DeepLab V3

2017 CVPR

Rethinking Atrous Convolution for Semantic Image Segmentation

Liang-Chieh Chen George Papandreou Florian Schroff Hartwig Adam
Google Inc.
{lcchen, gpapan, fschroff, hadam}@google.com

- 引入了Multi-grid
- 改进ASPP结构
- 移除CRFs后处理

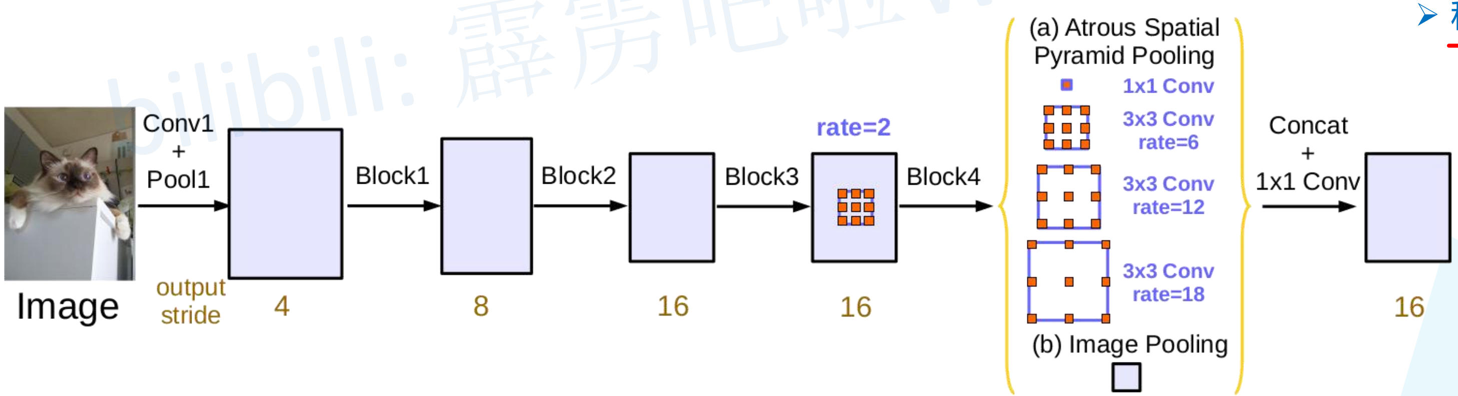
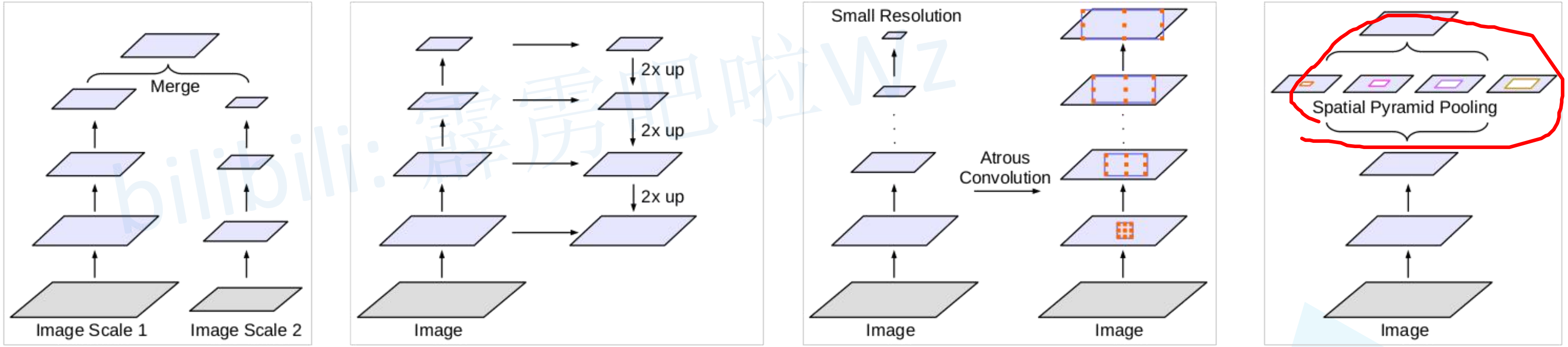


Figure 5. Parallel modules with atrous convolution (ASPP), augmented with image-level features.

论文下载地址：https://arxiv.org/abs/1706.05587

博文推荐：https://blog.csdn.net/qq_37541097/article/details/121797301

DeepLab V3



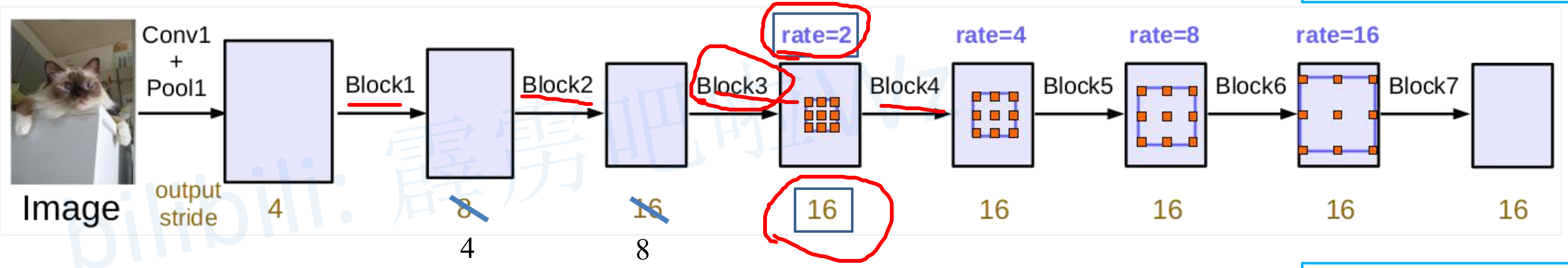
(a) Image Pyramid (b) Encoder-Decoder (c) Deeper w. Atrous Convolution (d) Spatial Pyramid Pooling

Figure 2. Alternative architectures to capture multi-scale context.

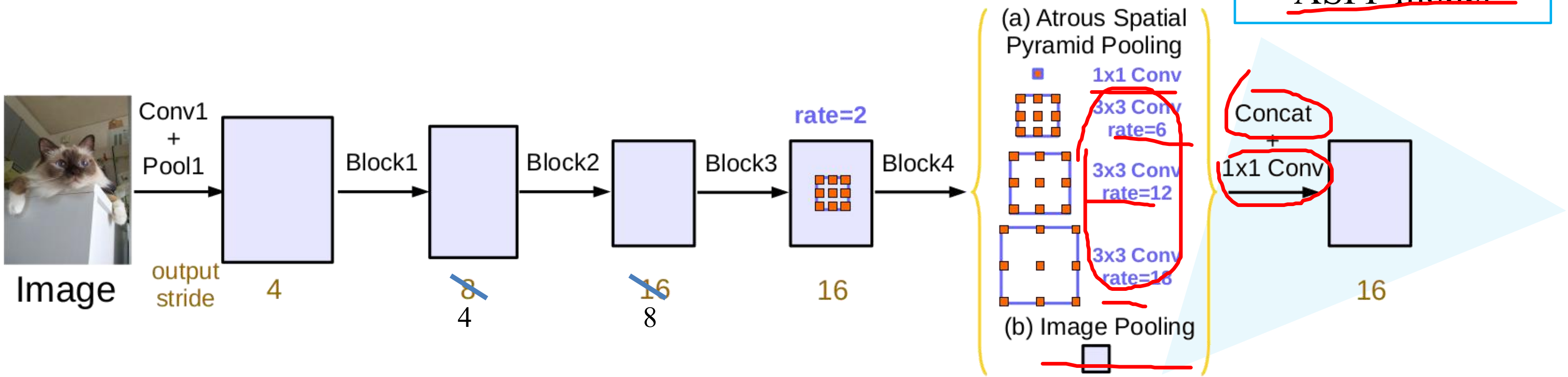
DeepLab V3

DeepLabV3两种模型结构

cascaded model

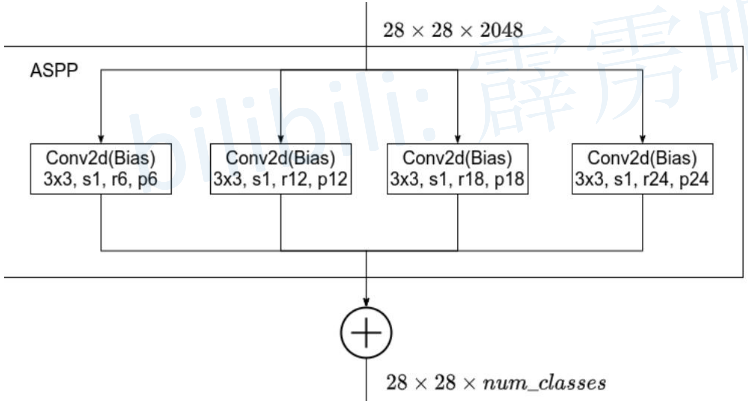


ASPP model

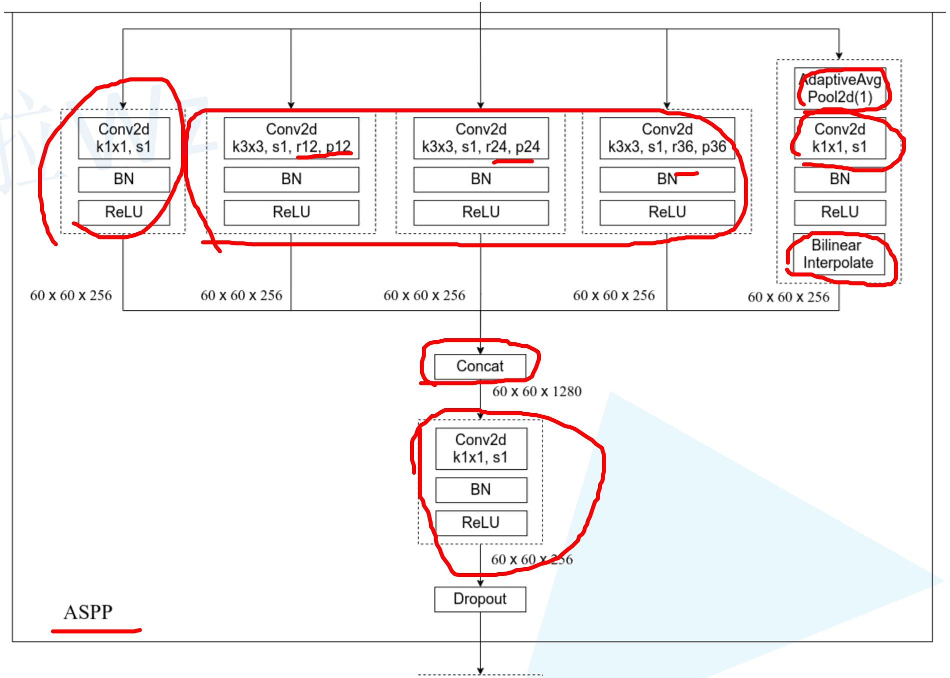


DeepLab V3

V2中的ASPP



V3中的ASPP



https://github.com/WZMIAOMIAO/deep-learning-for-image-processing/tree/master/pytorch_segmentation/deeplab_v3

DeepLab V3

Multi-grid

Multi-Grid	block4	block5	block6	block7
(1, 1, 1)	68.39	73.21	75.34	75.76
(1, 2, 1)	70.23	75.67	76.09	**76.66**
(1, 2, 3)	73.14	75.78	75.96	76.11
(1, 2, 4)	73.45	75.74	75.85	76.02
(2, 2, 2)	71.45	74.30	74.70	74.62

Table 3. Employing multi-grid method for ResNet-101 with different number of cascaded blocks at *output_stride* = 16. The best model performance is shown in bold.

Multi-grid: We apply the multi-grid method to ResNet-101 with several cascadedly added blocks in Tab. 3. The unit rates, $Multi\_Grid = (r_1, r_2, r_3)$, are applied to block4 and all the other added blocks. As shown in the table, we observe that (a) applying multi-grid method is generally better than the vanilla version where $(r_1, r_2, r_3) = (1, 1, 1)$, (b) simply doubling the unit rates (*i.e.*, $(r_1, r_2, r_3) = (2, 2, 2)$) is not effective, and (c) going deeper with multi-grid improves the performance. Our best model is the case where block7 and $(r_1, r_2, r_3) = (1, 2, 1)$ are employed.

The final atrous rate for the convolutional layer is equal to the multiplication of the unit rate and the corresponding rate. *For example, when output stride = 16 and Multi Grid = (1, 2, 4), the three convolutions will have rates = 2 · (1, 2, 4) = (2, 4, 8) in the block4, respectively.*

DeepLab V3

cascaded model消融实验

Method	OS=16	OS=8	MS	Flip	mIOU
block7 +	✓				76.66
MG(1, 2, 1)		✓			78.05
		✓	✓		78.93
		✓	✓	✓	79.35

Table 4. Inference strategy on the *val* set. **MG**: Multi-grid. **OS**: *output_stride*. **MS**: Multi-scale inputs during test. **Flip**: Adding left-right flipped inputs.

scales = {0.5, 0.75, 1.0, 1.25, 1.5, 1.75}

DeepLab V3

ASPP model消融实验

Method	OS=16	OS=8	MS	Flip	COCO	mIOU
MG(1, 2, 4) +	✓					77.21
ASPP(6, 12, 18) +		✓				78.51
Image Pooling		✓	✓			79.45
		✓	✓	✓		79.77
		✓	✓	✓	✓	82.70

Table 6. Inference strategy on the *val* set: **MG**: Multi-grid. **ASPP**: Atrous spatial pyramid pooling. **OS**: *output_stride*. **MS**: Multi-scale inputs during test. **Flip**: Adding left-right flipped inputs. **COCO**: Model pretrained on MS-COCO.

scales = {0.5, 0.75, 1.0, 1.25, 1.5, 1.75}

DeepLab V3

训练细节

A. Effect of hyper-parameters

As mentioned in the main paper, we change the training protocol in [10, 11] with three main differences:
(1) ***larger crop size****,*
(2) ***upsampling logits during training****, and*
(3) ***fine-tuning batch normalization****.*
Here, we quantitatively measure the effect of the changes.

Crop Size	UL	BN	mIOU
513	✓	✓	77.21
513	✓		75.95
513		✓	76.01
321		✓	67.22

Table 8. Effect of hyper-parameters during training on PASCAL VOC 2012 *val* set at *output_stride=16*. **UL**: Upsampling Logits. **BN**: Fine-tuning batch normalization.

Method	**mIOU**
Adelaide_VeryDeep_FCN_VOC [85]	79.1
LRR_4x_ResNet-CRF [25]	79.3
DeepLabv2-CRF [11]	79.7
CentraleSupelec Deep G-CRF [8]	80.2
HikSeg_COCO [80]	81.4
SegModel [75]	81.8
Deep Layer Cascade (LC) [52]	82.7
TuSimple [84]	83.1
Large_Kernel_Matters [68]	83.6
Multipath-RefineNet [54]	84.2
ResNet-38_MS_COCO [86]	84.9
PSPNet [95]	85.4
IDW-CNN [83]	86.3
CASIA_IVA_SDN [23]	86.6
DIS [61]	86.8
DeepLabv3	85.7
DeepLabv3-JFT	86.9

Table 7. Performance on PASCAL VOC 2012 *test* set.

DeepLab V3

Pytorch官方实现的DeepLabV3

- 没有使用Multi-Grid，有兴趣的同学可以自己动手加上试试。
- 多了一个FCNHead辅助训练分支，可以选择不使用。
- 无论是训练还是验证output_stride都使用的8。
- ASPP中三个膨胀卷积分支的膨胀系数是12, 24, 36