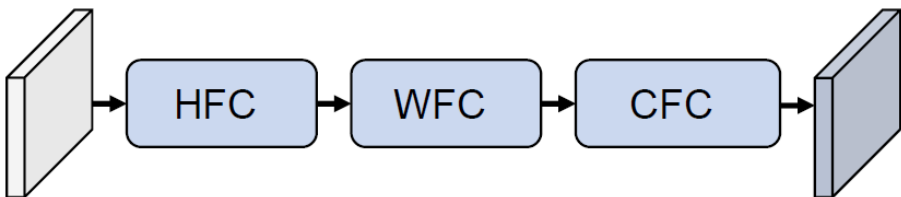
➤ Method

Different way to encode contextual information

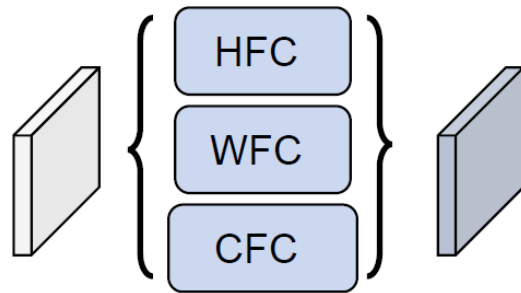


➤ Method

Different variants of multi-direction convolution modules



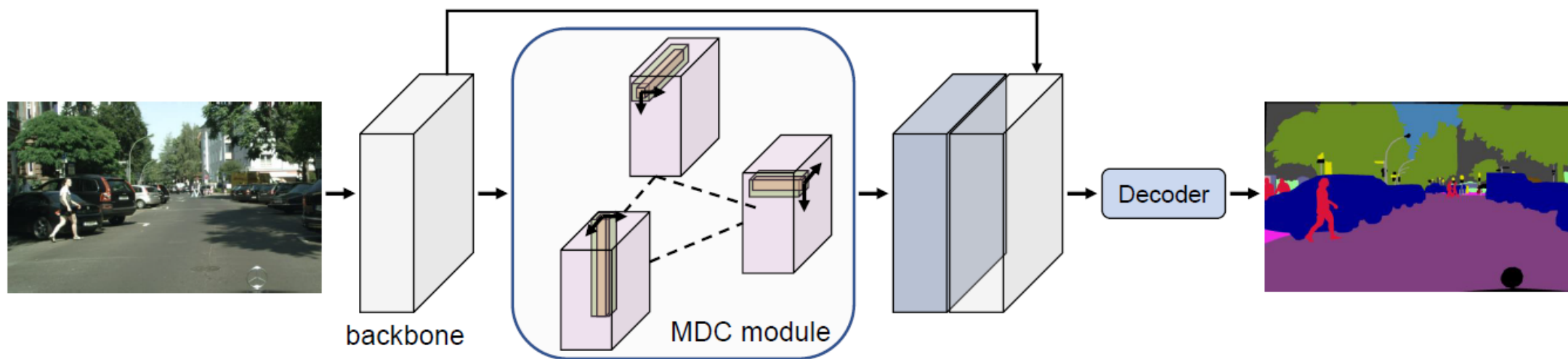
(a) MDC in cascade



(b) MDC in parallel

➤ Method

Overview of network embedding multi-direction convolution module



➤ Experiments

Compared segmentation results on Cityscapes validation set

Method	mIoU%	FLOPs	Params
ResNet-18 (baseline)	61.97	103.29G	11.85M
baseline + global	63.94	+0.48G	+0.33M
baseline + PSP [3]	68.02	+1.03G	+0.65M
baseline + ASPP [4]	68.04	+30.81G	+3.90M
baseline + dsASPP [26]	69.51	+36.20G	+4.44M
baseline + OC [8]	67.63	+22.90G	+2.67M
baseline + DA [9]	68.30	+11.78G	+1.45M
baseline + MDCM	69.04	+8.53G	+1.12M

➤ Experiments

Compared segmentation results on VOC2012 validation set

Method	mIoU%	FLOPs	Params
ResNet-50 (baseline)	66.60	104.41G	25.91M
baseline + global	67.17	+0.28G	+1.12M
baseline + PSP [3]	68.54	+0.60G	+2.23M
baseline + ASPP [4]	68.14	+59.74G	+15.11M
baseline + dsASPP [23]	68.50	+81.90G	+20.03M
baseline + OC [8]	67.99	+39.67G	+9.56M
baseline + DA [9]	68.48	+97.18G	+23.73M
baseline + MDCM	69.66	+14.18G	+3.47M

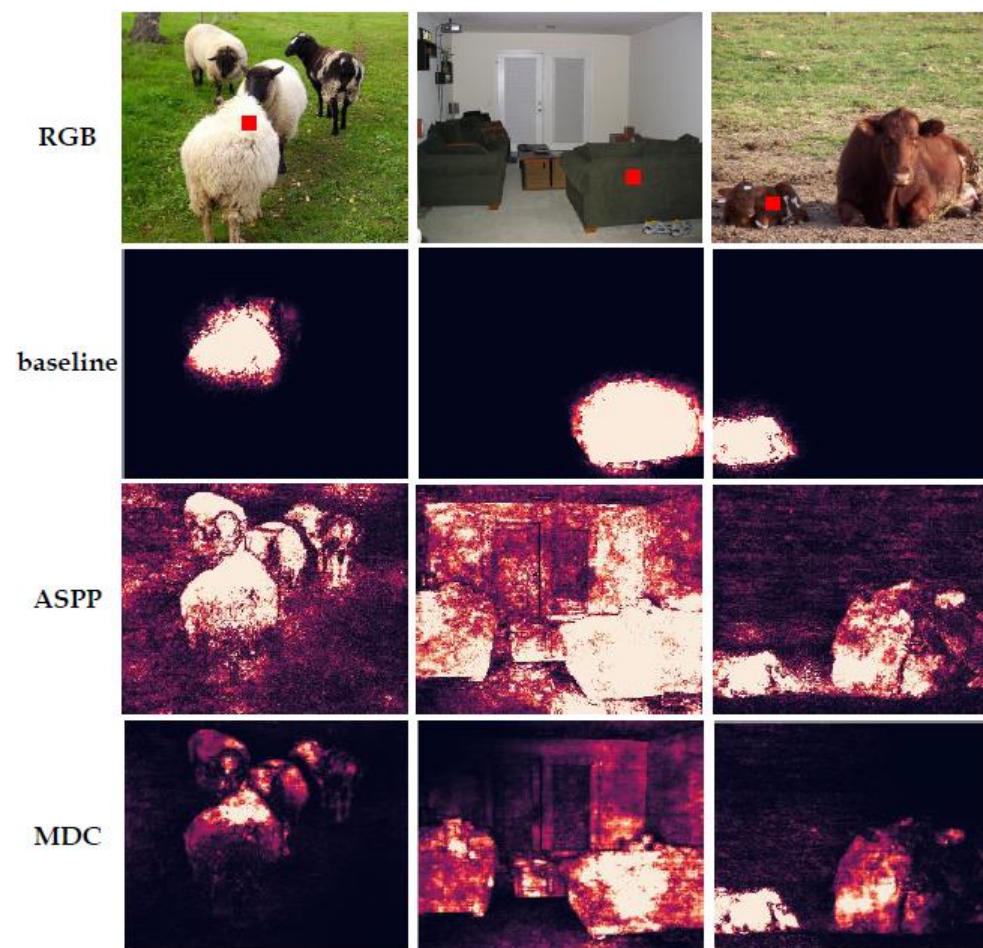
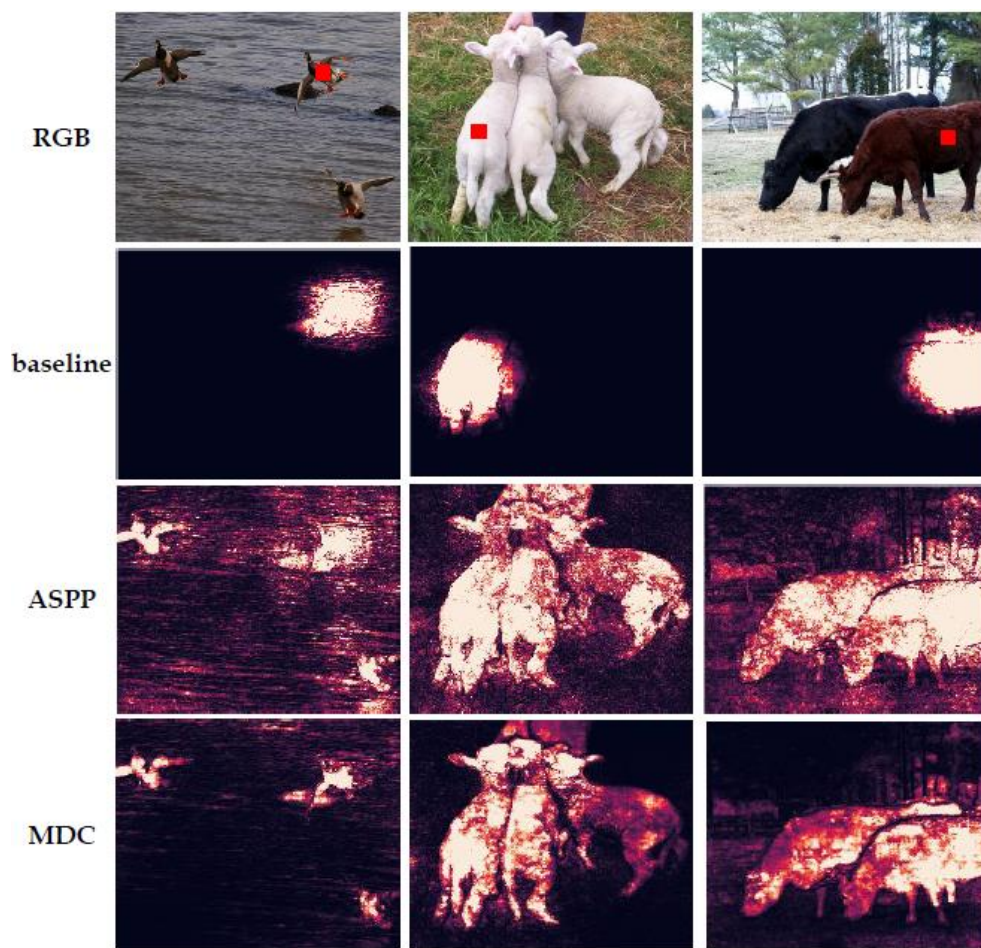
➤ Experiments

Different kernel sizes in multi-direction convolution modules

Kernel size	Mean IoU%	FLOPs	Params
(baseline)	61.97	103.29G	11.85M
1×1	68.50	+7.45G	+0.92M
3×3	69.04	+8.53G	+1.12M
5×5	69.21	+10.67G	+1.51M
7×7	69.53	+13.90G	+2.10M

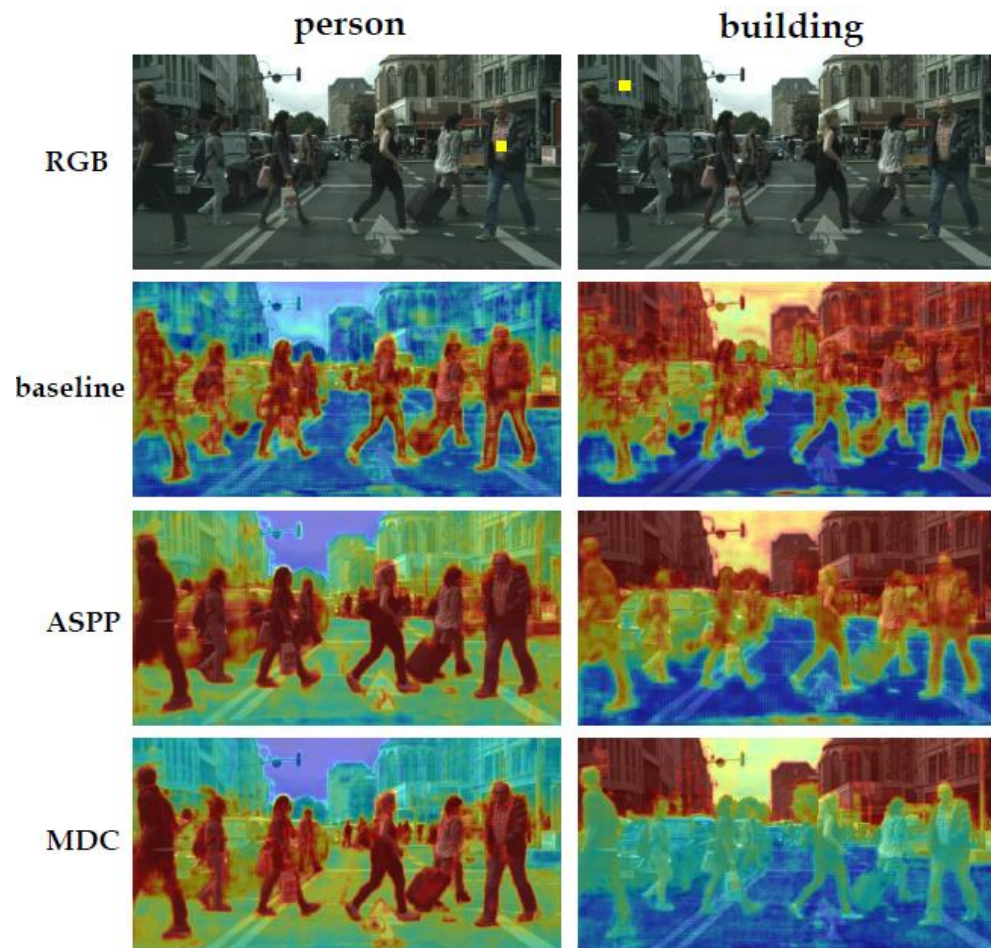
➤ Further Analysis and Discussions

Example of empirical receptive fields.



➤ Further Analysis and Discussions

Example of affinity maps.



➤ Conclusion

Multi-Direction Convolution is able to enlarge the receptive field and encode rich contextual information.

Multi-Direction Convolution is both effective and efficient compare with existing methods.



Thank you!