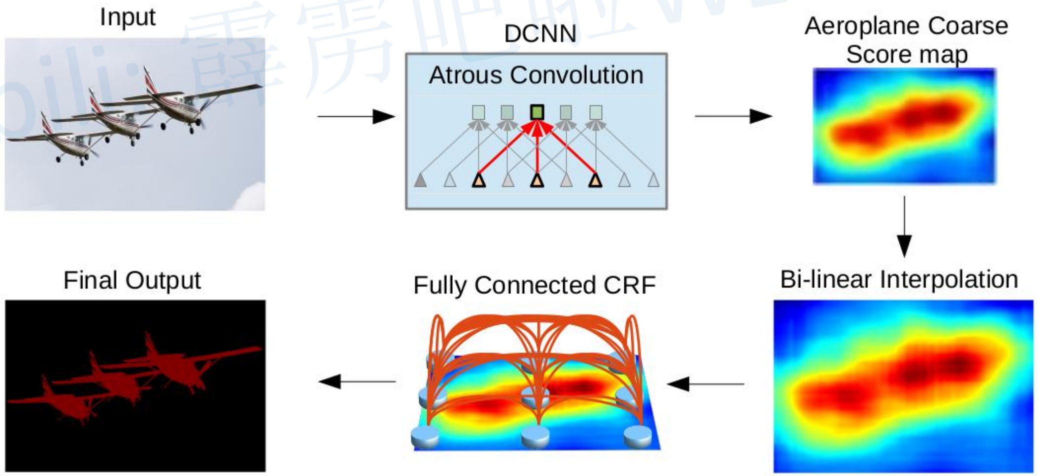
DeepLab V2

DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs

Liang-Chieh Chen, George Papandreou, *Senior Member, IEEE,* Iasonas Kokkinos, *Member, IEEE,*
Kevin Murphy, and Alan L. Yuille, *Fellow, IEEE*

2016 CVPR



论文下载地址：https://arxiv.org/abs/1606.00915

博文推荐：https://blog.csdn.net/qq_37541097/article/details/121752679

DeepLab V2

DCNNs应用在语义分割任务中问题

In particular we consider three challenges in the application of DCNNs to semantic image segmentation: (1) ***reduced feature resolution****, (2)* ***existence of objects at multiple scales****, and (3)* ***reduced localization accuracy due to DCNN invariance****.*

- 分辨率被降低（主要由于下采样stride>1的层导致）
- 目标的多尺度问题
- DCNNs的不变性(invariance)会降低定位精度

DeepLab V2

对应的解决方法

- **针对分辨率被降低的问题**，一般就是将最后的几个Maxpooling层的**stride设置成1**(如果是通过卷积下采样的，比如resnet，同样将stride设置成1即可)，**配合使用膨胀卷积。**
- **针对目标多尺度的问题**，最容易想到的就是将图像缩放到多个尺度分别通过网络进行推理，最后将多个结果进行融合即可。这样做虽然有用但是计算量太大了。为了解决这个问题，DeepLab V2 中提出了ASPP模块（atrous spatial pyramid pooling）。
- **针对DCNNs不变性导致定位精度降低的问题**，和DeepLab V1差不多还是**通过CRFs解决**，不过这里用的是fully connected pairwise CRF，相比V1里的fully connected CRF要更高效点。

DeepLab V2

网络优势

- 速度更快
- 准确率更高（当时的state-of-art）
- 模型结构简单，还是DCNNs和CRFs联级

DeepLab V2

ASPP(atrous spatial pyramid pooling)

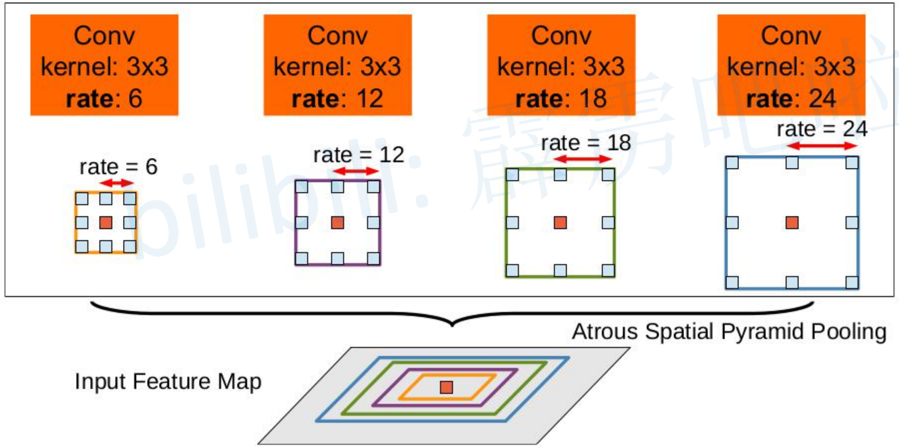


Fig. 4: Atrous Spatial Pyramid Pooling (ASPP). To classify the center pixel (orange), ASPP exploits multi-scale features by employing multiple parallel filters with different rates. The effective Field-Of-Views are shown in different colors.

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, $r = \{6, 12, 18, 24\}$.

DeepLab V2

ASPP(atrous spatial pyramid pooling)

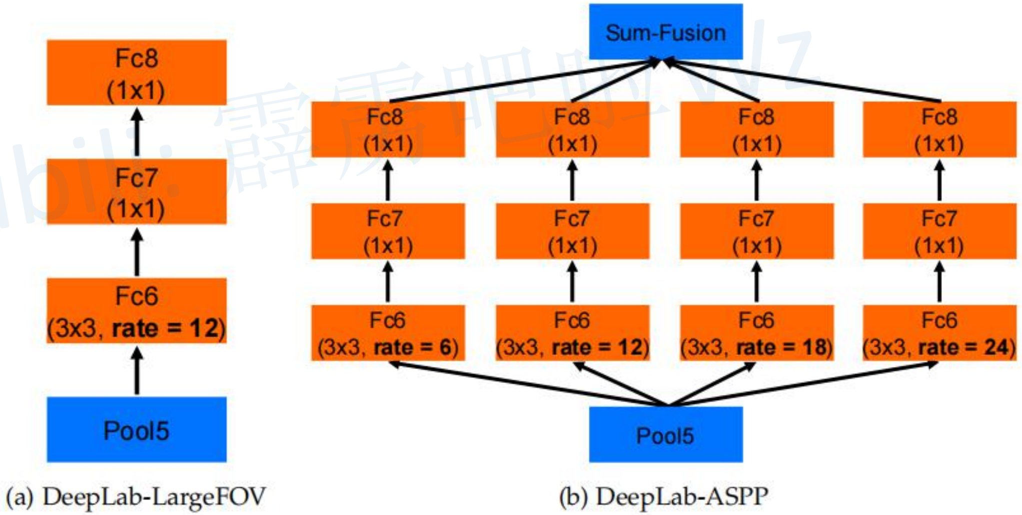


Fig. 7: DeepLab-ASPP employs multiple filters with different rates to capture objects and context at multiple scales.

DeepLab V2

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

DeepLab V2

Learning rate policy

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

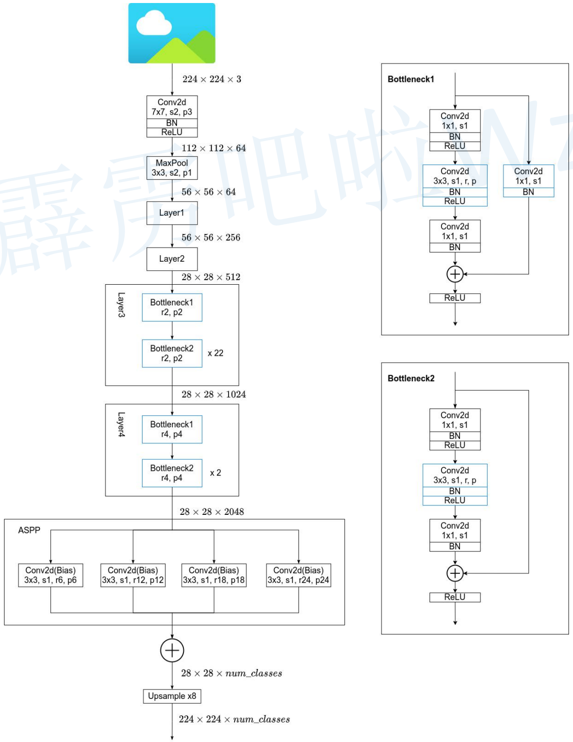
power = 0.9

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.

DeepLab V2

DeepLab V2网络结构



backbone：ResNet101

不考虑MSC的情况