浙江大学
ZHEJIANG UNIVERSITY

Deeplabv1论文分享

汇报人：王嘉恒

2022.10.15

问题背景

> There are two technical hurdles in the application of DCNNs to image labeling tasks: signal downsampling, and spatial 'insensitivity' (invariance).

在论文的引言部分抛出了两个问题（针对语义分割任务）： 信号下采样导致分辨率降低和空间“不敏感” 问题。

问题背景

论文贡献:

- 对于第一个问题信号下采样，作者说主要是采用Maxpooling导致的，为了解决这个问题作者引入了‘atrous’(with holes) algorithm（空洞卷积 / 膨胀卷积 / 扩张卷积）。
- 对于第二个问题空间“不敏感”，作者说分类器自身的问题（分类器本来就具备一定空间不变性），为了解决这个问题作者采用了fully-connected CRF(Conditional Random Field)方法

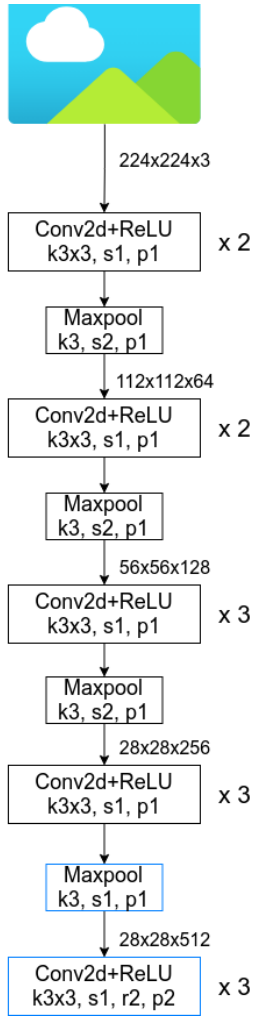
Deeplabv1的成果

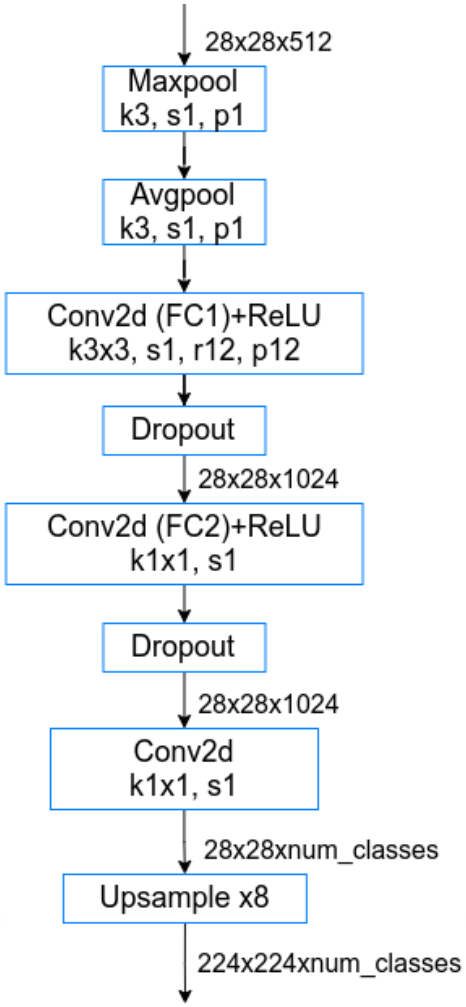
- 速度更快，论文中说是使用了膨胀卷积的原因（减小参数量）保持8FPS的速度
- 准确率更高，相比之前的最好模型提升了7.2个点，达到71.2%的Iou
- 模型很简单，主要由DCNN和CRF级联构成

Deeplabv1

网络细节-DCNN

- 使用VGG16的backbone，但是后两个池化层不进行下采样
- 后面3层改成卷积层
- 池化层中k=3，s=2，p=1
- 最后三个3×3的卷积层就使用了膨胀卷积 r=2
- 将全连接层通道数从4096下降到1024，减少参数并且没有损失性能
- 最后通过上采样8倍还原成原图大小



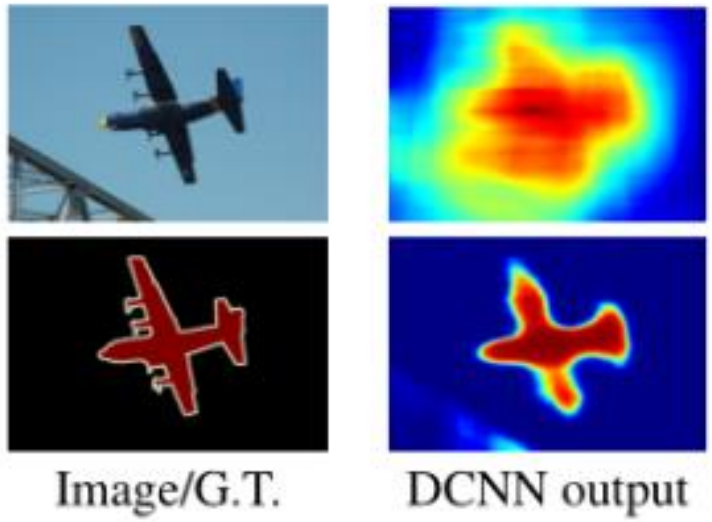


网络细节-DCNN

DCNN预测物体的轮廓是粗糙的，其分类精度和定位精度存在一个平衡，太大的感受野和网络的不变性导致无法得到精确信息。

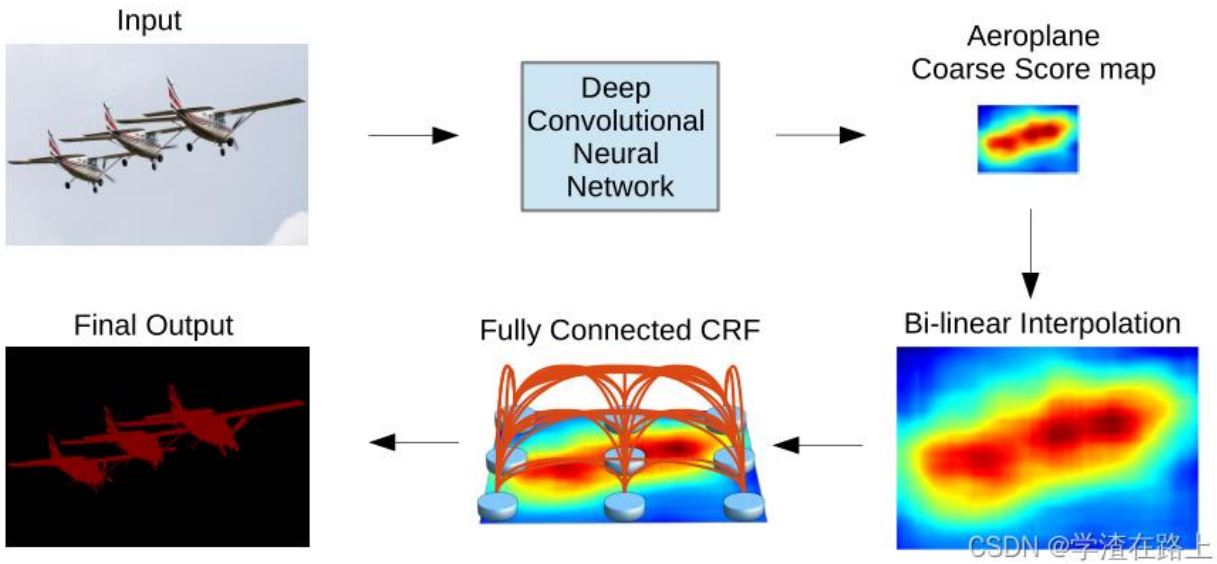
这时，作者讨论了三个方法来进行优化：

- 将CRF和DCNN结合起来解决
- 使用Multi-Scale的方法，和FCN中的思想类似，融合前四个maxpool层的输出
- 讨论控制感受野，使用空洞卷积带来的好处



网络细节-CRF

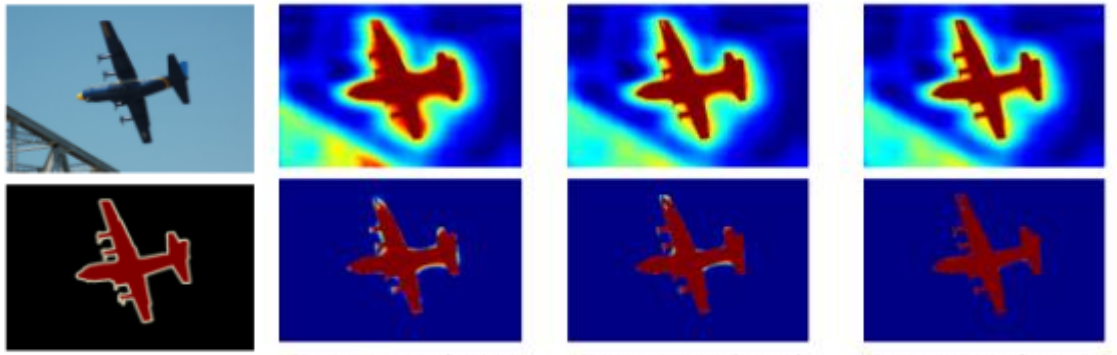
在deeplabv1中 作者使用了fully connected CRF（全连接条件随机场）
是将DCNN和CRF级联的一个结构



CSDN @学渣在路上

网络细节-CRF

相较于直接使用DCNN，后面加一个CRF的效果是会更好。
这里在训练CRF的过程中是固定了DCNN的网络

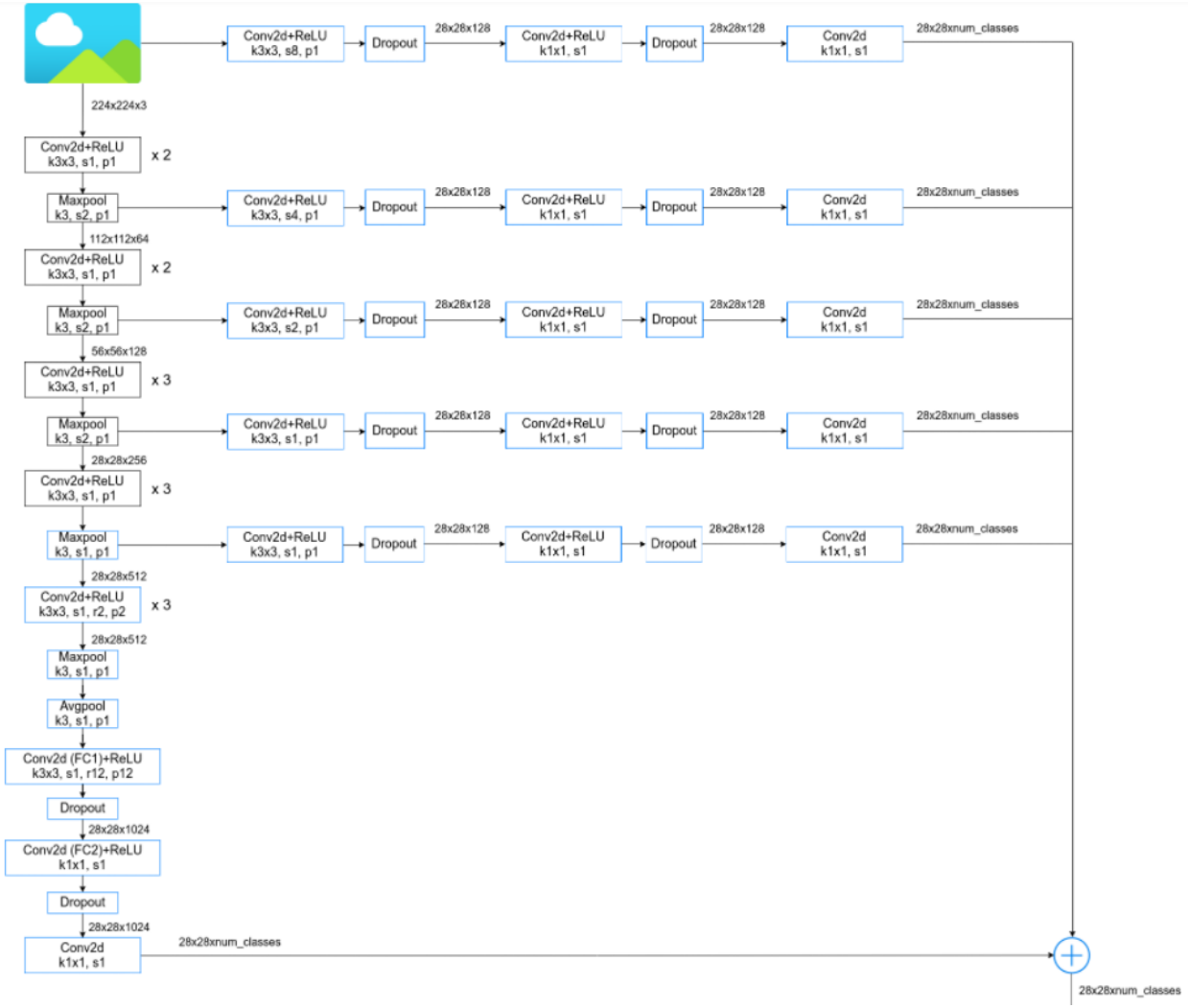


浙江大学 ZHEJIANG UNIVERSITY

Deeplabv1

网络细节-Msc

Multi-Scale Prediction，融合了多个特征层的输出。
融合了来自原图尺度以及前四个maxpool层的输出，将其通过一个两层的多层感知机。



网络细节-LargeFov

Method	kernel size	input stride	receptive field	# parameters	mean IOU (%)	Training speed (img/sec)
DeepLab-CRF-7x7	7×7	4	224	134.3M	67.64	1.44
DeepLab-CRF	4×4	4	128	65.1M	63.74	2.90
DeepLab-CRF-4x4	4×4	8	224	65.1M	67.14	2.90
DeepLab-CRF-LargeFOV	3×3	12	224	20.5M	67.64	4.84

Table 2: Effect of Field-Of-View. We show the performance (after CRF) and training speed on the PASCAL VOC 2012 'val' set as the function of (1) the kernel size of first fully connected layer, (2) the input stride value employed in the atrous algorithm.

DeepLab-CRF-7×7：效果不错但参数量大 且缓慢
DeepLab-CRF ：训练结果下降
DeepLab-CRF-4×4：效果上升，直观体现空洞卷积的作用
DeepLab-CRF-LargeFOV：效果最好，且训练速度最快

实验结果

在PASCAL VOC 2012上跑出的结果如右图：

- 添加CRF 可以提高3个百分点
- 用MSc方法 提高1.5个百分点
- 用LargeFOV方法，提高2个百分点

Method	mean IOU (%)
DeepLab	59.80
DeepLab-CRF	63.74
DeepLab-MSc	61.30
DeepLab-MSc-CRF	65.21
DeepLab-7x7	64.38
DeepLab-CRF-7x7	67.64
DeepLab-LargeFOV	62.25
DeepLab-CRF-LargeFOV	67.64
DeepLab-MSc-LargeFOV	64.21
DeepLab-MSc-CRF-LargeFOV	68.70

实验结果

在PASCAL VOC 2012与其他方法进行比较：

可以看到，DeepLabv1跑出最好的结果比当时最好的方法高出7个百分点。

Method	mean IOU (%)
MSRA-CFM	61.8
FCN-8s	62.2
TTI-Zoomout-16	64.4
DeepLab-CRF	66.4
DeepLab-MSc-CRF	67.1
DeepLab-CRF-7x7	70.3
DeepLab-CRF-LargeFOV	70.3
DeepLab-MSc-CRF-LargeFOV	71.6

总结

优点：使用空洞卷积减少了参数量，更快、使用CRF得到了更准的结果，比之前最好的结果高了7个百分点。

不足：

DeepLabv1中选择的backboneVGG16性能一般，还有待改进的空间使用

CRF占用了很多的时间，所以到v3中都不再使用了

Deeplabv2

问题背景

- 连续下采样导致空间分辨率降低
- 目标的多尺度问题
- DCNNs的不变性会降低定位精度

相较于DeepLabv1的改进：

- 把backbone换成了Resnet
- 对于多尺度问题，引入了ASPP模块
- 把CRF做了一个优化，换成了fully connected pairwise CRF

ASPP:空洞卷积池化金字塔

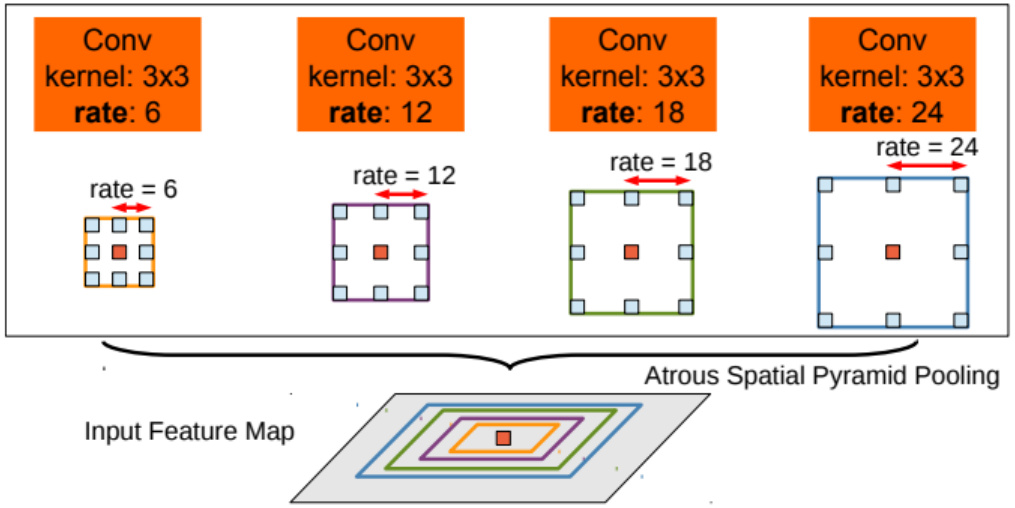


Fig. 4: Atrous Spatial Pyramid Pooling (ASPP). To classify the center pixel (orange), ASPP exploits multi-scale features by employing multiple parallel filters with different rates. The effective Field-Of-Views are shown in different colors.

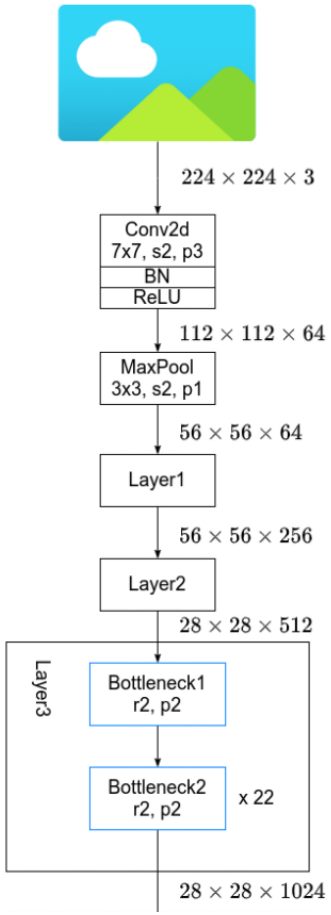
Method	before CRF	after CRF
LargeFOV	65.76	69.84
ASPP-S	66.98	69.73
ASPP-L	68.96	71.57

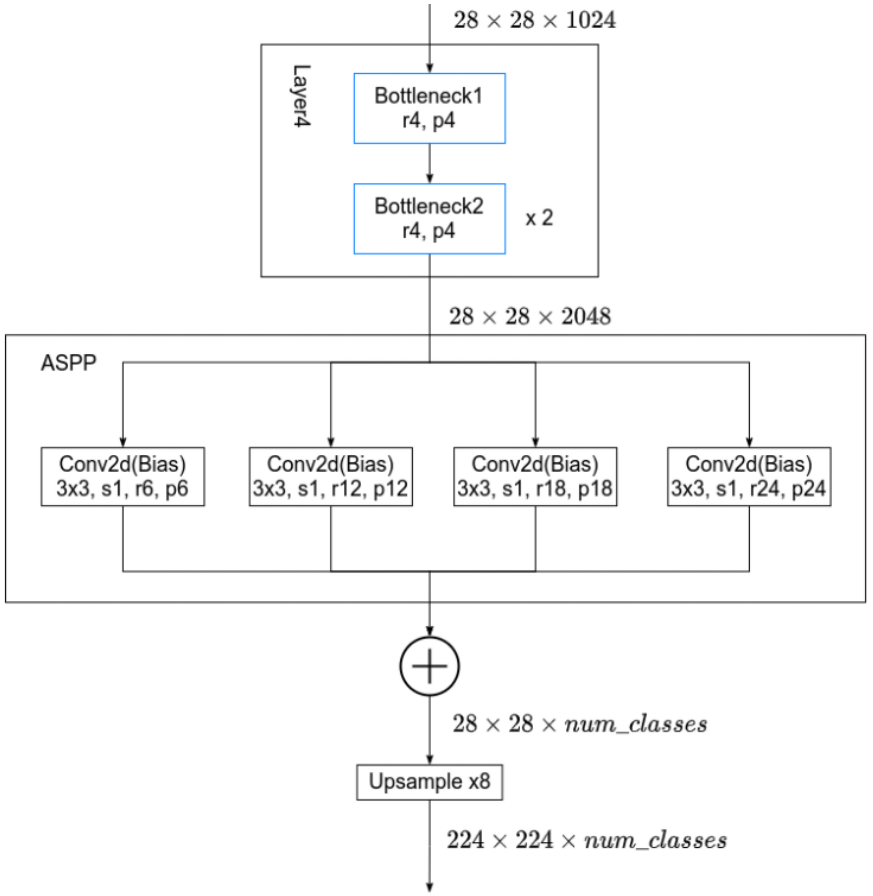
TABLE 3: Effect of ASPP on PASCAL VOC 2012 *val* set performance (mean IOU) for VGG-16 based DeepLab model. **LargeFOV**: single branch, $r = 12$. **ASPP-S**: four branches, $r = \{2, 4, 8, 12\}$. **ASPP-L**: four branches, $r = \{6, 12, 18, 24\}$.

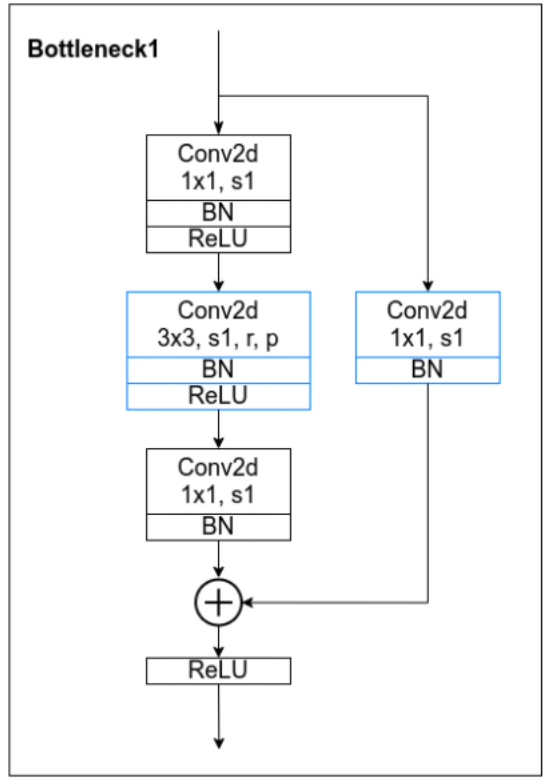
浙江大学 ZHEJIANG UNIVERSITY

Deeplabv2

DeepLabv2结构







Learning Rate policy

Learning policy	Batch size	Iteration	mean IOU
step	30	6K	62.25
poly	30	6K	63.42
poly	30	10K	64.90
poly	10	10K	64.71
poly	10	20K	65.88

TABLE 2: PASCAL VOC 2012 *val* set results (%) (before CRF) as different learning hyper parameters vary. Employing "poly" learning policy is more effective than "step" when training DeepLab-LargeFOV.

$$lr \times (1 - \frac{iter}{max\_iter})^{power}$$

Power=0.9

消融实验

MSC	COCO	Aug	LargeFOV	ASPP	CRF	mIOU
						68.72
✓						71.27
✓	✓					73.28
✓	✓	✓				74.87
✓	✓	✓	✓			75.54
✓	✓	✓		✓		76.35
✓	✓	✓		✓	✓	77.69

TABLE 4: Employing ResNet-101 for DeepLab on PASCAL VOC 2012 *val* set. **MSC**: Employing mutli-scale inputs with max fusion. **COCO**: Models pretrained on MS-COCO. **Aug**: Data augmentation by randomly rescaling inputs.

of [17], [18], [39], [40], [58], [59], [62]: (1) Multi-scale inputs: We separately feed to the DCNN images at scale = {0.5, 0.75, 1}, fusing their score maps by taking the maximum response across scales for each position separately [17]. (2) Models pretrained on MS-COCO [87]. (3) Data augmentation by randomly scaling the input images (from 0.5 to 1.5) during training. In Tab. 4, we evaluate how each of these factors, along with LargeFOV and atrous spatial pyramid pooling (ASPP), affects *val* set performance. Adopting ResNet-101 instead of VGG-16 significantly improves DeepLab performance (*e.g.*, our simplest ResNet-101 based model attains 68.72%, compared to 65.76% of our DeepLab-LargeFOV VGG-16 based variant, both before CRF). Multiscale fusion [17] brings extra 2.55% improvement, while pretraining the model on MS-COCO gives another 2.01% gain. Data augmentation during training is effective (about 1.6% improvement). Employing LargeFOV (adding an atrous convolutional layer on top of ResNet, with 3×3 kernel and rate = 12) is beneficial (about 0.6% improvement). Further 0.8% improvement is achieved by atrous spatial pyramid pooling (ASPP). Post-processing our best model by dense CRF yields

Deeplabv2

与其他算法对比：

Method	mIOU
DeepLab-CRF-LargeFOV-COCO [58]	72.7
MERL_DEEP_GCRF [88]	73.2
CRF-RNN [59]	74.7
POSTECH_DeconvNet_CRF_VOC [61]	74.8
BoxSup [60]	75.2
Context + CRF-RNN [76]	75.3
QO_4^{mres} [66]	75.5
DeepLab-CRF-Attention [17]	75.7
CentraleSuperBoundaries++ [18]	76.0
DeepLab-CRF-Attention-DT [63]	76.3
H-ReNet + DenseCRF [89]	76.8
LRR_4x_COCO [90]	76.8
DPN [62]	77.5
Adelaide_Context [40]	77.8
Oxford_TVG_HO_CRF [91]	77.9
Context CRF + Guidance CRF [92]	78.1
Adelaide_VeryDeep_FCN_VOC [93]	79.1
DeepLab-CRF (ResNet-101)	79.7

基于vgg

TABLE 5: Performance on PASCAL VOC 2012 *test* set. We have added some results from recent arXiv papers on top of the official leadearboard results.

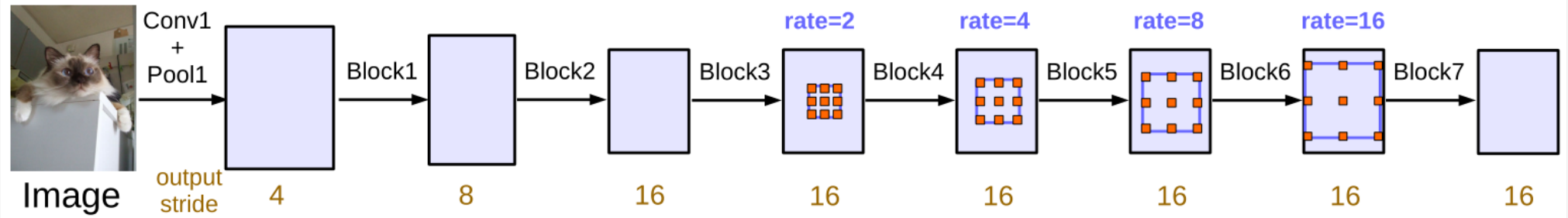
https://blog.csdn.net/magic_ll

问题背景

Deeplabv3是在Deeplabv1、2上的再扩展

- Deeplabv1是将VGG中的后两个池化层stride=1，然后采用了空洞卷积来扩大感受野，上采样使用双线性插值
- Deeplabv2是在将backbone的输出后增加一个ASPP结构，通过不同的空洞卷积率得到不同尺度的特征图再将其进行融合
- Deeplabv3则是对ASPP的结构再进行一个升级，获得更好的性能

网络结构-cascaded model



(b) Going deeper with atrous convolution. Atrous convolution with $rate > 1$ is applied after block3 when $output\_stride = 16$.

Figure 3. Cascaded modules without and with atrous convolution.

网络结构-cascaded model

级联结构下，使用空洞卷积的块级联，块越多效果越好，但是提升的幅度越来越小。

Network	block4	block5	block6	block7
ResNet-50	64.81	72.14	74.29	73.88
ResNet-101	68.39	73.21	75.34	75.76

Table 2. Going deeper with atrous convolution when employing ResNet-50 and ResNet-101 with different number of cascaded blocks at *output_stride* = 16. Network structures 'block4', 'block5', 'block6', and 'block7' add extra 0, 1, 2, 3 cascaded modules respectively. The performance is generally improved by adopting more cascaded blocks.

网络结构-cascaded model

将Multi-grid的策略运用到Resnet101的级联结构上：

使用(1,2,1)的策略最好

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.

网络结构-cascaded model

消融实验：

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

网络结构-ASPP model

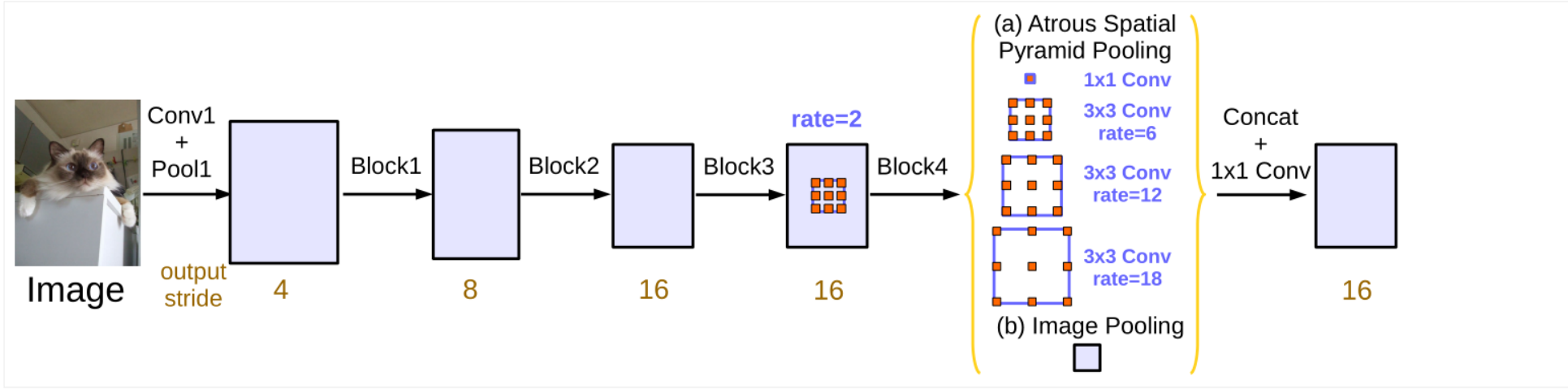
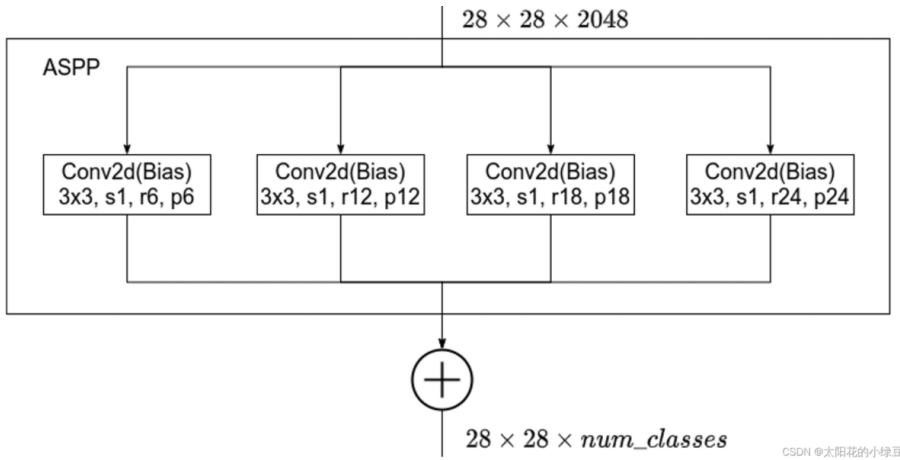
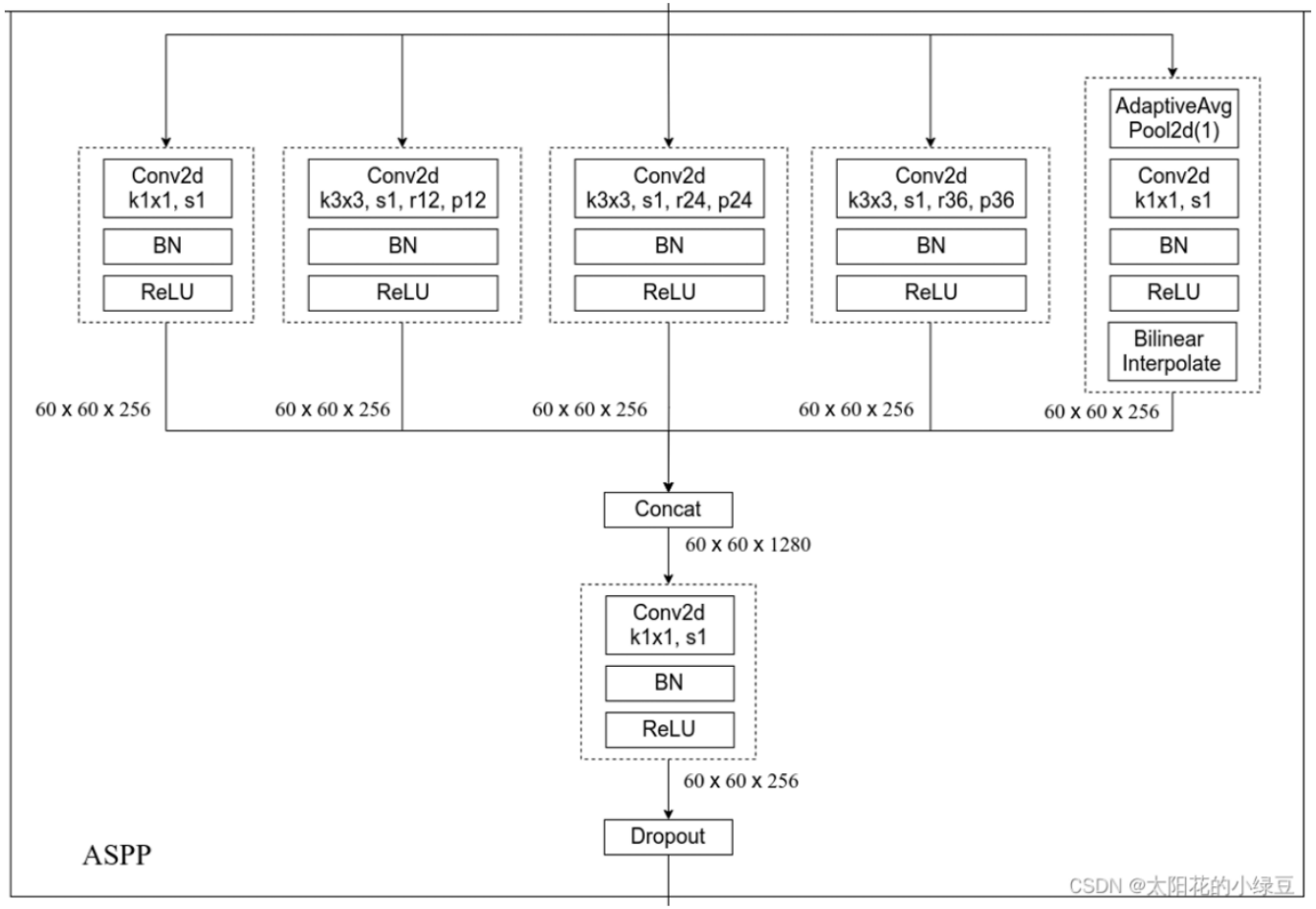


Figure 5. Parallel modules with atrous convolution (ASPP), augmented with image-level features.

Deeplabv3

网络结构-ASPP model





CSDN @太阳花的小绿豆

网络结构-ASPP model

Multi-grid对ASPP model的影响：

对于6，12，18采用(1，2，4)效果是最好的

再加一个r=24的分支，性能反而会下降

对ASPP进行Image Pooling是有效的

Multi-Grid			ASPP		Image	
(1, 1, 1)	(1, 2, 1)	(1, 2, 4)	(6, 12, 18)	(6, 12, 18, 24)	Pooling	mIOU
✓			✓			75.36
	✓		✓			75.93
		✓	✓			76.58
		✓		✓		76.46
		✓	✓		✓	77.21

Table 5. Atrous Spatial Pyramid Pooling with multi-grid method and image-level features at *output_stride* = 16.

网络结构-ASPP model

消融实验：

使用OS=16来进行训练，OS=8来进行验证会比较好。最好的性能达到79.77%，优于级联模型的79.35%。

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.

DeepLabv3

与其他算法对比：

DeepLabv3相较于v2提升了6个点。

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.