Frequency-Adaptive Dilated Convolution for Semantic Segmentation

Linwe Chen[1] Lin Gu[2,3] Dezhi Zheng[1] Ying Fu[1*]

[1] Beijing Institute of Technology, [2]The University of Tokyo, [3]RIKEN AIP

CVPR JUNE 17-21, 2024 SEATTLE, WA

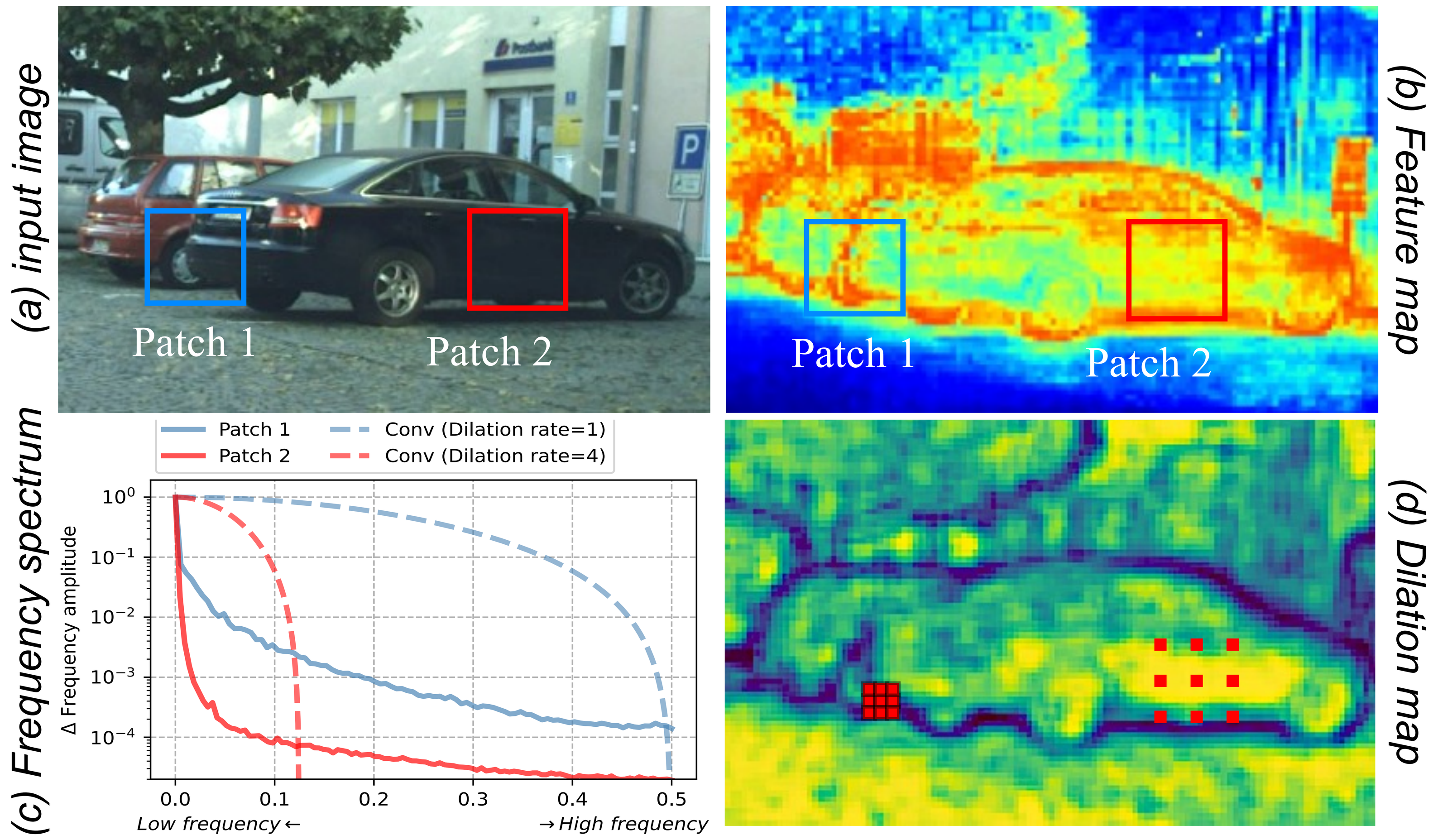
Author WeChat

GitHub Code

Introduction

Motivation

- Dilated convolution expands the receptive field without significantly increasing computational load, widely used in semantic segmentation and object detection.
- However, high dilation rates limit the capture of high frequencies, leading to gridding artifacts and reduced performance.
- Increasing the dilation rate from 1 to D expands the convolution kernel by through zero-insertion. This scaling reduces the frequency response and bandwidth of the kernel to 1/D.



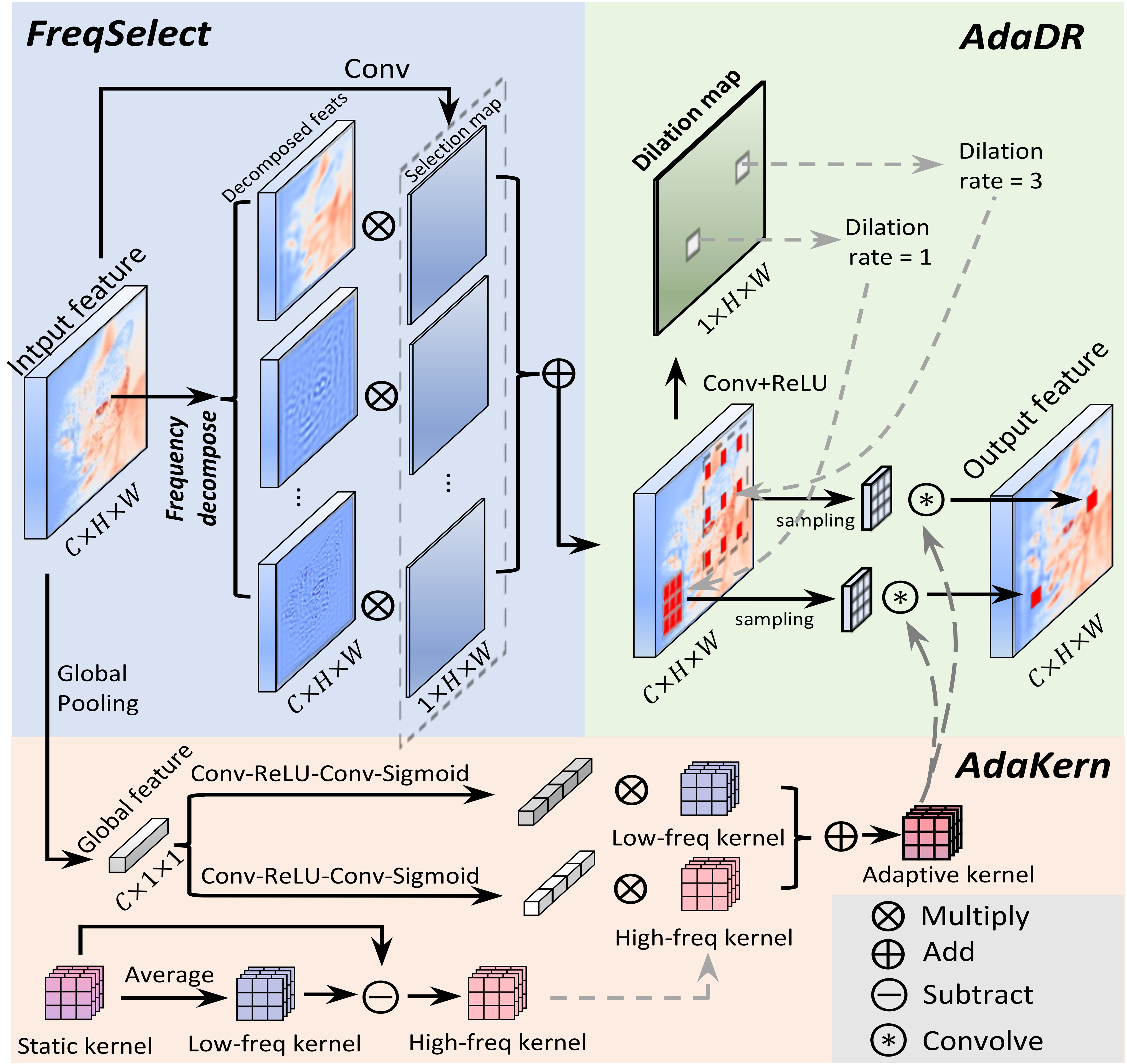
The frequency distribution is spatially variant, the dilation rate should be adjusted to adapt to it.

Contributions

- **In-depth analysis.** We conduct an in-depth exploration of dilated convolution using frequency analysis, reframing the assignment of dilation as a trade-off problem that involves balancing effective bandwidth and receptive field.
- **New method.** We introduce FADC to balance effective bandwidth and receptive field for each position.
- **Consistent improvement.** We validate our approach through comprehensive experiments in the semantic segmentation.

Our Approach

Frequency-Adaptive Dilated Convolution (FADC)



Adaptive Dilation Rate (AdaDR)

$$\mathbf{Y}(p) = \sum_{i=1}^{K\times K} \mathbf{W}_i \mathbf{X}(p + \Delta p_i \times \hat{\mathbf{D}}(p))$$

Dynamically adjusts dilation rates based on the feature frequency content. Small/ Large dilation rates for high/low-frequency patches.

Adaptive Kernel (AdaKern)

$$\mathbf{W} = \bar{\mathbf{W}} + \hat{\mathbf{W}}$$
$$\mathbf{W}' = \lambda_l \bar{\mathbf{W}} + \lambda_h \hat{\mathbf{W}}$$

Decomposes convolution weights into low- and high-frequency parts. Dynamically adjusts low&high components to optimize frequency response.

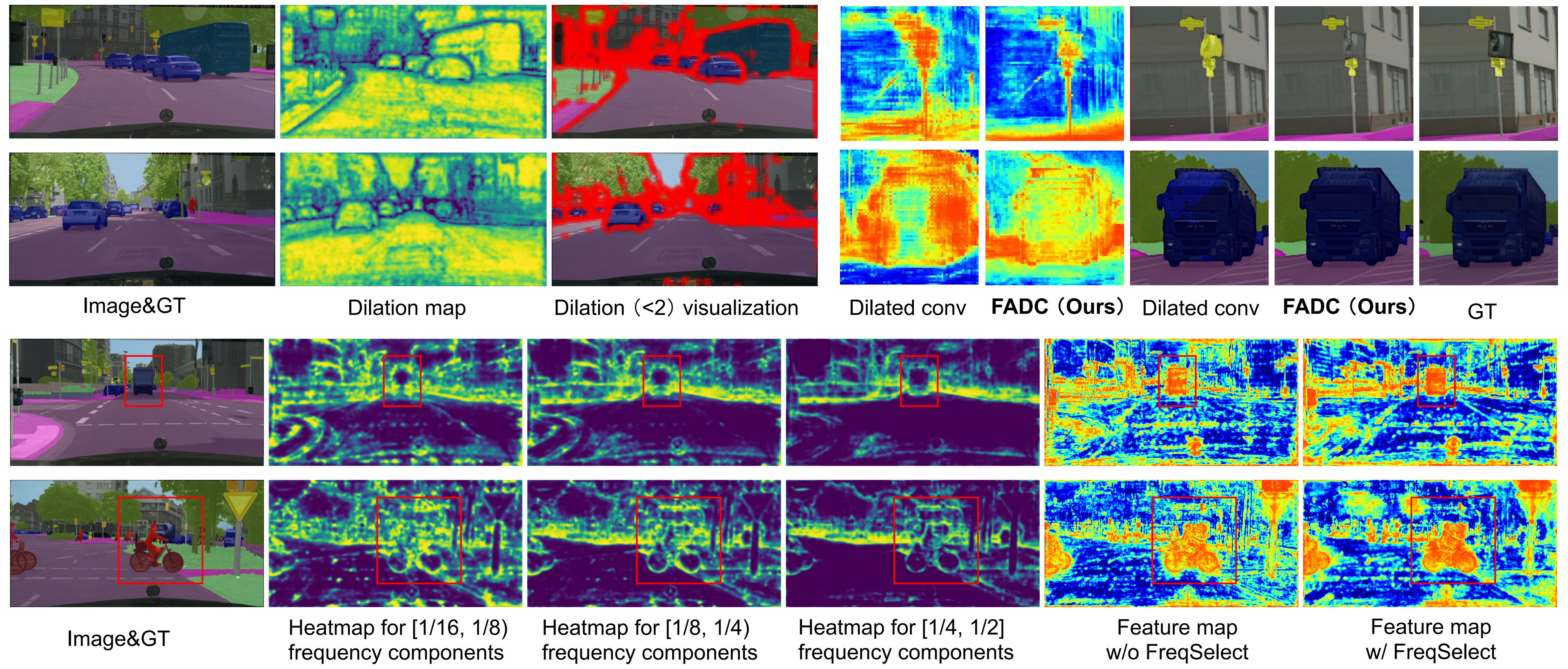
Frequency Selection (FreqSelect)

$$\hat{\mathbf{X}}(i,j) = \sum_{b=0}^{B-1} \mathbf{A}_b(i,j)\mathbf{X}_b(i,j)$$

Decomposes features into different frequency bands and dynamically weights them. It enhances the receptive field by suppressing high-frequency power.

Experimental Results

Visualized Results



Quantitative Results

Tab.1 *Semantic segmentation on Cityscapes.*

Method	#Params	#FLOPS	mIoU
Backbone: Dilated-ResNet-50 [82]			
PSPNet [88]	49.0M	1427.5G	77.8
PSPNet [88] + DCNv2 [91]	+0.7M	+24.5G	79.7
PSPNet [88] + FADC (Ours)	+0.5M	+9.2G	**80.4**
Backbone: Dilated-ResNet-101 [82]			
DeepLabV3+ [9] + ADC [78]	62.8M	2032.3G	80.7
DeepLabV3+ [9] + FADC (Ours)	63.9M	2067.0G	**81.5**
Backbone: ResNet-50 [25]			
Mask2Former [10]	44.0M	-	79.4
Mask2Former [10] + DCNv2 [91]	+0.9M	+7.7G	80.4
Mask2Former [10] + FADC (Ours)	+0.5M	+4.3G	**80.6**

Tab.3 *Semantic segmentation on ADE20k.*

Method	#Params	#FLOPS	mIoU SS	mIoU MS
ResNet-50 [25]	66M	947G	40.7	41.8
ResNet-101 [25]	85M	1029G	42.9	44.0
ResNet-50-FADC (Ours)	67M	949G	**44.4**	**45.5**
Swin-B [43]	121M	1188G	48.1	49.7
NAT-B [23]	123M	1137G	48.5	49.7
ConvNeXt-B [44]	122M	1170G	49.1	49.9
ConvNeXt-B-dcls [31]	122M	1170G	49.3	-
DAT-B [73]	121M	1212G	49.4	50.6
DiNAT-B [22]	123M	1137G	49.6	50.4
Focal-B [77]	126M	1354G	49.0	50.5
InternImage-B [70]	128M	1185G	50.8	51.3
HorNet-B [54]	126M	1171G	50.5	50.9
HorNet-B-FADC (Ours)	128M	1176G	**51.1**	**51.5**

Tab. 2 *Real-time Semantic segmentation on Cityscapes.*

Model	#Params	#FLOPs	#FPS	Val	Test
DF2-Seg1 [35]	-	-	67.2	75.9	74.8
DF2-Seg2 [35]	-	-	56.3	76.9	75.3
SwiftNetRN-18 [48]	11.8M	104.0G	39.9	75.5	75.4
SwiftNetRN-18 ens [48]	24.7M	218.0G	18.4	-	76.5
CABiNet [32]	2.64M	12.0G	76.5	76.6	75.9
BiSeNet(Res18)[81]	49M	55.3G	65.5	74.8	74.7
BiSeNetV2-L[80]	-	118.5G	47.3	75.8	75.3
STDC1-Seg75 [16]	-	-	74.8	74.5	75.3
STDC2-Seg75 [16]	-	-	58.2	77.0	76.8
PP-LiteSeg-T2 [50]	-	-	96.0	76.0	74.9
PP-LiteSeg-B2 [50]	-	-	68.2	78.2	77.5
HyperSeg-M [47]	10.1M	7.5G	59.1	76.2	75.8
HyperSeg-S [47]	10.2M	17.0G	45.7	78.2	78.1
SFNet(DF2)[34]	10.53M	-	87.6	-	77.8
SFNet(ResNet-18)[34]	12.87M	247.0G	30.4	-	78.9
SFNet(ResNet-18)† [34]	12.87M	247.0G	30.4	-	80.4
DDRNet-23-S[26]	5.7M	36.3G	108.1	77.8	77.4
DDRNet-23 [26]	20.1M	143.1G	51.4	79.5	79.4
DDRNet-39 [26]	32.3M	281.2G	30.8	-	80.4
PIDNet-S [76]	7.6M	47.6G	93.2	78.8	78.6
PIDNet-M [76]	34.4M	197.4G	39.8	80.1	80.1
PIDNet-L [76]	36.9M	275.8G	31.1	80.9	80.6
PIDNet-M-FADC (Ours)	34.6M	198.4G	37.7	**81.0**	**80.6**