DeepLab v3 paper

INTroduction

语义分割两大挑战（DCNN深度卷积网络）

① 因为连续池化和卷积，导致降低的特征分辨率

让 DCNN 只能学习到越来越抽象的特征

↓这种图像局部变换不变性影响密集检测任务

需要详细的空间信息

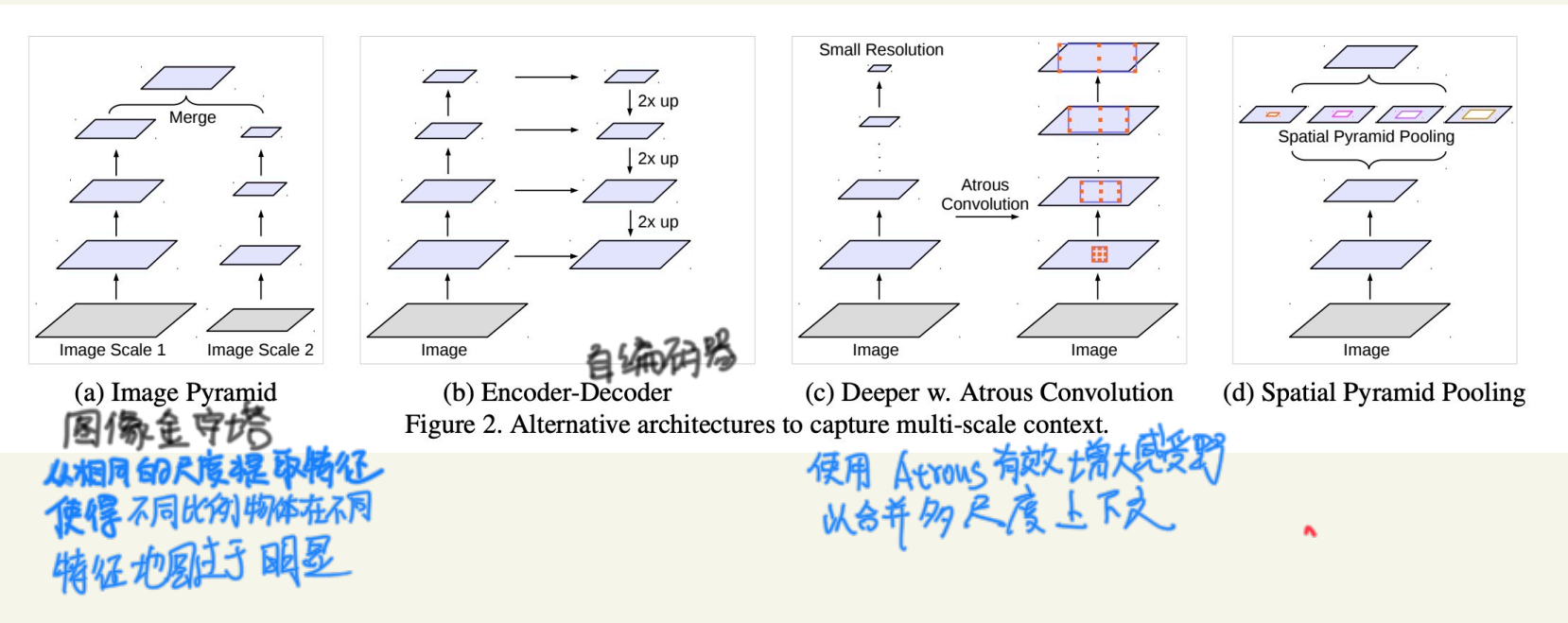
解决方法：基于Atrous卷积（萎缩的卷积）

Atrous conv 又是扩张性卷积，1.通过删除ImageNet中最后几层的下采样，提取更密集的特征图 2.在相应的滤波器上进行上采样，即为在滤波器权重之间插入孔

效果：①控制某个特征的分辨率，在DCNN模型中，无需额外学习参数

② Another difficulty come from 多尺度物体

有一些解决方法，比如



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

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

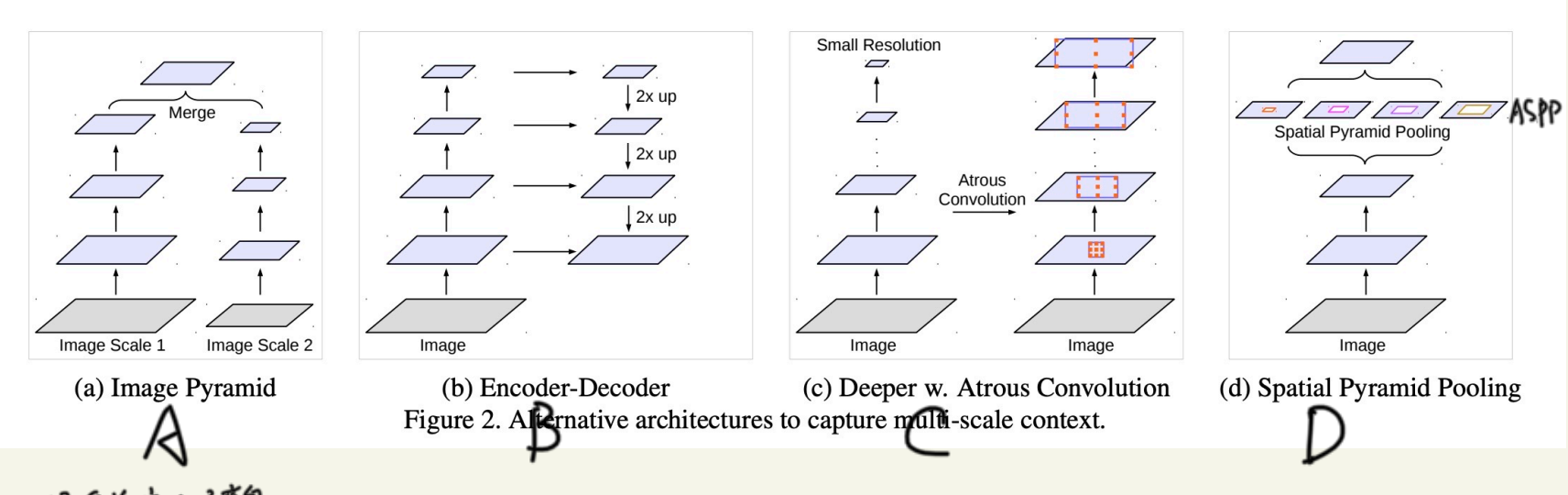
自编码器

图像金字塔

从相同的尺度提取特征，使得不同比例物体在不同特征地图中明显

使用 Atrous 有效增大感受野以合并多尺度上下文

Related Work



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

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

A B C D

无法很好适应大尺度的DCNN

Method

① Atrous Convolution for Dense feature Extraction

考虑二维信号，对于输出y和滤波器w上的每个位置i，在输入特征映射x上应用卷积

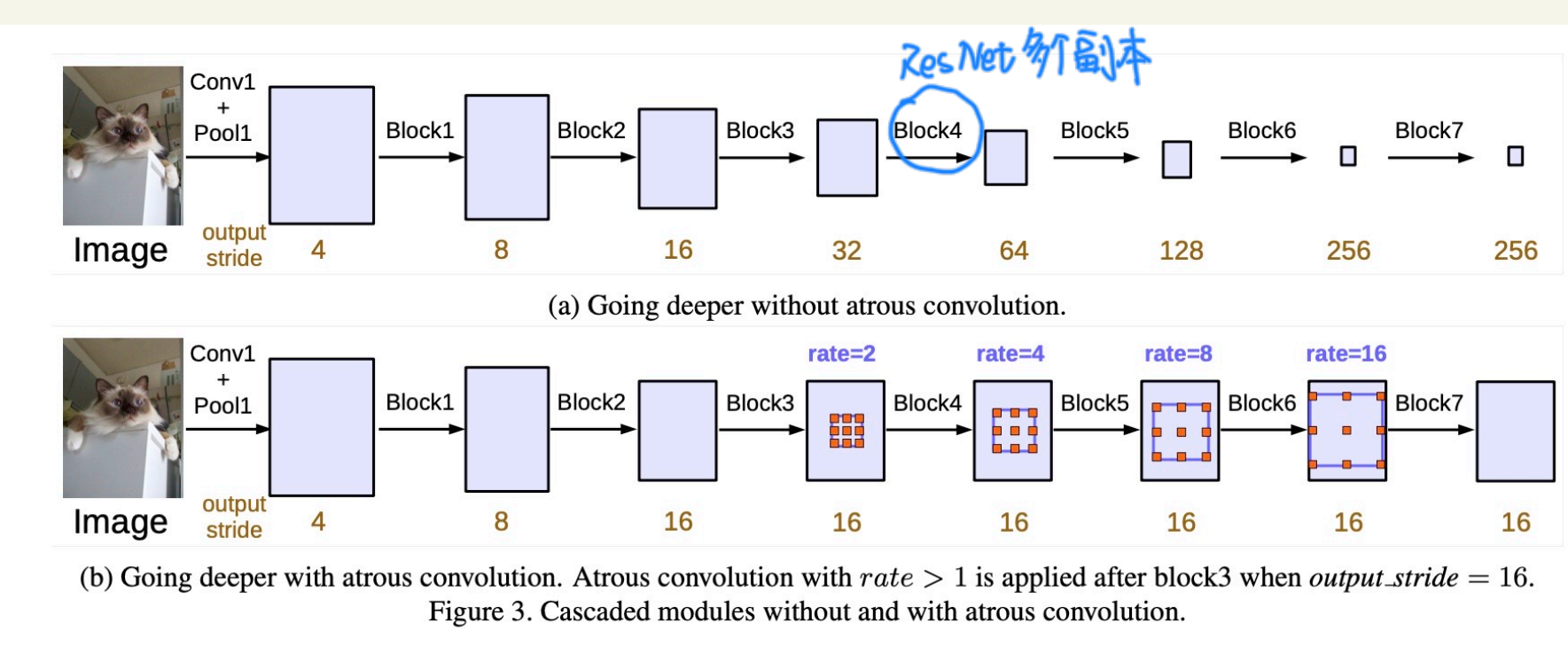
$y[i] = \sum_k x[i + r \cdot k] w[k]$

↑萎缩率（相当于采样输入信号的步长，也等于将输入x与通过插入r产生的上采样滤波器进行卷积

标准为r=1，atrous 允许改变速率自适应修改滤波器视野值

还允许我们显式控制在 CNN 中计算特征响应的密度

② Going Deeper with Atrous Convolution



(a) Going deeper without atrous convolution.

(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.

3.2.1 Mutli-grid Method 多网格方法，采用不同的网格分层

3.3 Atrous Spatial Pyramid Pooling

4 Experimental Evaluation

ImageNet 预训练 RestNet 语义分割，通过 Atrous 提取密集特征

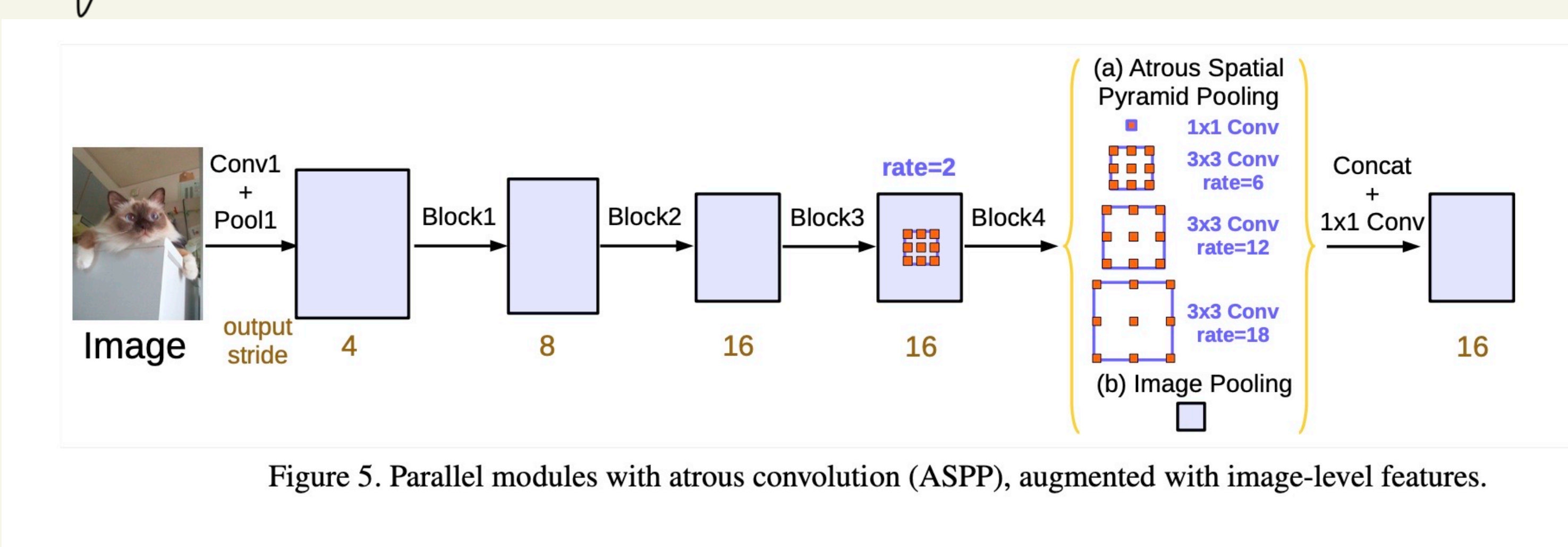


Figure 5. Parallel modules with atrous convolution (ASPP), augmented with image-level features.

4.1 Training Protocol

Learning rate policy: initial lr is $(1 - \frac{iter}{max\_iter})$, power=0.9

Copy size: For astrous conv with large rates to be effective large crop size is required

set crop size to be 513

Batch normalization: add modules on top of ResNet, employ output_stride=16 and batch_size=16

the batch normalization param decay = 0.9997

initial_lr = 0.007, 30k iterations

↓

freeze params, out_stride = 8

↓

train set for 30k small base lr = 0.001

tips:

speed(out_stries=16) > speed(8)

aber lose acc

Upsamling logits: upsamle the final logits

remove the fine annotations

4.2 Going Deeper with Atrous Convolution (block5, 6, 7)

ResNet-50: atrous conv is essential when building more blocks cascadelly for semantic segmentation

ResNet-50 vs ResNet-101: Notely in ResNet50 use block7 decrease the performance, but not ResNet-101

Multi-grid: best model block7 $(r_1, r_2, r_3) = (1, 2, 1)$

Inference strategy on val set, { Multi-scale input, left-right flipped } performance better

4.3 Astrous Spatial Pyramid Pooling

ASPP: final 77%

Inference stategy on val set: ASPP (79.77%)

Comparison with DeepLab v2: all improve but for the bact normalization parameters