

通用视觉框架OpenMMLab
第5讲 语义分割与MMSegmentation

刘子纬 教授
2021年5月



任务： 将图像按照物体的类别分割成不同的区域

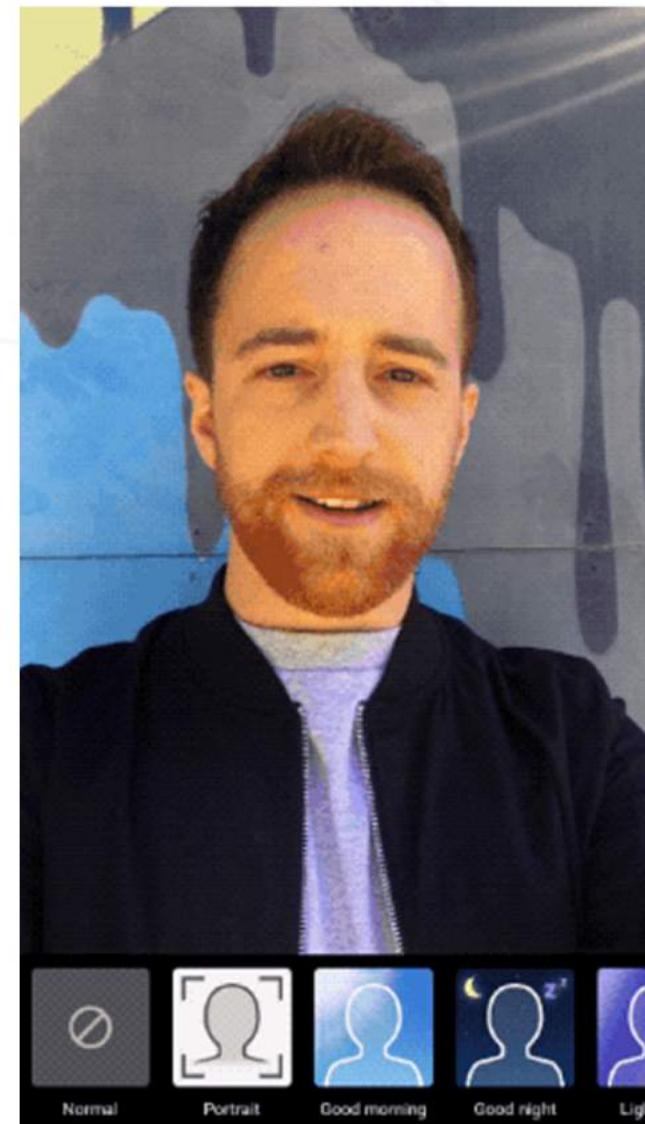
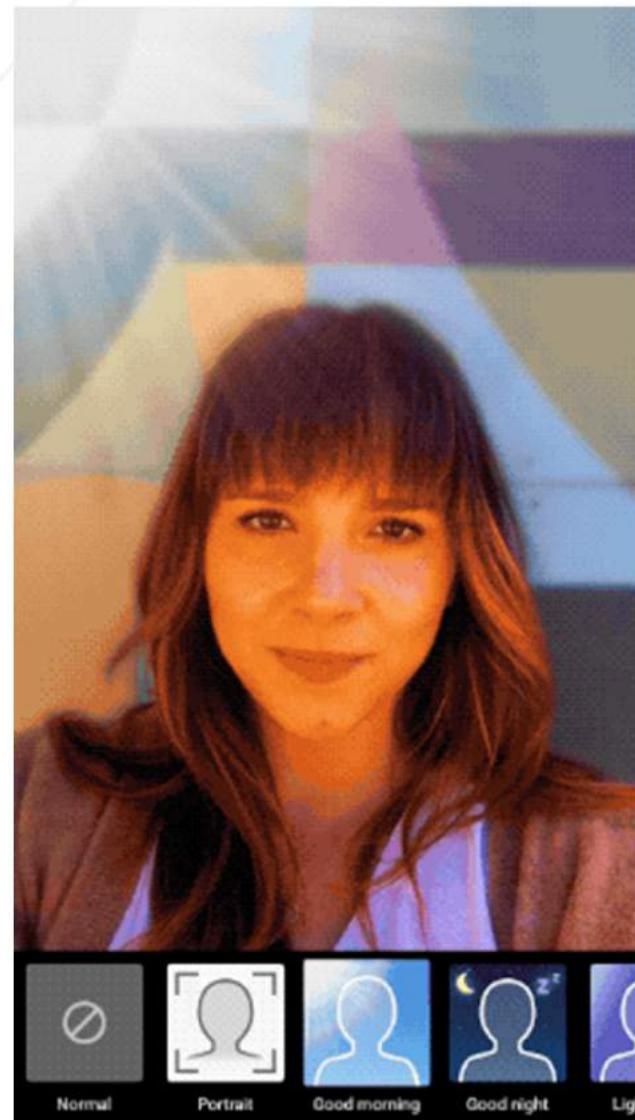
等价于： 对每个像素进行分类

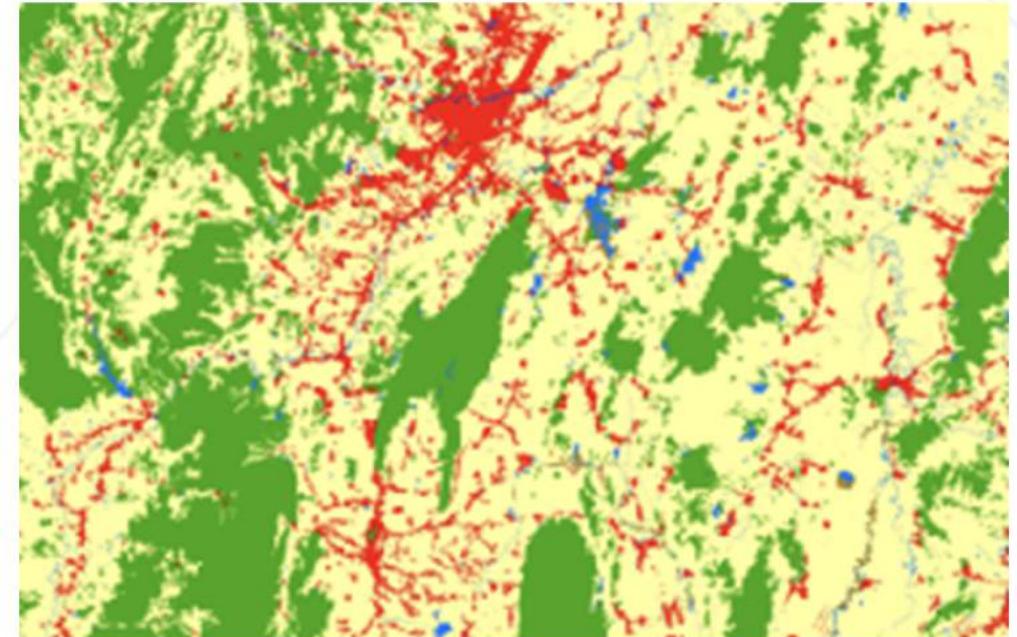
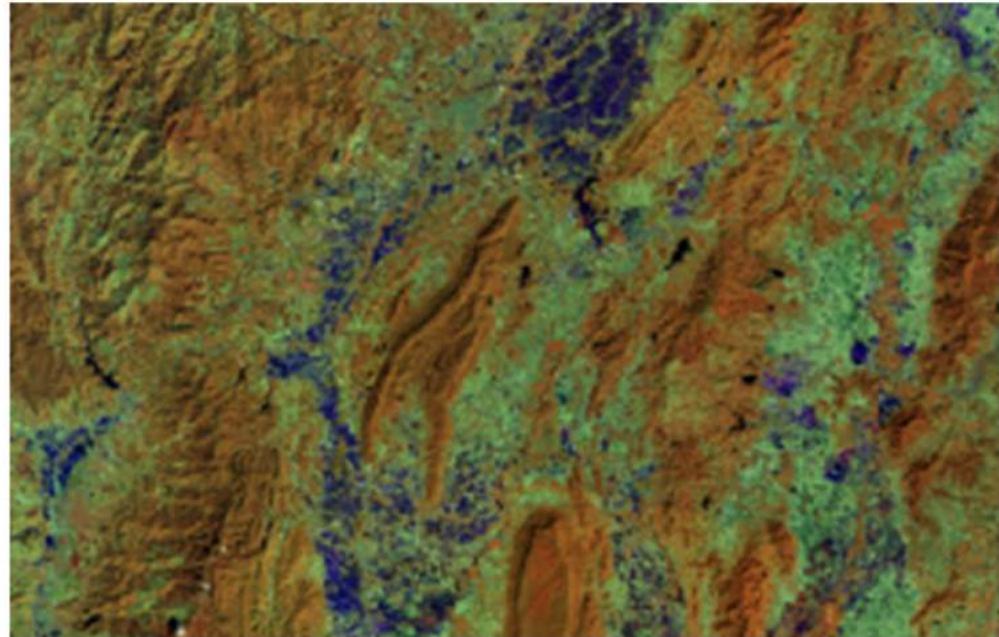


自动驾驶车辆，会将行人，其他车辆，行车道，人行道、交通标志、房屋、草地与树木等按照类别在图像中分割出来，从而辅助车辆对道路的情况进行识别与认知。

实时替换视频的背景

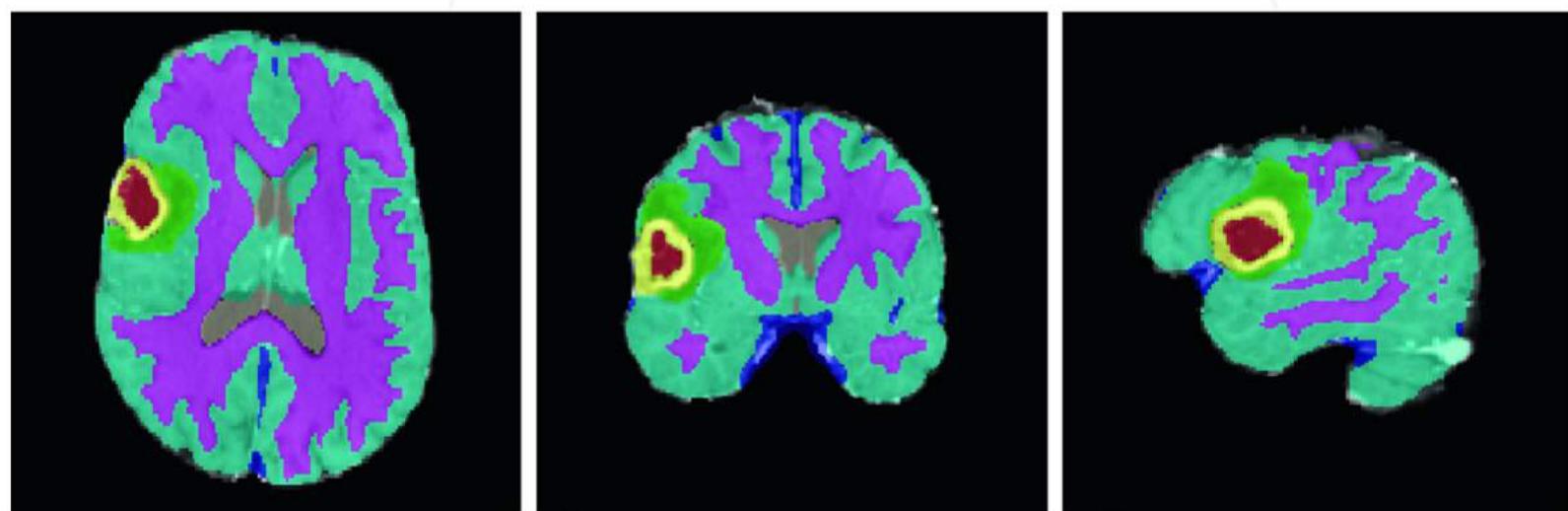
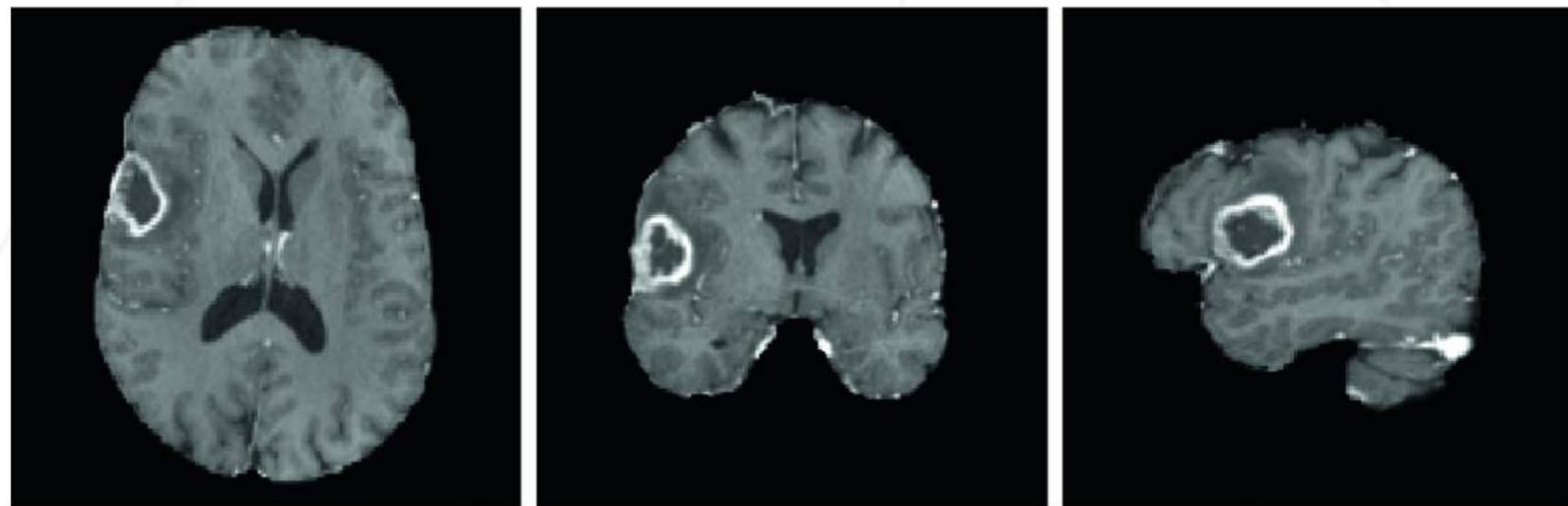
在智慧互娱和智能会议场景中，可以通过这种方法增加交互的多样性





分辨地表物体的类别，通过右侧分割之后的图像可以看到，红色的部分属于湖泊水流。通过智能遥感能够监测不同季节地表水域的变化，从而辅助农业生产，以及旱灾洪灾的预测等等。

通过图像分割技术，辅助进行医疗诊断。如右图，识别脑部肿瘤异物的位置。





语义分割

仅考虑像素的类别

不分割同一类的不同实体



实例分割

分割不同的实体

仅考虑前景物体



全景分割

背景仅考虑类别

前景需要区分实体



本节内容：

- 语义分割的基本思路
- 深度学习下的语义分割模型
 - 全卷积网络
 - 空洞卷积与 DeepLab 模型
 - 上下文信息与 PSPNet 模型
- 分割模型的评估方法
- 实践 MMSegmentation

先验知识

物体内部颜色相近，物体交界颜色变化

基于图像处理方法 按照颜色分割



需要额外手段确定物体类别

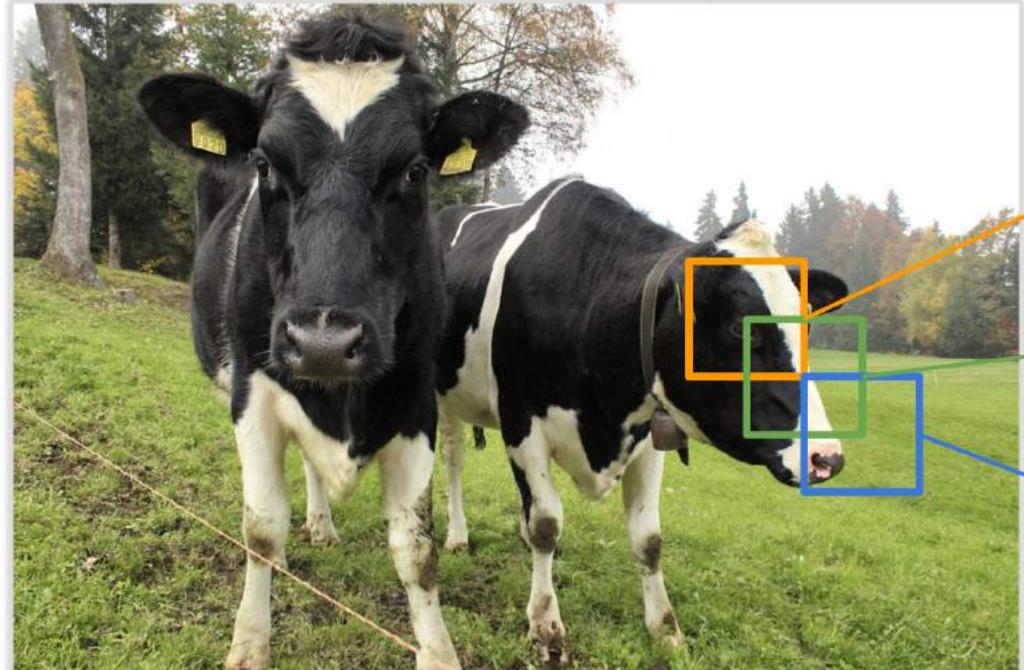


问题

先验知识不完全准确：

不同物体颜色可能相近，物体内也会包含多种颜色

最终性能依赖初步分割结果



滑窗



分类模型

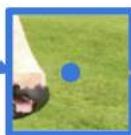


中心像素
的类别

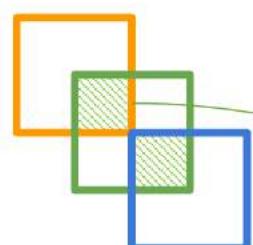
奶牛



奶牛

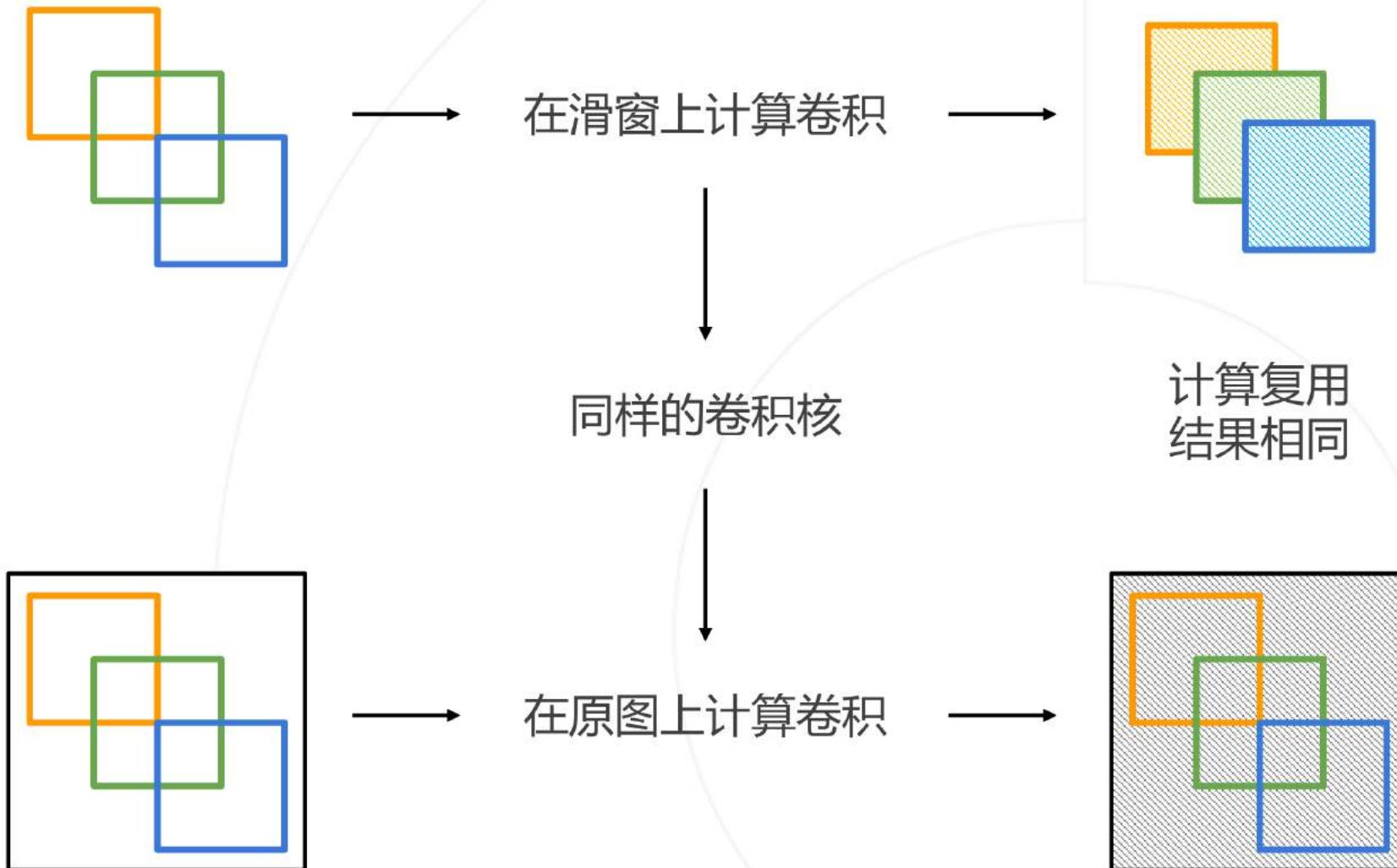


草地



优势：可以充分利用已有的图像分类模型

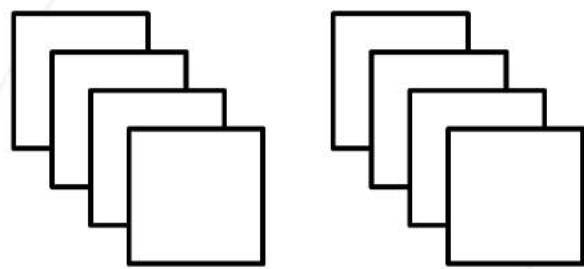
问题：效率低下，重叠区域重复计算卷积



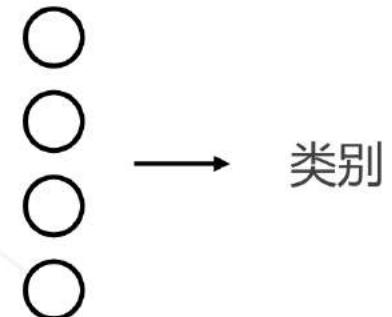
固定大小的区块



卷积层



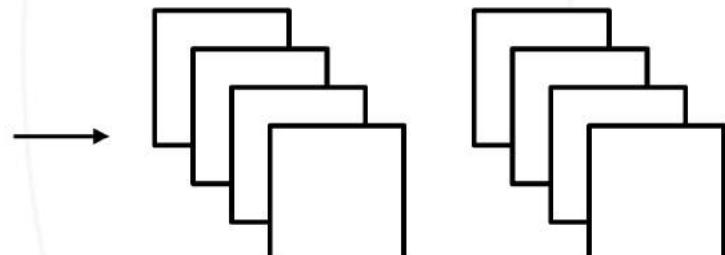
全连接层



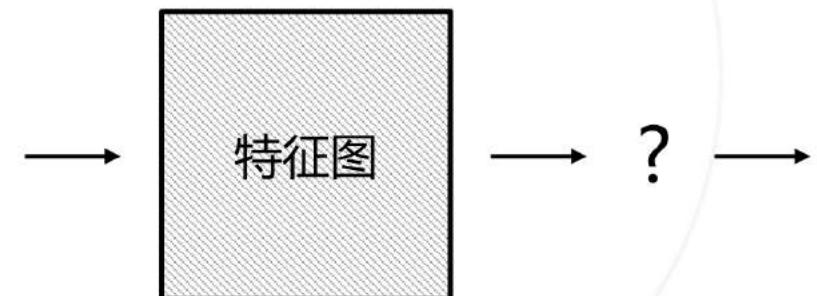
任意大小的原图



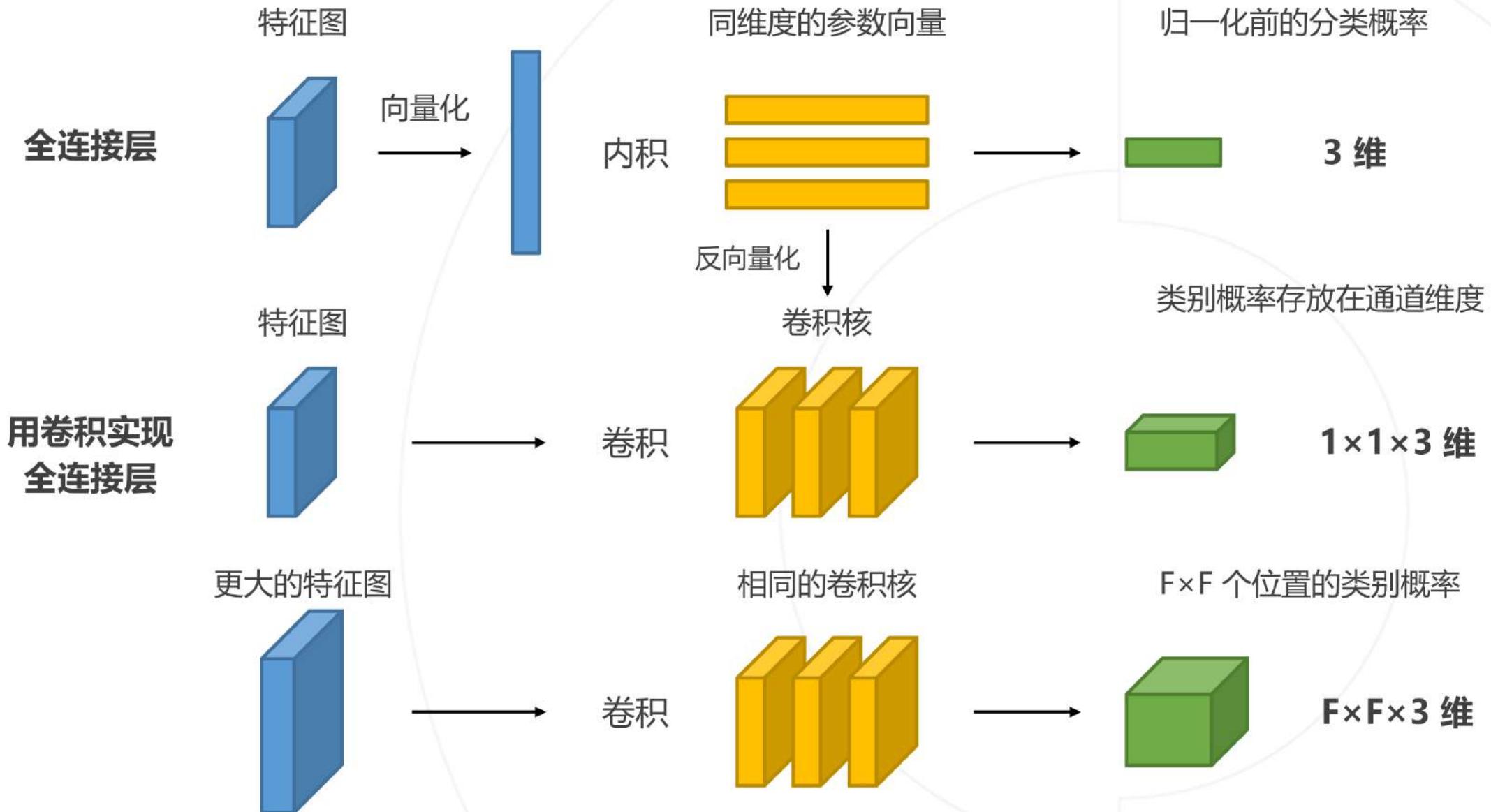
问题：全连接层要求固定输入大小



特征图

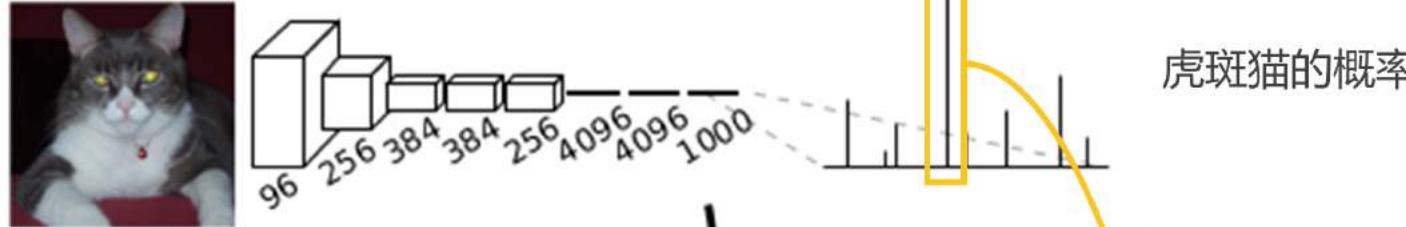


所有位置
的类别



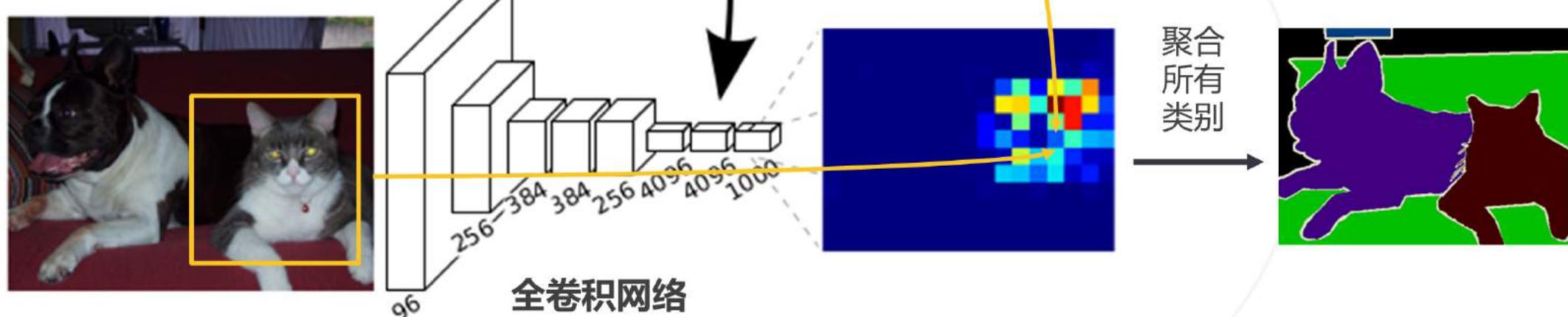
图像分类

固定尺寸输入



语义分割

任意尺寸输入



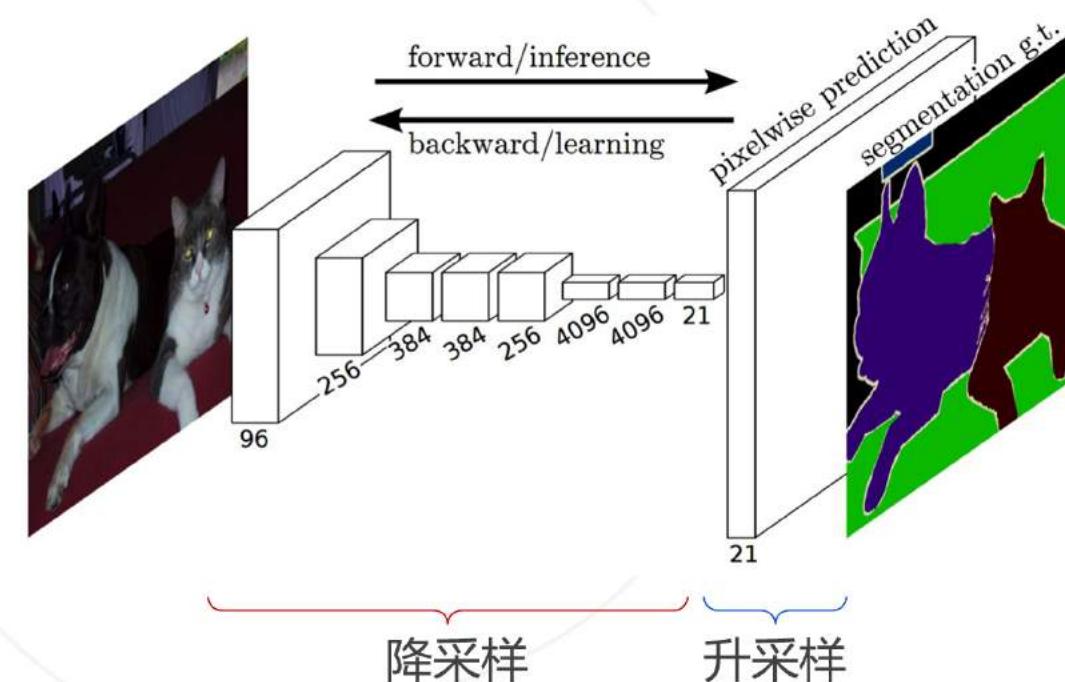
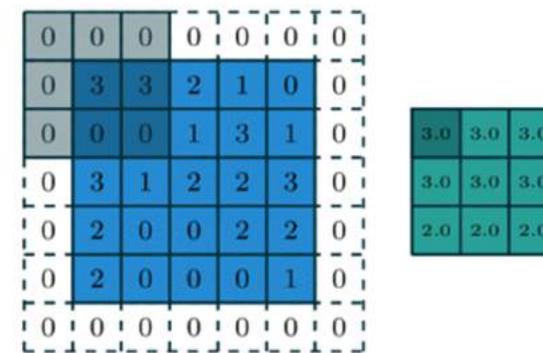
问题：

图像分类模型使用降采样层（步长卷积或池化）获得高层次特征，导致全卷积网络输出尺寸小于原图，而分割要求同尺寸输出

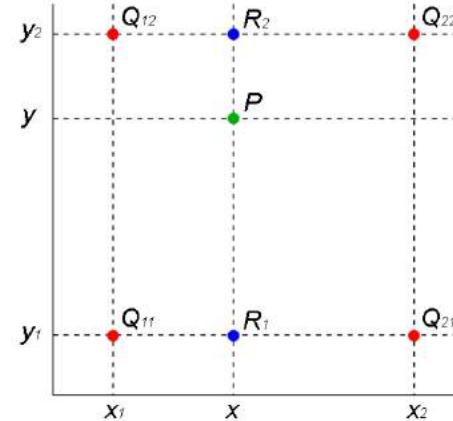
解决方法：

对预测的分割图升采样，恢复原图分辨率，升采样方案：

1. 双线性插值
2. 转置卷积：可学习的升采样层



双线性插值 Bilinear Interpolation



$$f(x, y) = \frac{1}{(x_2 - x_1)(y_2 - y_1)} \begin{bmatrix} x_2 - x & x - x_1 \end{bmatrix} \begin{bmatrix} f(Q_{11}) & f(Q_{12}) \\ f(Q_{21}) & f(Q_{22}) \end{bmatrix} \begin{bmatrix} y_2 - y \\ y - y_1 \end{bmatrix}.$$

1	2	3
4	5	6
7	8	9



1.0	1.5	2.0	2.5	3.0
2.5	3.0	3.5	4.0	4.5
4.0	4.5	5.0	5.5	6.0
5.5	6.0	6.5	7.0	7.5
7.0	7.5	8.0	8.5	9.0

卷积实现双线性插值

1	2	3
4	5	6
7	8	9

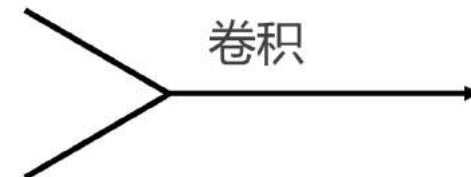


零插值

1.0		2.0		3.0
4.0		5.0		6.0
7.0		8.0		9.0

双线性插值对应的卷积核

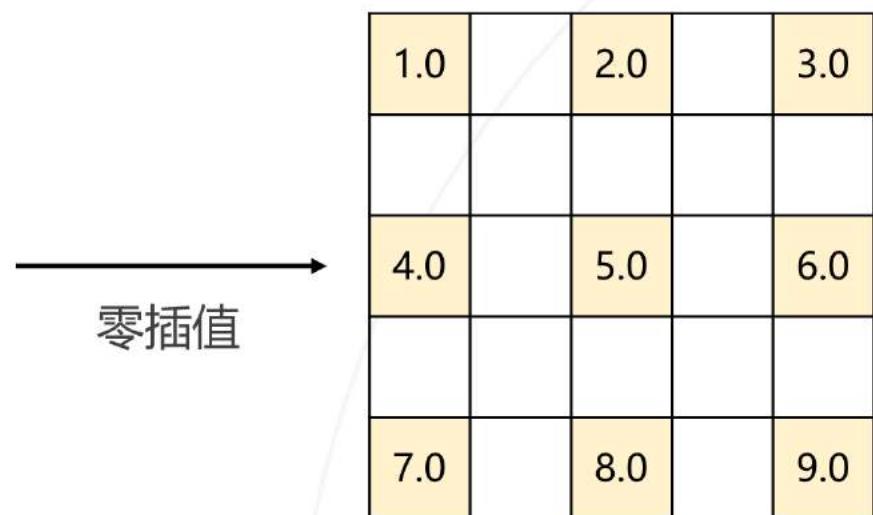
0.25	0.5	0.25
0.5	1	0.5
0.25	0.5	0.25



卷积

1.0	1.5	2.0	2.5	3.0
2.5	3.0	3.5	4.0	4.5
4.0	4.5	5.0	5.5	6.0
5.5	6.0	6.5	7.0	7.5
7.0	7.5	8.0	8.5	9.0

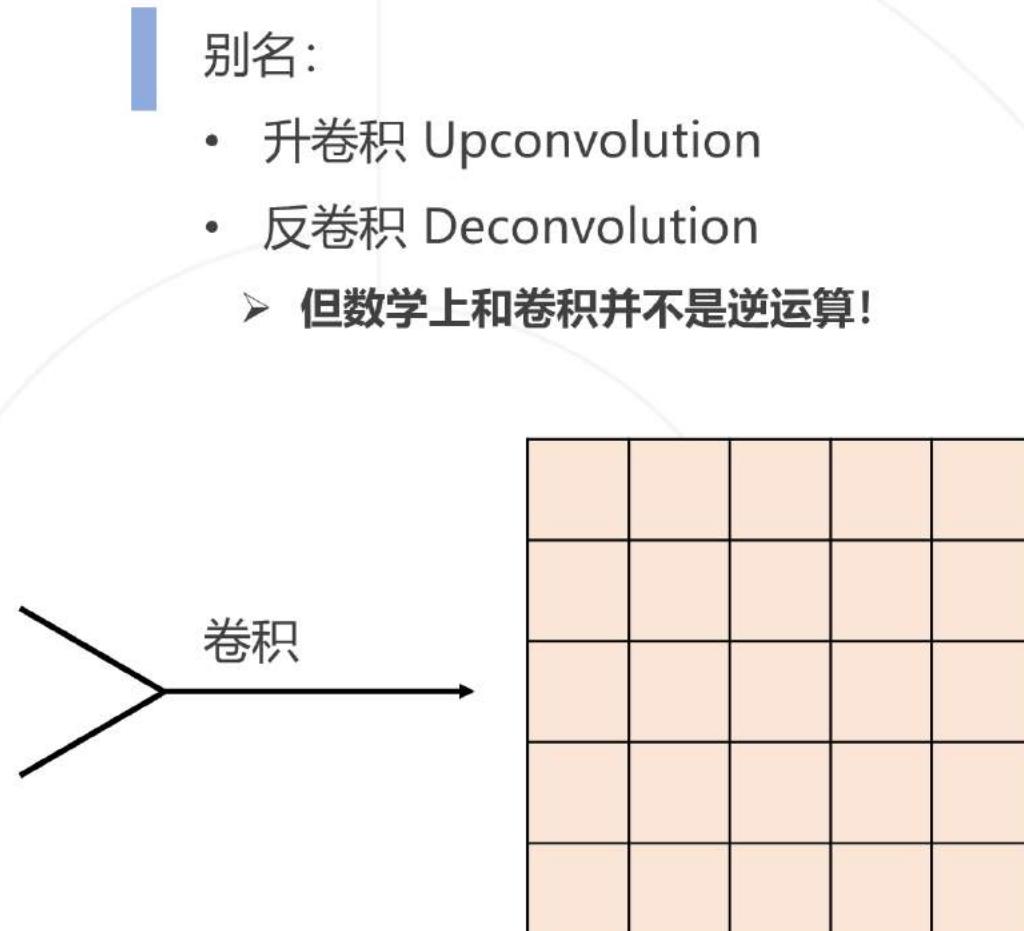
1	2	3
4	5	6
7	8	9



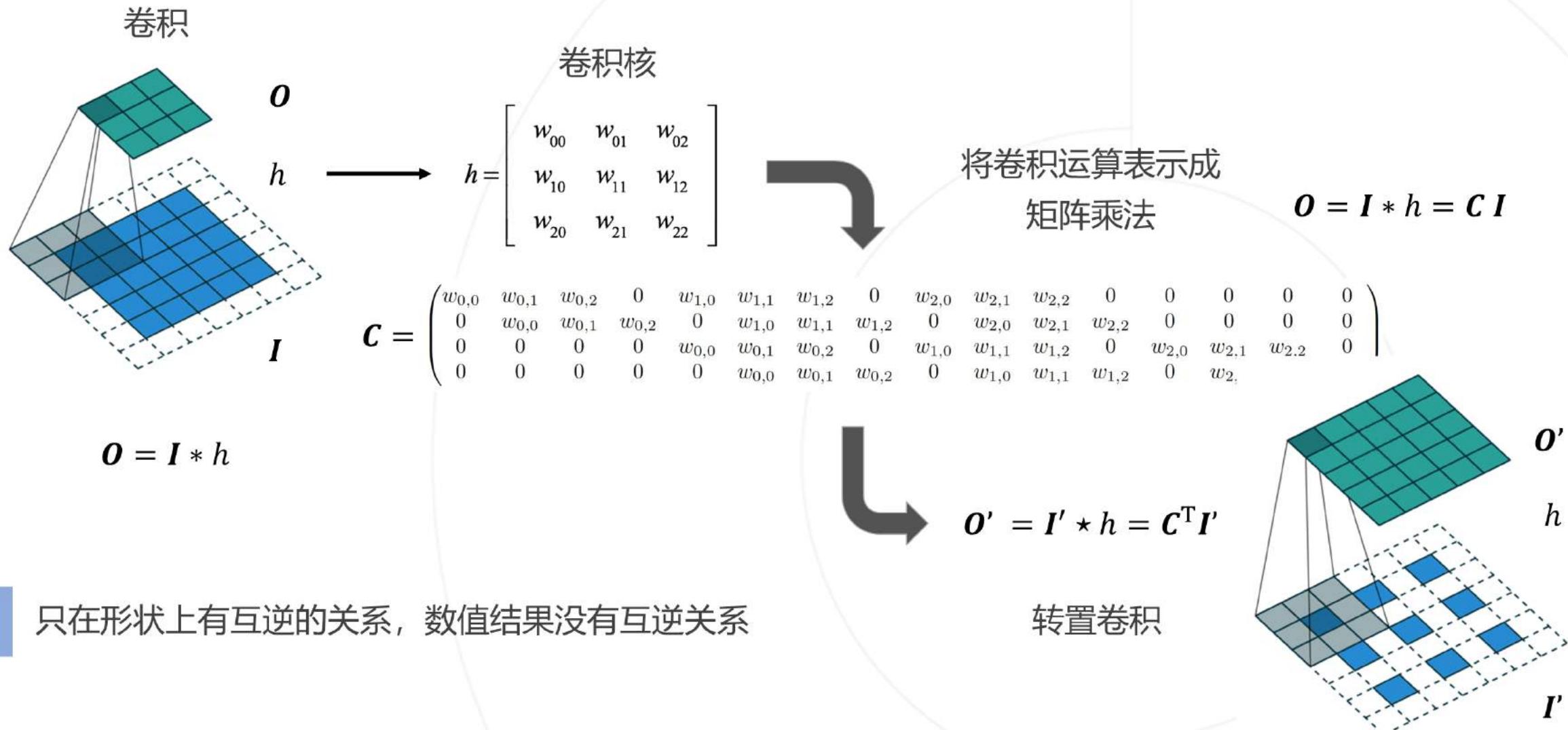
零插值

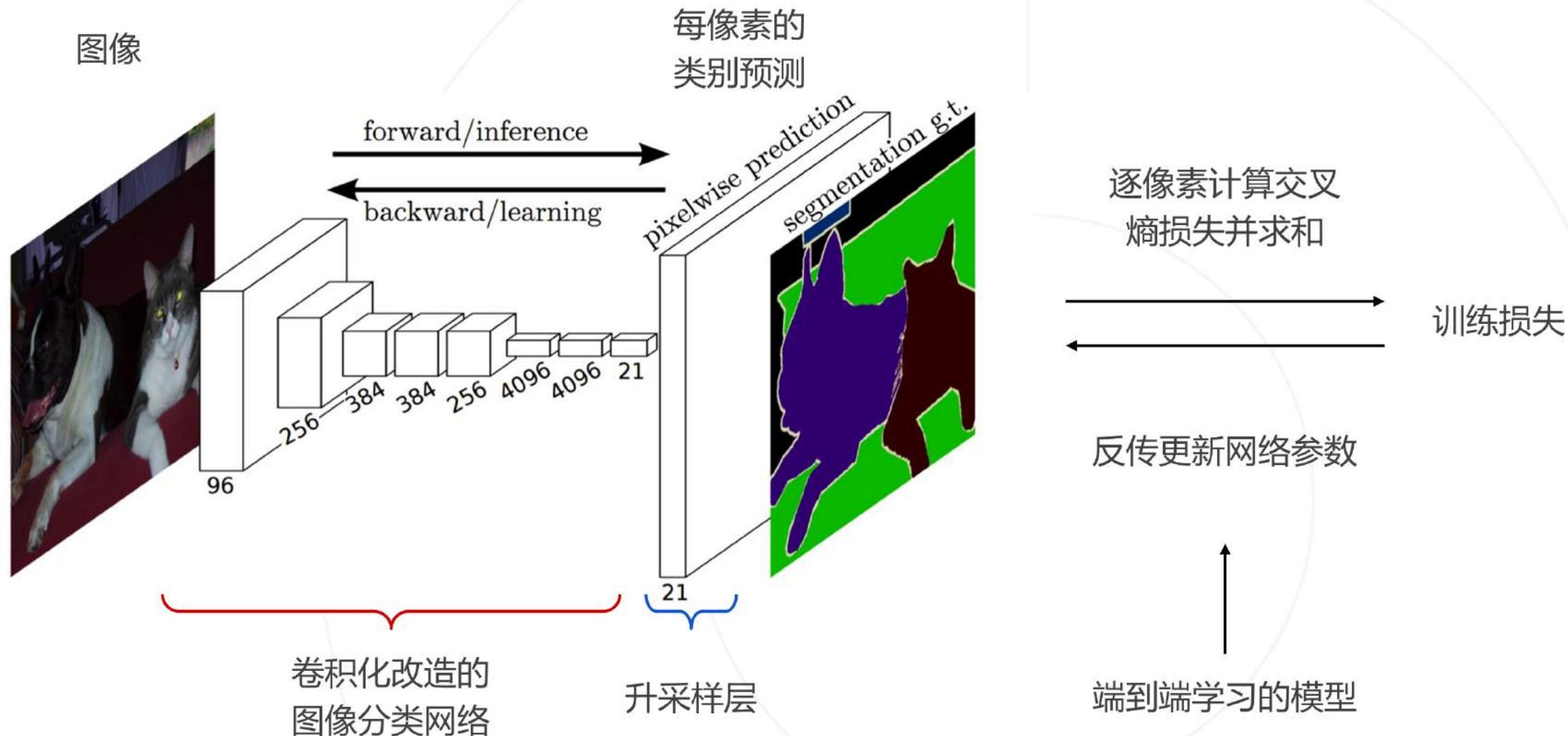
可学习的卷积核

a_{11}	a_{12}	a_{13}
a_{21}	a_{22}	a_{23}
a_{31}	a_{32}	a_{33}



升采样的特征图

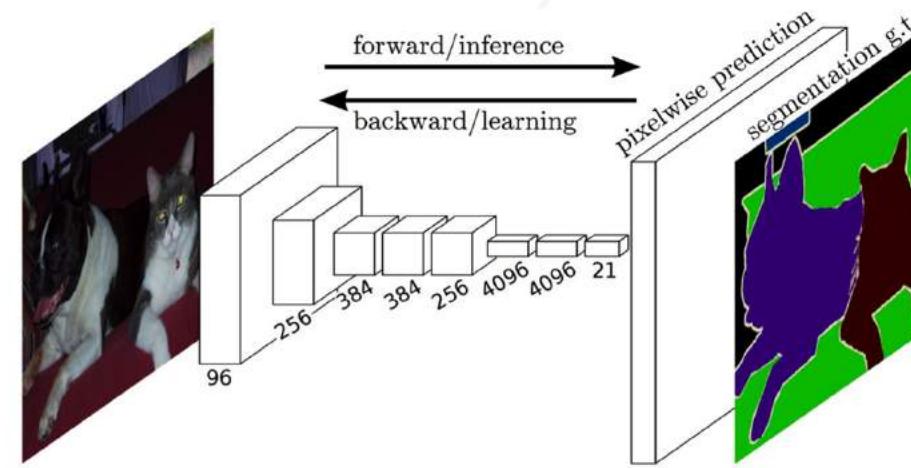




问题：基于顶层特征预测，再升采样 32 倍得到的预测图较为粗糙。

分析：高层特征经过多次降采样，细节丢失严重。

解决思路：结合低层次和高层次特征图。



细节信息丰富
语义信息贫乏

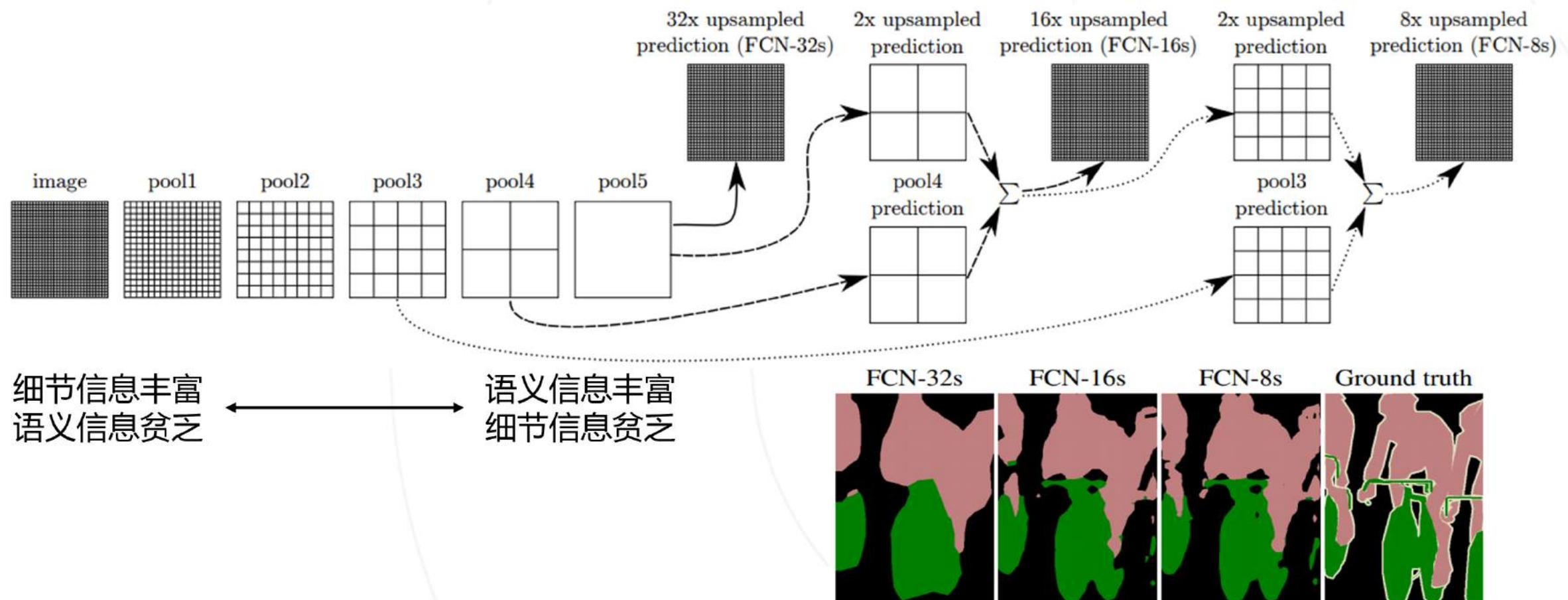


语义信息丰富
细节信息贫乏



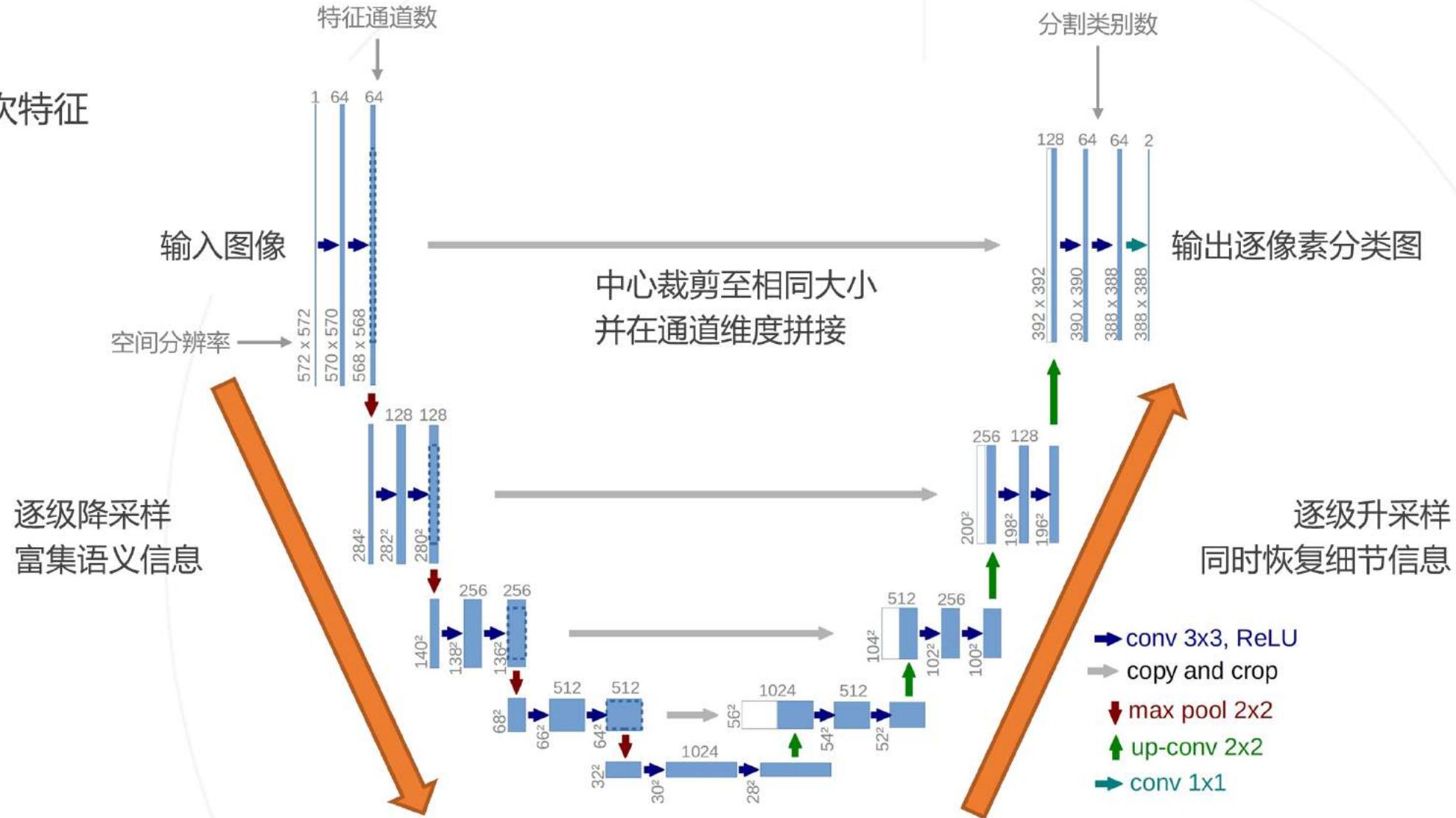
解决方案 FCN：

基于低层次和高层次特征图分别产生类别预测，升采样到原图大小，再平均得到最终结果



解决方案 UNet:

逐级融合高低层次特征



上下文信息



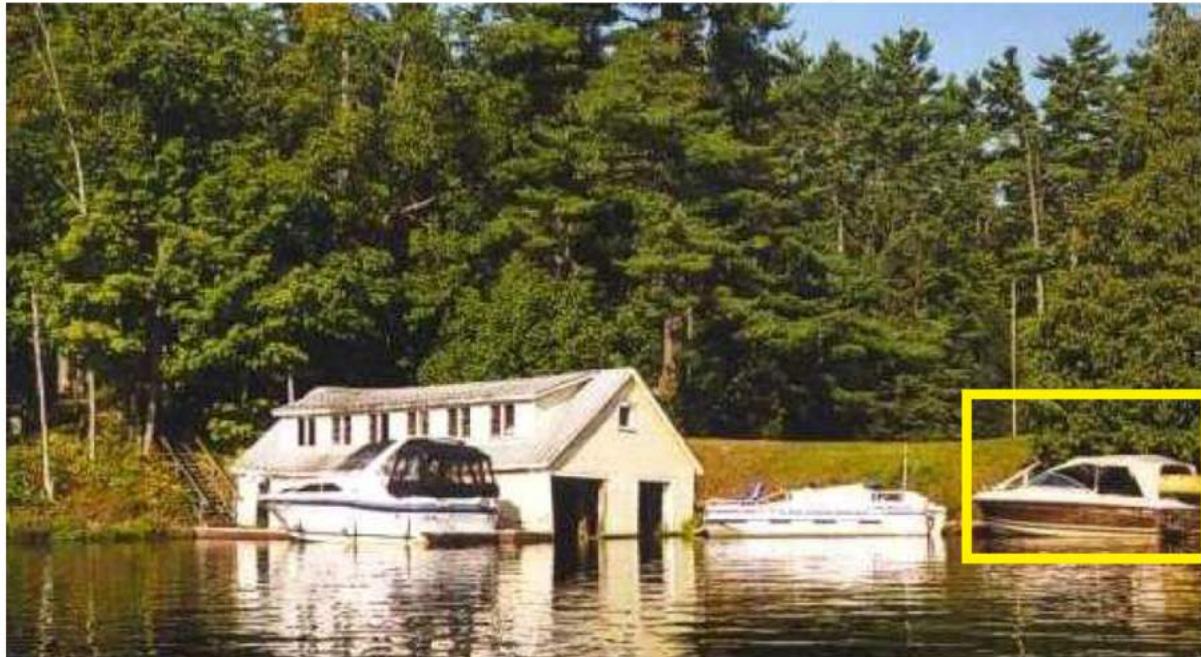
船?

汽车?

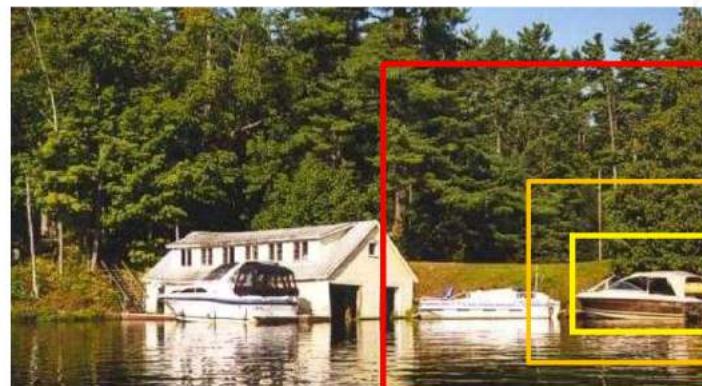
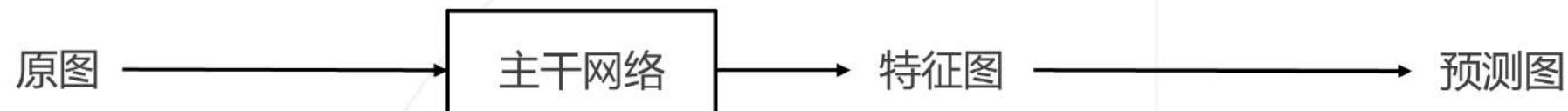


枕头?

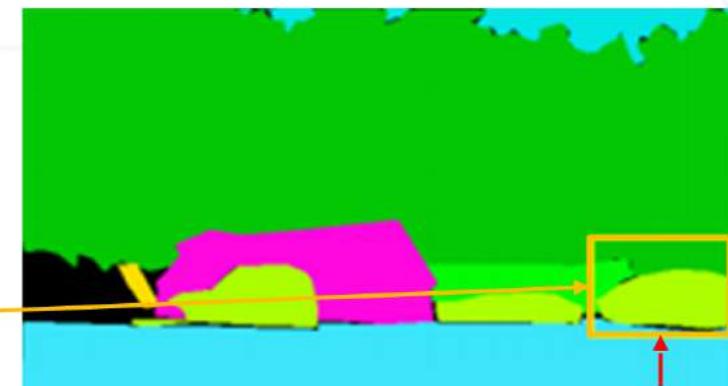
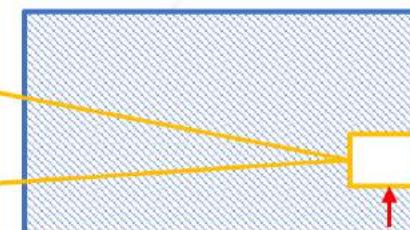
被子?



图像周围的内容（也称上下文）可以帮助我们做出更准确的判断。



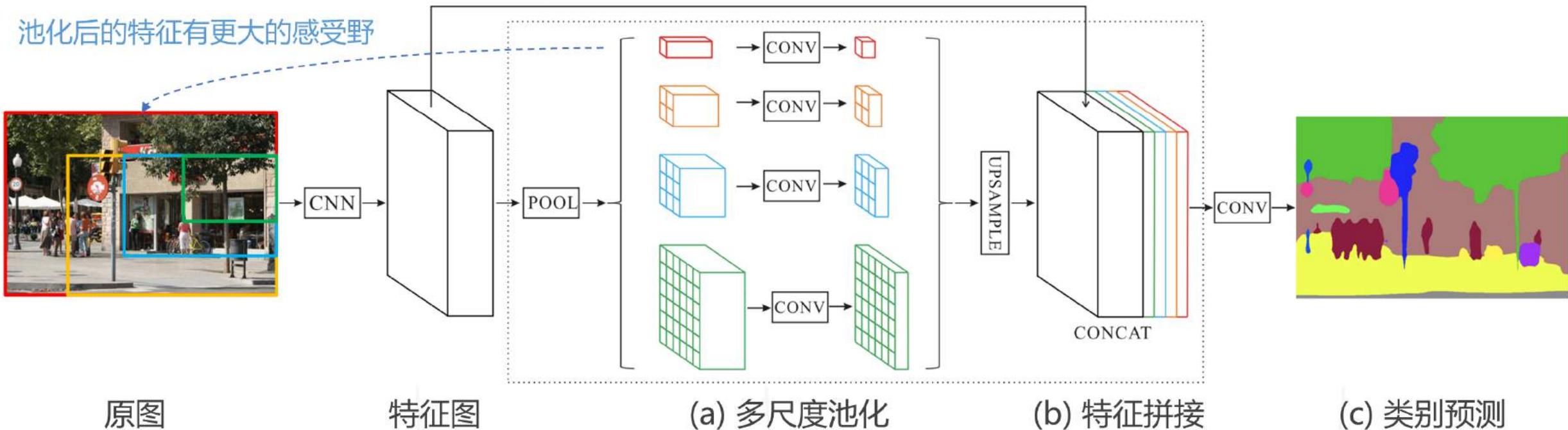
感受野受主干网络结构限制



如何在预测过程中使用上下文信息?

方案: 增加感受野更大的网络分支,
将上下文信息导入局部预测中

池化后的特征有更大的感受野



(a) 对特征图进行不同尺度的池化，得到不同尺度的上下文特征

(b) 上下文特征经过通道压缩和空间上采样之后拼接回原特征图 → 同时包含局部和上下文特征

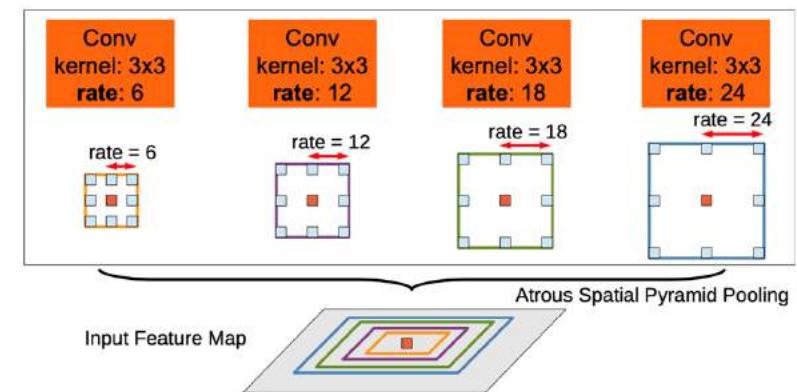
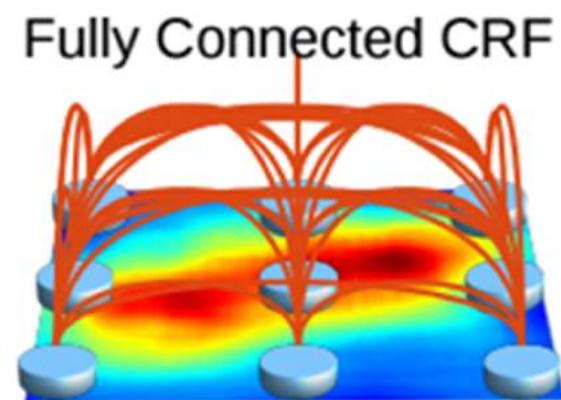
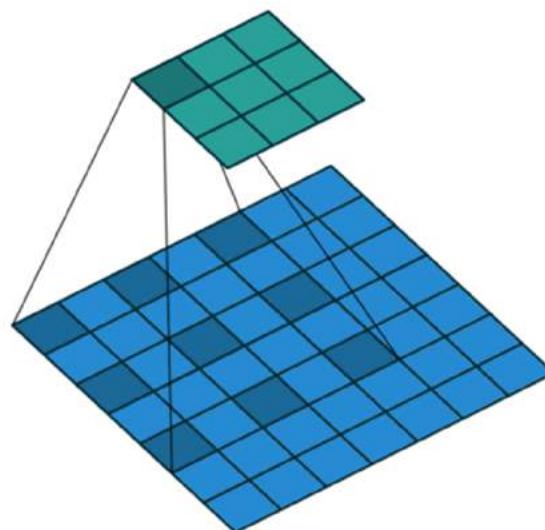
(c) 基于融合的特征产生预测图

空洞卷积与 DeepLab 系列算法

DeepLab 是语义分割的又一系列工作，其主要贡献为：

- 使用空洞卷积解决网络中的下采样问题
- 使用条件随机场 CRF 作为后处理手段，精细化分割图
- 使用多尺度的空洞卷积（ASPP 模块）捕捉上下文信息

DeepLab v1 发表于 2014 年，后于 2016、2017、2018 年提出 v2、v3、v3+ 版本。



图像分类模型中的下采样层使输出尺寸变小

如果将池化层和卷积中的步长去掉：

- 可以减少下采样的次数；
- 特征图就会变大，需要对应增大卷积核，以维持相同的感受野，但会增加大量参数
- 使用空洞卷积 (Dilated Convolution/Atrous Convolution)，在不增加参数的情况下增大感受野

标准卷积

1.0		2.0		3.0
4.0		5.0		6.0
7.0		8.0		9.0

下采样

1.0	2.0	3.0
4.0	5.0	6.0
7.0	8.0	9.0

卷积核

a_{11}	a_{12}	a_{13}
a_{21}	a_{22}	a_{23}
a_{31}	a_{32}	a_{33}

结果



卷积运算

膨胀卷积核
不产生额外参数

a_{11}		a_{12}		a_{13}
a_{21}		a_{22}		a_{23}
a_{31}		a_{32}		a_{33}

相同的计算结果

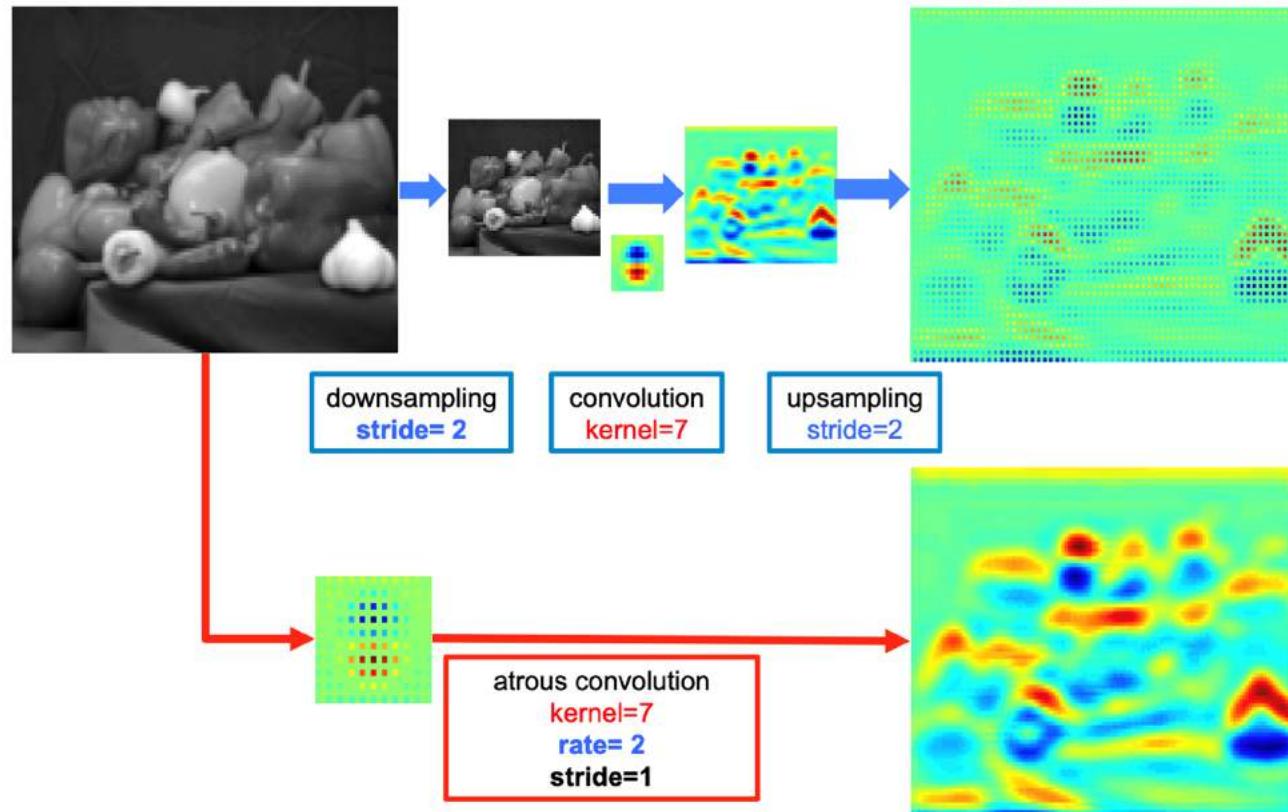


空洞卷积

1.0		2.0		3.0
4.0		5.0		6.0
7.0		8.0		9.0

特征图不变
膨胀卷积核
再进行卷积运算

**下采样加标准卷积
等价于空洞卷积**



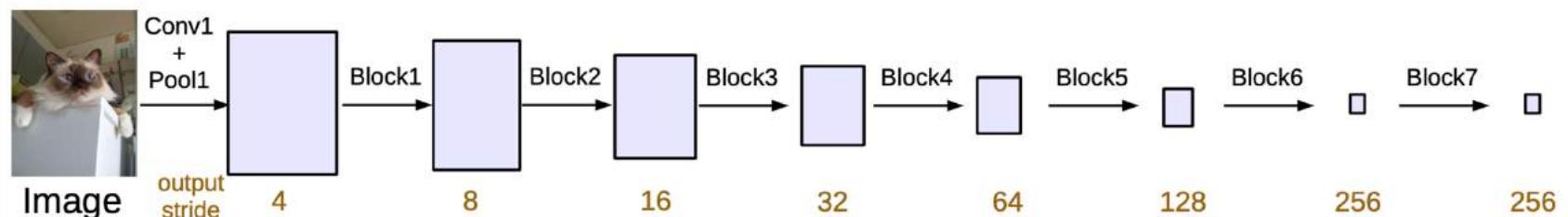
rate = 膨胀倍率，即参数之间的距离

使用升采样方案得到的特征图只有原图 1/4 位置的响应，需要配合插值

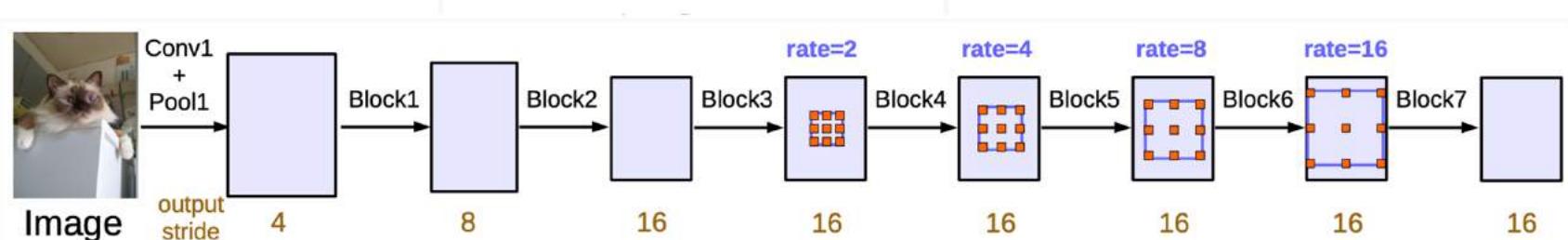
使用空洞卷积可以得到相同分辨率的特征图，且无需额外插值操作

DeepLab 在图像分类网络的基础上做了修改：

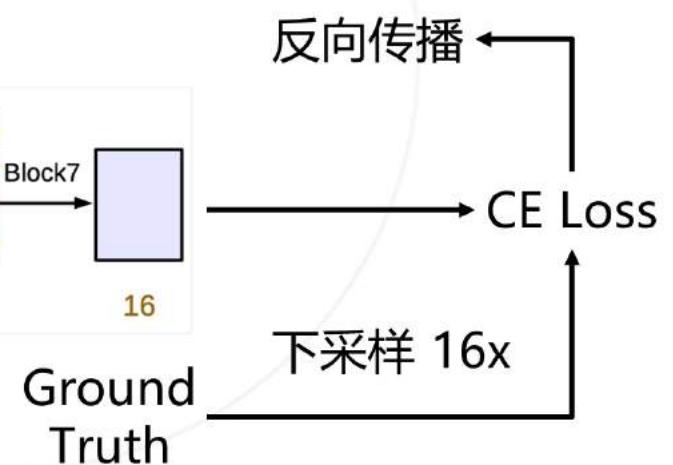
- 去除分类模型中的后半部分的下采样层
- 后续的卷积层改为膨胀卷积，并且逐步增加rate来维持原网络的感受野



含有降采样的卷积网络

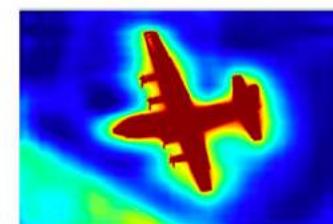
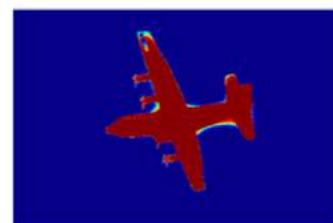
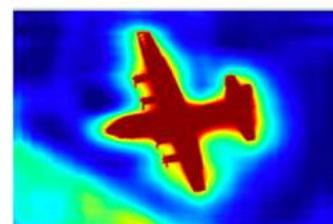
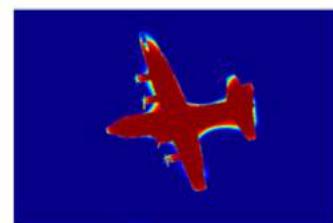
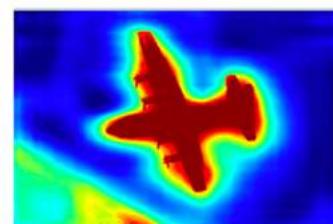
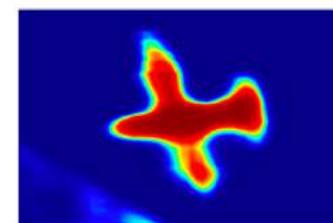
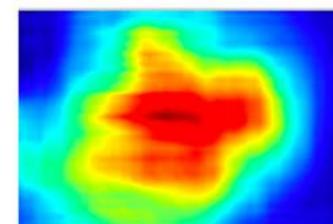


去除降采样并使用空洞卷积的网络



模型直接输出的分割图较为粗糙，尤其在物体边界处不能产生很好的分割结果。

DeepLab v1&v2 使用条件随机场 (CRF) 作为后处理手段，结合原图颜色信息和神经网络预测的类别得到精细化分割结果。



输入/真实分割

网络输出

CRF 第一次迭代

CRF 第二次迭代

CRF 第十次迭代

分割边界模糊

边界分割精确

CRF 是一种概率模型。DeepLab 使用 CRF 对分割结果进行建模，用能量函数用来表示分割结果优劣，通过最小化能量函数获得更好的分割结果。

能量函数
$$E(\boldsymbol{x}) = \sum_i \theta_i(x_i) + \sum_{ij} \theta_{ij}(x_i, x_j)$$

x_i, x_j	特定像素的预测结果（向量化后只有1维坐标）
\boldsymbol{x}	全部像素的预测结果
$\theta_i(x_i)$	单个预测对能量函数的贡献
$\theta_{i,j}(x_i, x_j)$	一对预测对能量函数的贡献

能量函数

$$E(\mathbf{x}) = \sum_i \theta_i(x_i) + \sum_{ij} \theta_{ij}(x_i, x_j)$$

鼓励后处理结果符合网络给出的结果

$P(x_i)$ = 网络输出的对应类别的概率

鼓励产生更好的分割边界

$$\theta_{ij}(x_i, x_j) = \mu(x_i, x_j) \left[w_1 \exp \left(-\frac{\|p_i - p_j\|^2}{2\sigma_\alpha^2} \right) - \frac{\|I_i - I_j\|^2}{2\sigma_\beta^2} \right] + w_2 \exp \left(-\frac{\|p_i - p_j\|^2}{2\sigma_\gamma^2} \right)$$

仅当类别不同时产生惩罚

$$\mu(x_i, x_j) = [x_i \neq x_j]$$

位置相近且颜色相近时惩罚

鼓励仅在原图颜色边界处
产生类别变化

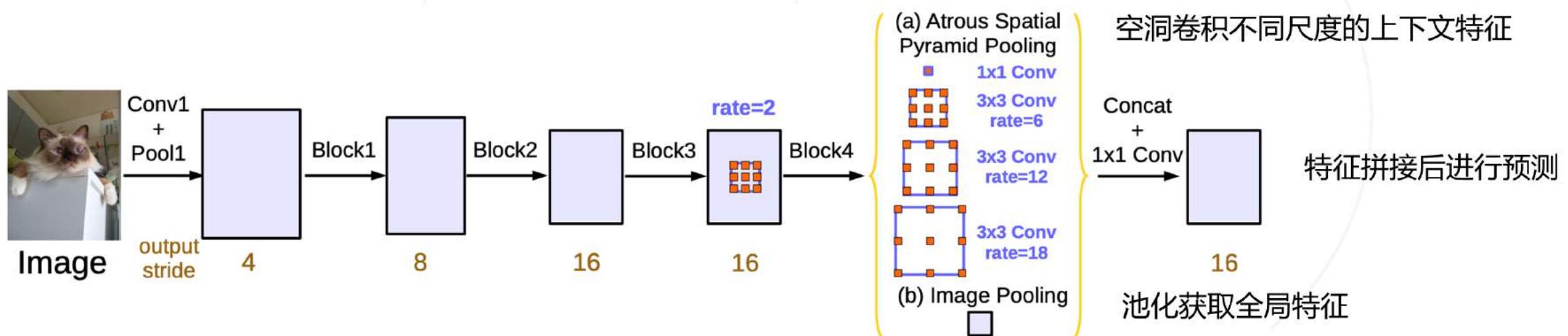
位置相近时惩罚

鼓励产生平滑的结果

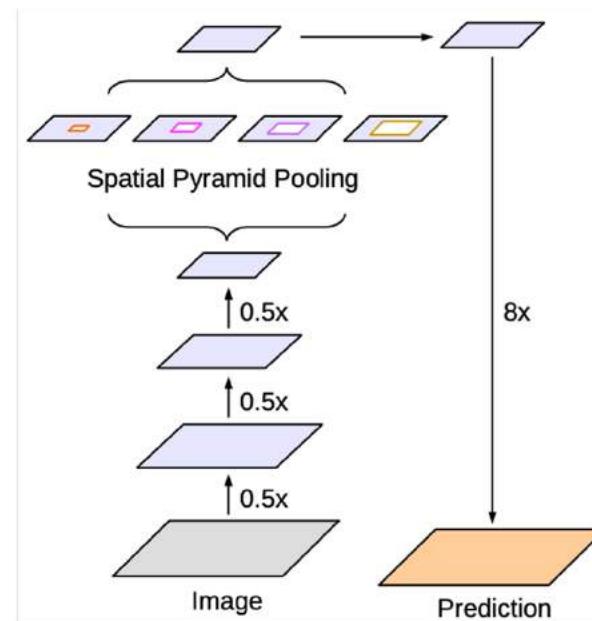
PSPNet 使用不同尺度的池化来获取不同尺度的上下文信息

DeepLab v2 & v3 使用不同尺度的空洞卷积达到类似的效果

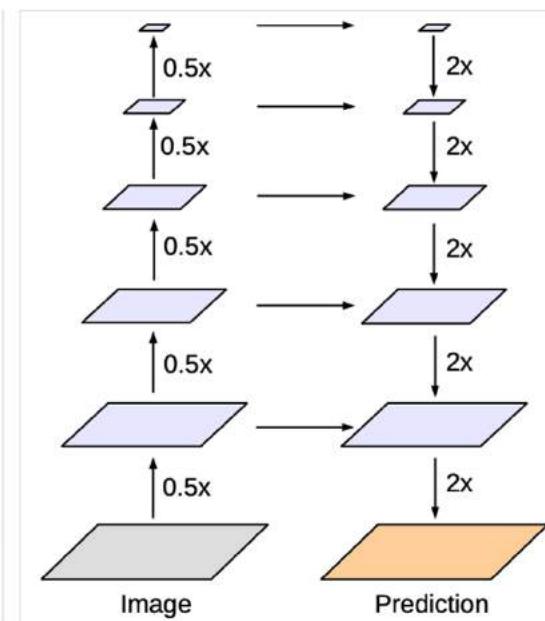
更大膨胀率的空洞卷积 → 更大的感受野 → 更多的上下文特征



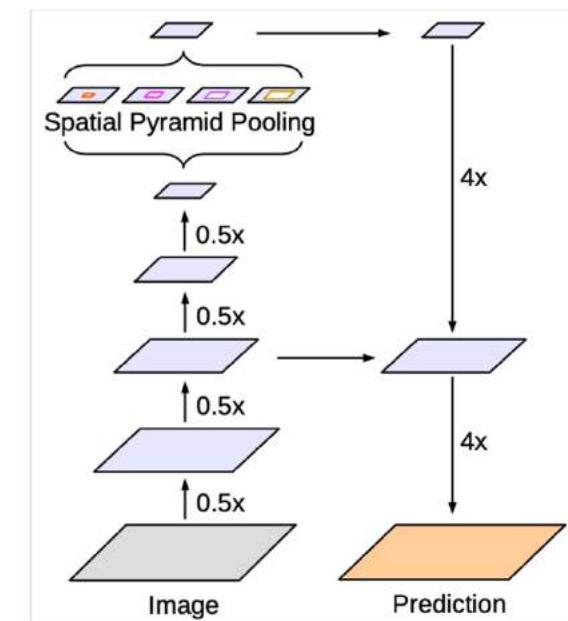
- DeepLab v2 / v3 模型使用 ASPP 捕捉上下文特征
- Encoder / Decoder 结构 (如 UNet) 在上采样过程中融入低层次的特征图，以获得更精细的分割图
- DeepLab v3+ 将两种思路融合，在原有模型结构上增加了一个简单的 decoder 结构



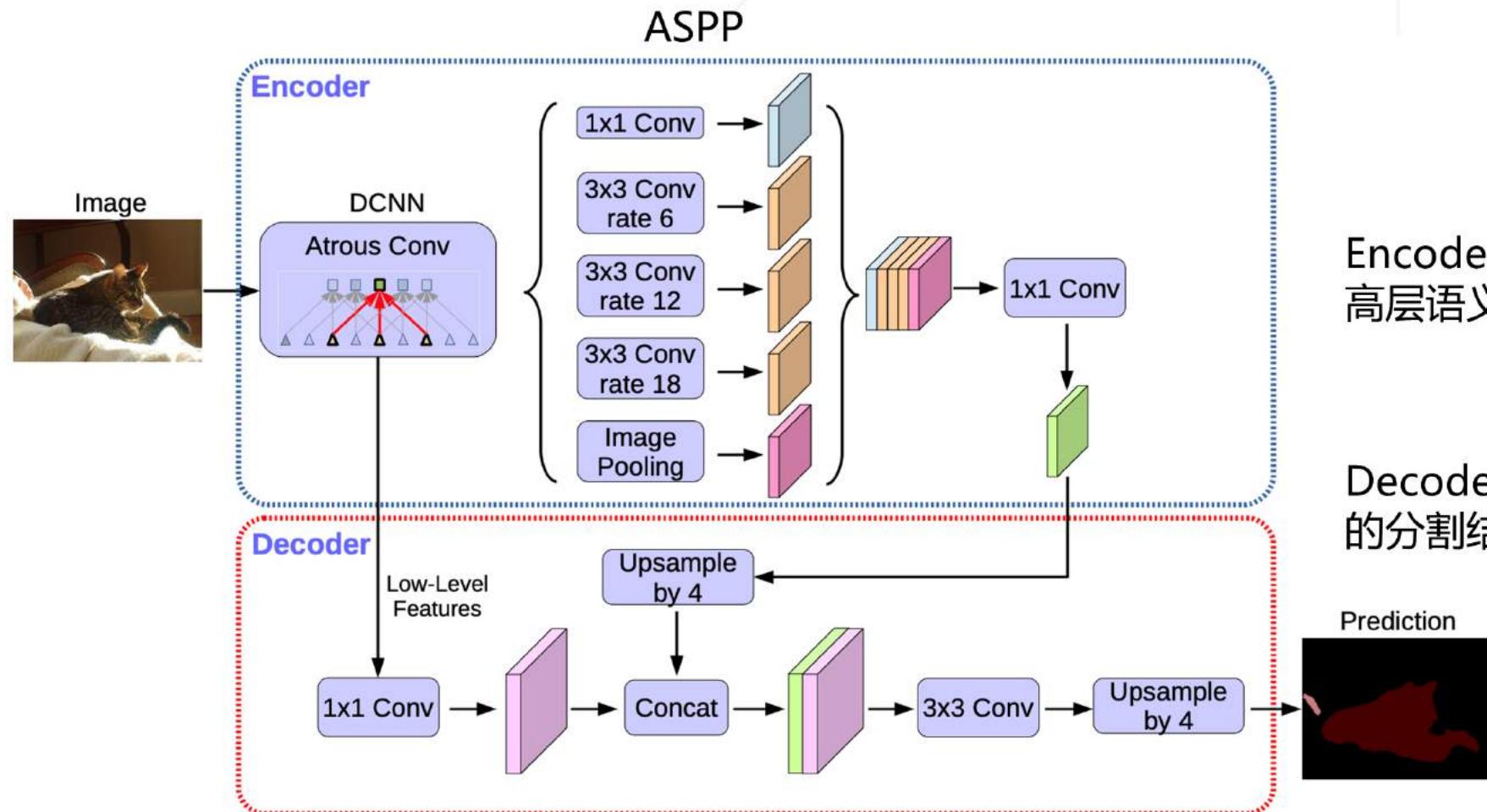
(a) ASPP
DeepLab v2 / v3



(b) Encoder / Decoder
UNet



(c) DeepLab v3+



Encoder 通过 ASPP 产生多尺度的高层语义信息

Decoder 主要融合低层特征产生惊喜的分割结果

Prediction

基本思想

滑窗预测

全卷积网络 FCN

恢复网络中的降采样

朴素思路
双线性插值

可学习的升采样
转置卷积

更大感受野
空洞卷积

FCN UNet

DeepLab 系列

恢复预测图的细节

分别基于高低层特征预测

融合高低层特征再预测

CRF 后处理

FCN

UNet DeepLab v3+

DeepLab v1/v2

利用上下文信息

池化金字塔

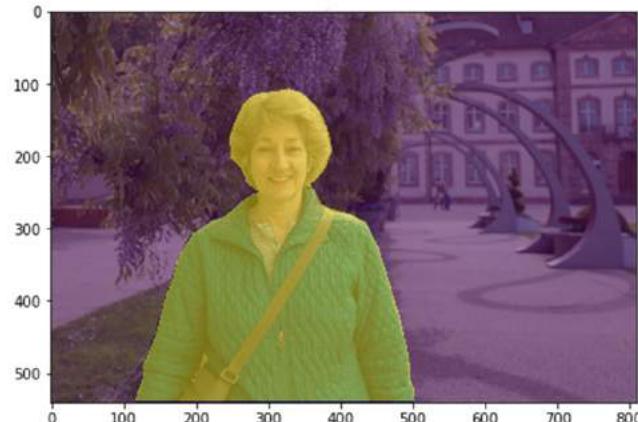
多尺度空洞卷积 ASPP

PSPNet

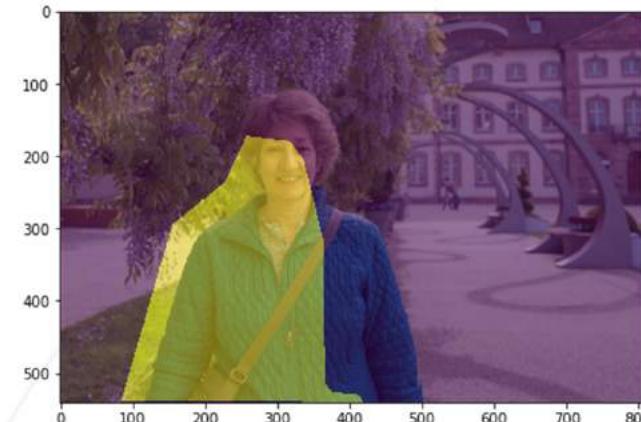
DeepLab v3/v3+

语义分割模型的评估

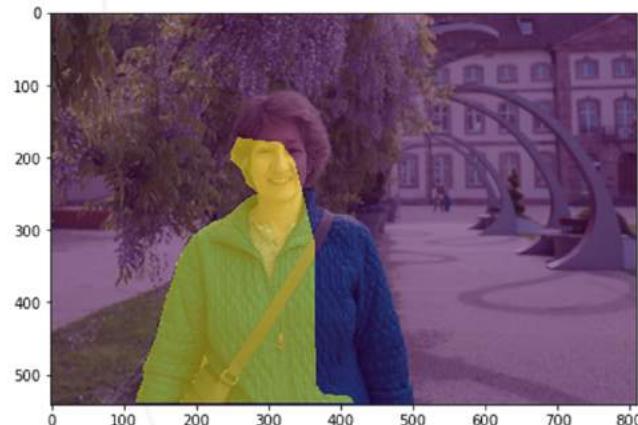
真实分割图



预测分割图



交集



并集



$$\text{Accuracy} = \frac{\text{面积比值}}{\text{GT}}$$

$$\text{IoU} = \frac{\text{面积比值}}{\text{GT} + \text{Pred}}$$

$$\text{Dice} = \frac{2 \times \text{面积比值}}{\text{GT} + \text{Pred}}$$

mAcc
 mIoU
 mDice = 对每类计算指标
 再按类别平均

并集面积的计算方法

$$\text{并集面积} = \text{GT} + \text{Pred} - \text{交集面积}$$

谢谢大家