

法律声明

□ 本课件包括：演示文稿，示例，代码，题库，视频和声音等，小象学院拥有完全知识产权的权利；只限于善意学习者在本课程使用，不得在课程范围外向任何第三方散播。任何其他人或机构不得盗版、复制、仿造其中的创意，我们将保留一切通过法律手段追究违反者的权利。

□ 课程详情请咨询

■ 微信公众号：小象

■ 新浪微博：ChinaHadoop



第6课 图像分割（下）

Image Segmentation

主讲人：张宗健

悉尼科技大学博士

主要研究方向： 计算机视觉、视觉场景理解、图像&语言、深度学习
图像检索CbIR、Human ReID等

本章结构

□ 语义分割 (Semantic Segmentation)

- DeepLab

□ 三大数据集介绍 (Pascal VOC, MSCOCO, Cityscapes)

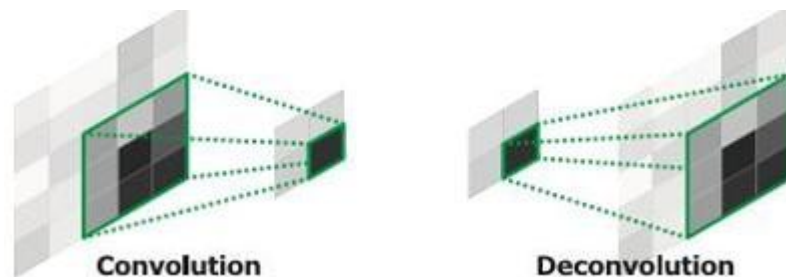
□ 代码演示

- DeepLab

语义分割（Semantic Segmentation）

FCN-反卷积（Deconvolution）

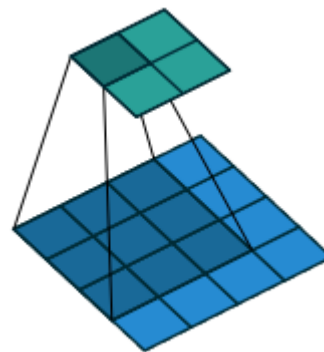
- 一对多操作
- 卷积的逆操作
 - 小数步长 $1/f$
 - 卷积核尺寸不变
- 前向和后向传播
 - 对应于卷积操作的后向和前向传播，优化上做颠倒
 - 在矩阵化操作中，反卷积核矩阵是卷积核矩阵的转置，学习率为0
- 准确叫法：转置卷积（Transposed convolution）



语义分割（Semantic Segmentation）

FCN-反卷积（Deconvolution）

- 卷积操作矩阵化
 - 卷积参数
 - 输入→输出：4x4→2x2
 - 核尺寸：3x3
 - Padding/步长：0/1
 - 矩阵乘
 - 输入展开为16维向量：x
 - 输出展开为4维向量：y
 - 卷积→ $y=Cx$



$$\begin{pmatrix} w_{0,0} & w_{0,1} & w_{0,2} & 0 & w_{1,0} & w_{1,1} & w_{1,2} & 0 & w_{2,0} & w_{2,1} & w_{2,2} & 0 & 0 & 0 & 0 & 0 \\ 0 & w_{0,0} & w_{0,1} & w_{0,2} & 0 & w_{1,0} & w_{1,1} & w_{1,2} & 0 & w_{2,0} & w_{2,1} & w_{2,2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & w_{0,0} & w_{0,1} & w_{0,2} & 0 & w_{1,0} & w_{1,1} & w_{1,2} & 0 & w_{2,0} & w_{2,1} & w_{2,2} & 0 \\ 0 & 0 & 0 & 0 & 0 & w_{0,0} & w_{0,1} & w_{0,2} & 0 & w_{1,0} & w_{1,1} & w_{1,2} & 0 & w_{2,0} & w_{2,1} & w_{2,2} \end{pmatrix}$$

语义分割（Semantic Segmentation）

FCN-反卷积（Deconvolution）

- 反卷积操作矩阵化

- 卷积参数

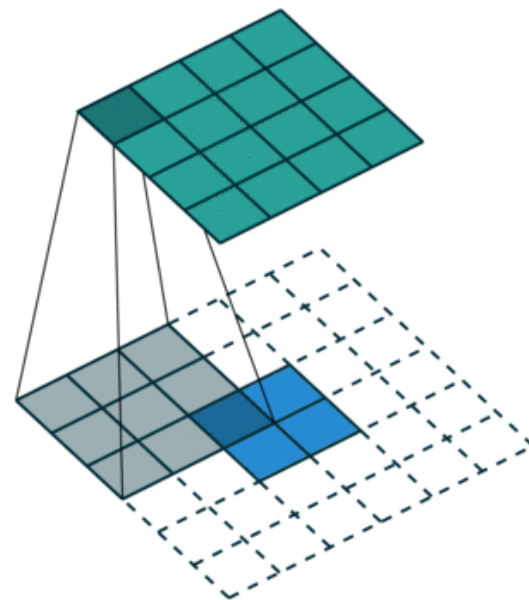
- 输入→输出：2x2→4x4
- 核尺寸：3x3
- Padding/步长：2/1

- 矩阵乘

- 输出展开为4维向量：y
- 输入展开为16维向量：x
- 卷积→ $x=C^T y$

$$\frac{\partial \text{Loss}}{\partial x_j} = \sum_i \frac{\partial \text{Loss}}{\partial y_i} \frac{\partial y_i}{\partial x_j} = \sum_i \frac{\partial \text{Loss}}{\partial y_i} C_{i,j} = \frac{\partial \text{Loss}}{\partial y} \cdot C_{*,j} = C_{*,j}^T \frac{\partial \text{Loss}}{\partial y}$$

$$\frac{\partial \text{Loss}}{\partial x} = \begin{bmatrix} \frac{\partial \text{Loss}}{\partial x_1} \\ \frac{\partial \text{Loss}}{\partial x_2} \\ \dots \\ \frac{\partial \text{Loss}}{\partial x_n} \end{bmatrix} = \begin{bmatrix} C_{*,1}^T \frac{\partial \text{Loss}}{\partial y} \\ C_{*,2}^T \frac{\partial \text{Loss}}{\partial y} \\ \dots \\ C_{*,n}^T \frac{\partial \text{Loss}}{\partial y} \end{bmatrix} = \begin{bmatrix} C_{*,1}^T \\ C_{*,2}^T \\ \dots \\ C_{*,n}^T \end{bmatrix} \frac{\partial \text{Loss}}{\partial y} = C^T \frac{\partial \text{Loss}}{\partial y}$$

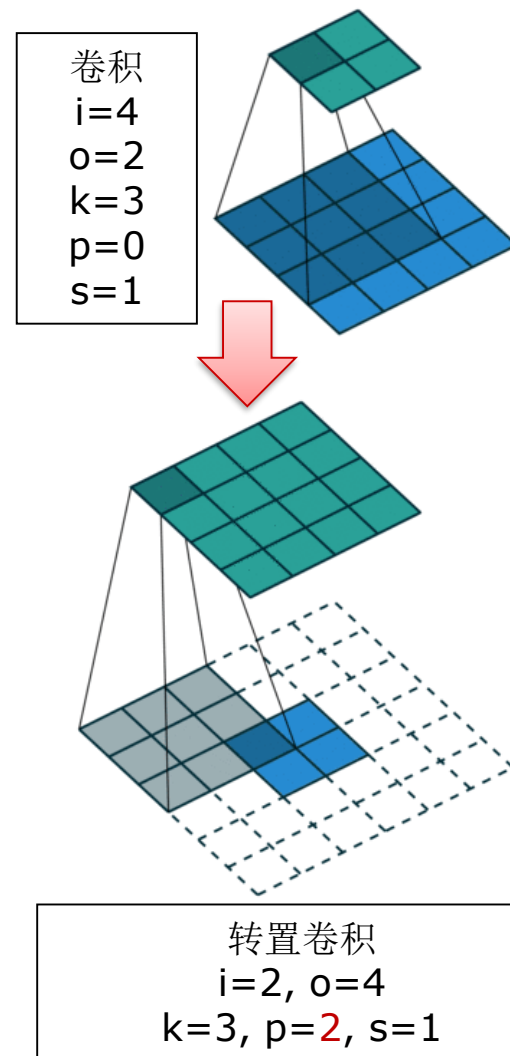


语义分割（Semantic Segmentation）

FCN-卷积/转置卷积的参数关系

- 步长: 1
- padding: 0

	卷积	转置卷积
核尺寸	k	$k' = k$
步长	$s = 1$	$s' = s$
Padding	$p = 0$	$p' = k - 1$
输入	i	$i' = o$
输出	o	$o' = i' + (k-1)$

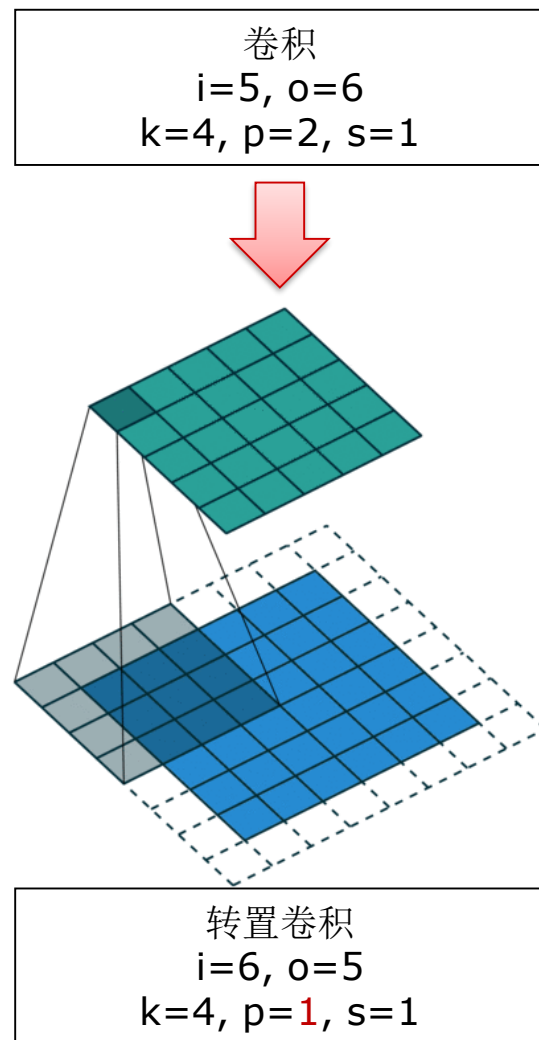


语义分割（Semantic Segmentation）

FCN-卷积/转置卷积的参数关系

- 步长: 1
- padding: > 0

	卷积	转置卷积
核尺寸	k	$k' = k$
步长	$s = 1$	$s' = s$
Padding	p	$p' = k - p - 1$
输入	i	$i' = o$
输出	o	$o' = i' + (k-1) - 2p$

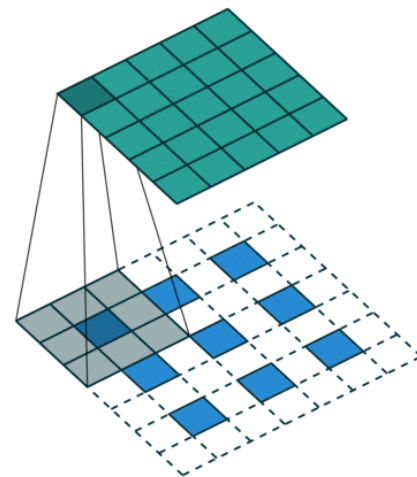
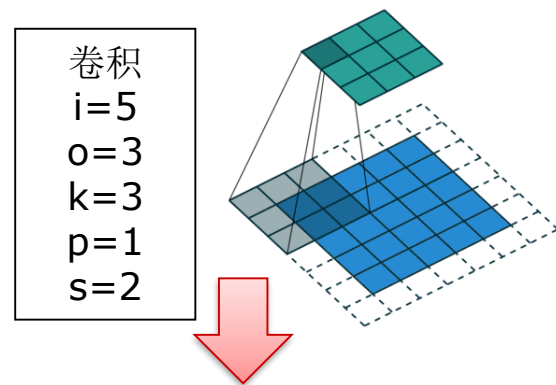


语义分割（Semantic Segmentation）

FCN-卷积/转置卷积的参数关系

- 步长: > 1
- padding: > 0
- $a = i + 2p - k$ 整除 s

	卷积	转置卷积
核尺寸	k	$k' = k$
步长	$s > 1$	$s' = 1$
Padding	p	$p' = k - p - 1$
输入	i	$i' = o$ 插零输入: $i'' = i' + (s-1)(i'-1)$
输出	o	$o' = s(i'-1) + k - 2p$



转置卷积
 $i=5, o=3$
 $k=3, p=1, s=1$

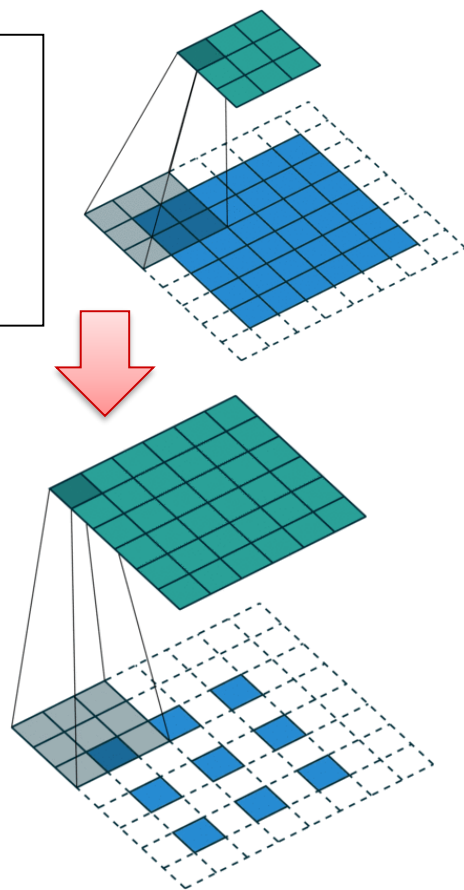
语义分割（Semantic Segmentation）

FCN-卷积/转置卷积的参数关系

- 步长: > 1
- padding: > 0
- $a = i + 2p - k$ 不整除 s

	卷积	转置卷积
核尺寸	k	$k' = k$
步长	$s > 1$	$s' = 1$
Padding	p	$p' = k - p - 1$ 上/右: $p'' = a \bmod s$
输入	i	$i' = o$ 插零输入: $i'' = i' + (s-1)(i'-1)$
输出	o	$o' = s(i'-1) + a + k - 2p$

卷积
 $i=6$
 $o=3$
 $k=3$
 $p=1$
 $s=2$



转置卷积
 $i=5, o=6$
 $k=3, p=(1,2,2,1), s=1$

语义分割（Semantic Segmentation）

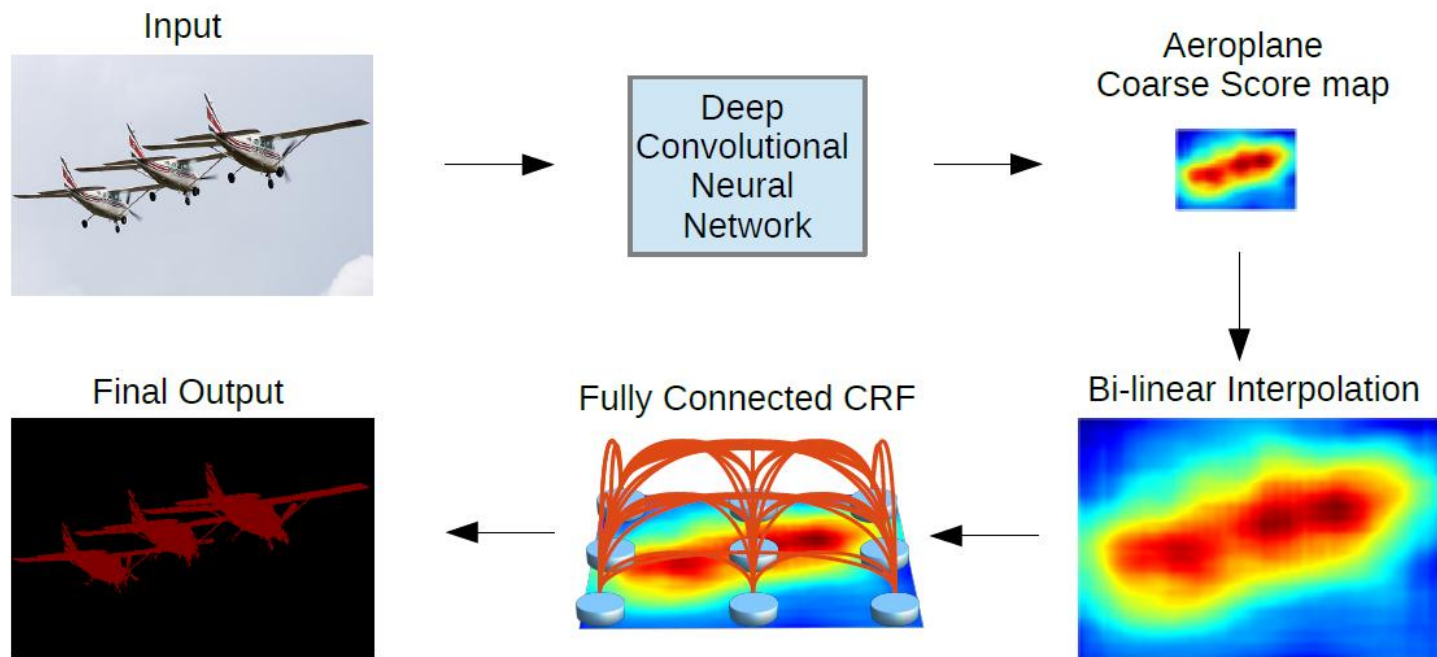
DeepLab全卷积网络

- 基本结构
 - 优化后的DCNN+传统的CRF图模型
- 新的上采样卷积方案
 - 带孔（hole）结构的膨胀卷积（Atrous/Dilated convolution）
- 多尺度图片表达
 - Atrous空间金字塔池化（Atrous Spatial Pyramid Pooling）
- 边界分割的优化
 - 使用全连接条件随机场CRF进行迭代优化

语义分割（Semantic Segmentation）

DeepLab全卷积网络

- 模块1：DCNN输出粗糙的分割结果
- 模块2：全连接CRF精化分割结果



语义分割（Semantic Segmentation）

DeepLab-DCNN

- 孔（Hole）算法
 - 解决原始FCN网络的输出低分辨率问题（100padding）
 - 降低池化层的降采样倍数
 - VGG16网络Pool4和Pool5层的步长：2→1
 - 减小降采样倍数：32→8
 - 后续卷积核的感受野（Field-Of-View）会受影响（变小）
 - 这些卷积核无法用来fine-tune
 - 更改卷积核的结构→加孔（Hole）
 - 无上采样功能
 - 恢复感受野，可以用来fine-tune
 - 保证了网络最终的高分辨率输出（仅8倍降采样）

语义分割（Semantic Segmentation）

DeepLab-DCNN

- 孔（Hole）算法

- 卷积核结构

- 尺寸不变（ 3×3 ），元素间距变大（ $1 \rightarrow 2$ ）

- 步长不变（1）

- 优势

- 参数数量不变

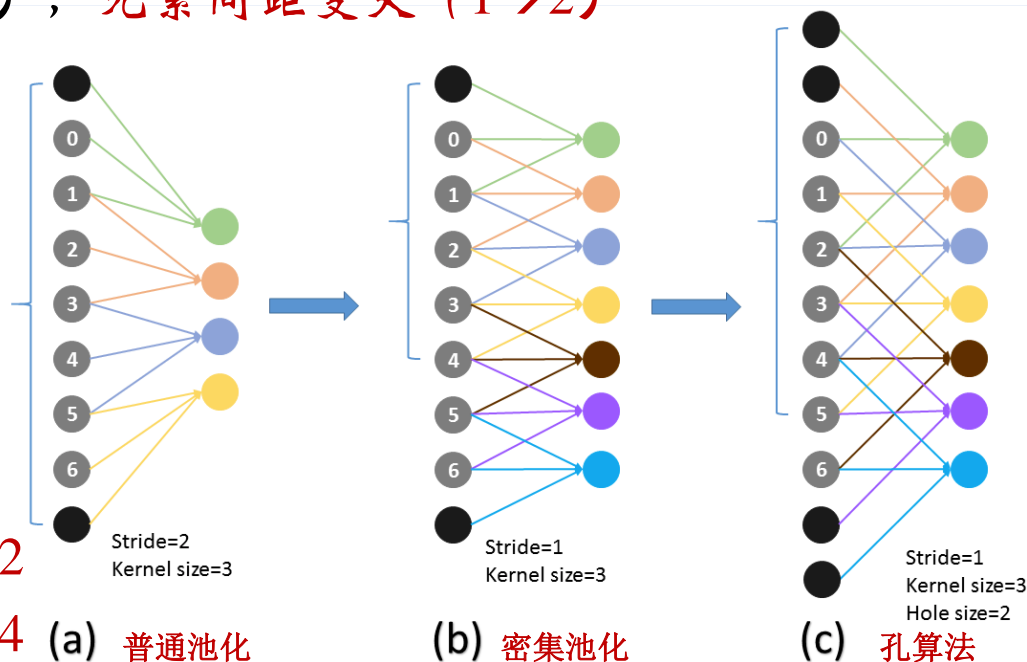
- 计算量不变

- 高分辨输出

- 采用层

- Conv5：孔尺寸2

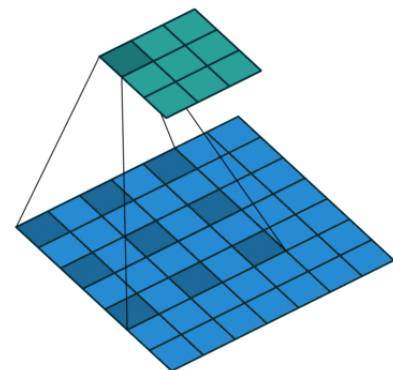
- Conv6：孔尺寸4



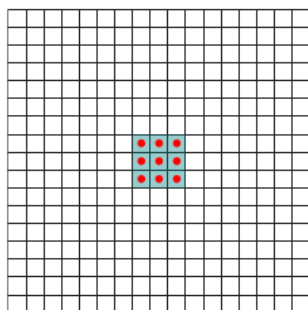
语义分割（Semantic Segmentation）

DeepLab-DCNN

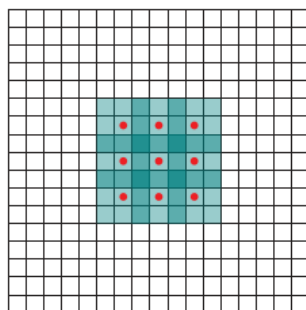
- 膨胀卷积（Atrous/Dilated convolution）
 - 孔算法的正式名称
 - 与降低池化层步长配对使用，以取代上采样反卷积
 - 孔尺寸 \rightarrow Rate
 - Rate 越大，感受野越大



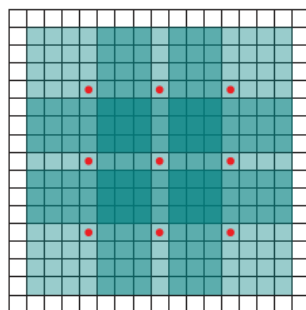
卷积核尺寸 3×3



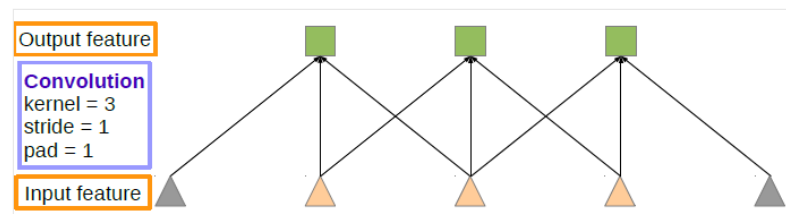
Rate = 1
无插零



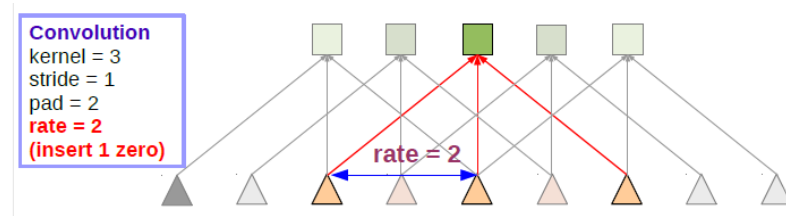
Rate = 2
插1个零



Rate = 4
插3个零



(a) Sparse feature extraction

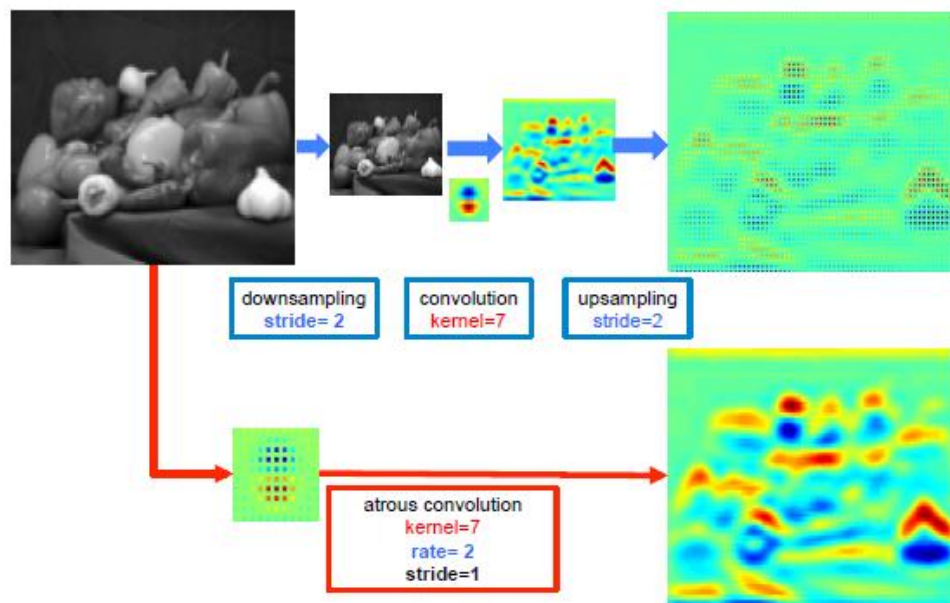


(b) Dense feature extraction

语义分割（Semantic Segmentation）

DeepLab-DCNN

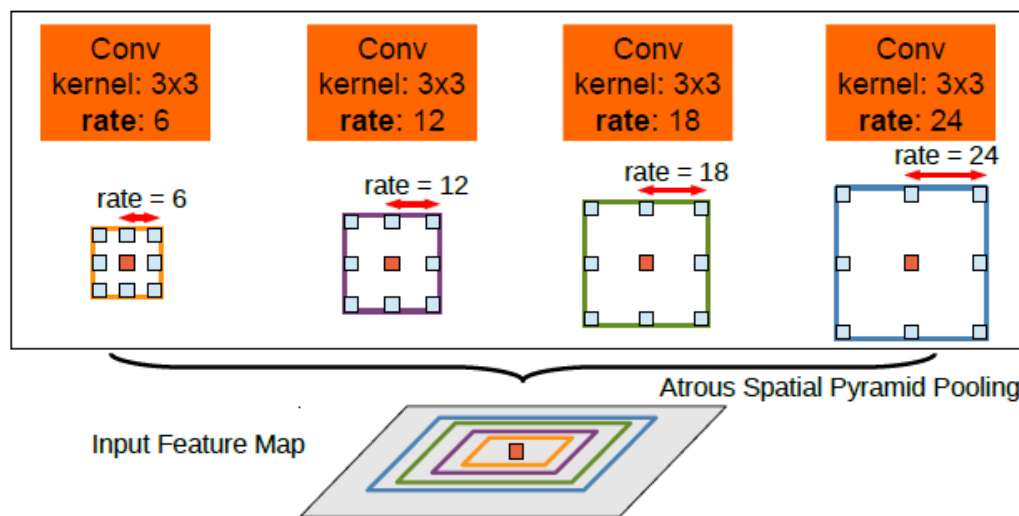
- 膨胀卷积效果
 - 稀疏特征提取：x2降采样 \rightarrow 7x7卷积 \rightarrow x2上采样
 - 稠密特征提取：7x7膨胀卷积
- 优势
 - 参数&计算量一样
 - 灵活控制分辨率



语义分割（Semantic Segmentation）

DeepLab-DCNN

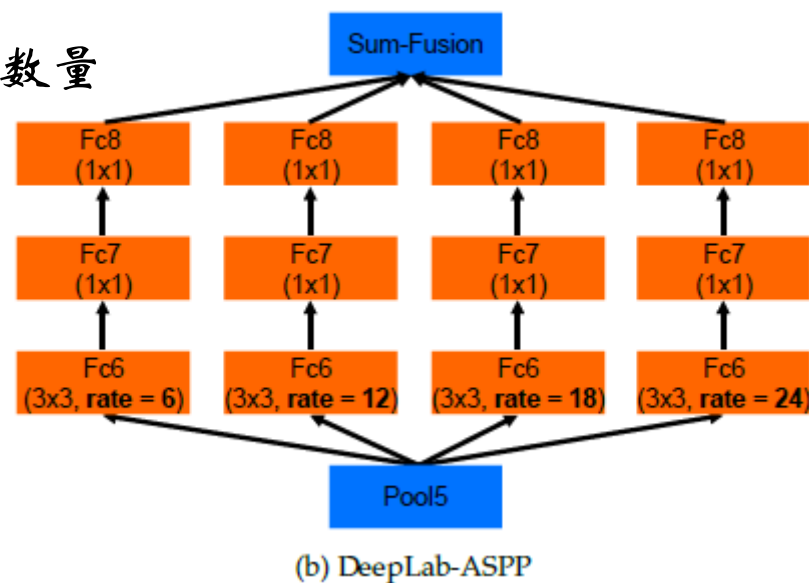
- **Atrous 空间金字塔池化**（Atrous Spatial Pyramid Pooling）
 - 不同感受野（rate）捕捉不同尺度上的特征
 - 在Conv6层引入4个并行膨胀卷积
 - Rate: 6, 12, 18, 24



语义分割（Semantic Segmentation）

DeepLab-DCNN

- Atrous 空间金字塔池化（Atrous Spatial Pyramid Pooling）
 - 4个并行膨胀卷积
 - 感受野：13x13, 25x25, 37x37, 49x49
 - $Fc6 \rightarrow Fc7 \rightarrow Fc8$
 - 深度：4096 \rightarrow 2014 \rightarrow 类别数量
 - 卷积核：3x3 \rightarrow 1x1 \rightarrow 1x1
 - 融合：概率相加



语义分割（Semantic Segmentation）

DeepLab-全连接CRF

- 作用：通过迭代精化分割结果（恢复精确边界）
- 输入
 - 首次：FCN网络输出结果的8倍双线性插值
 - 非首次：上一轮迭代结果
- 能量计算基于图片RGB像素值

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

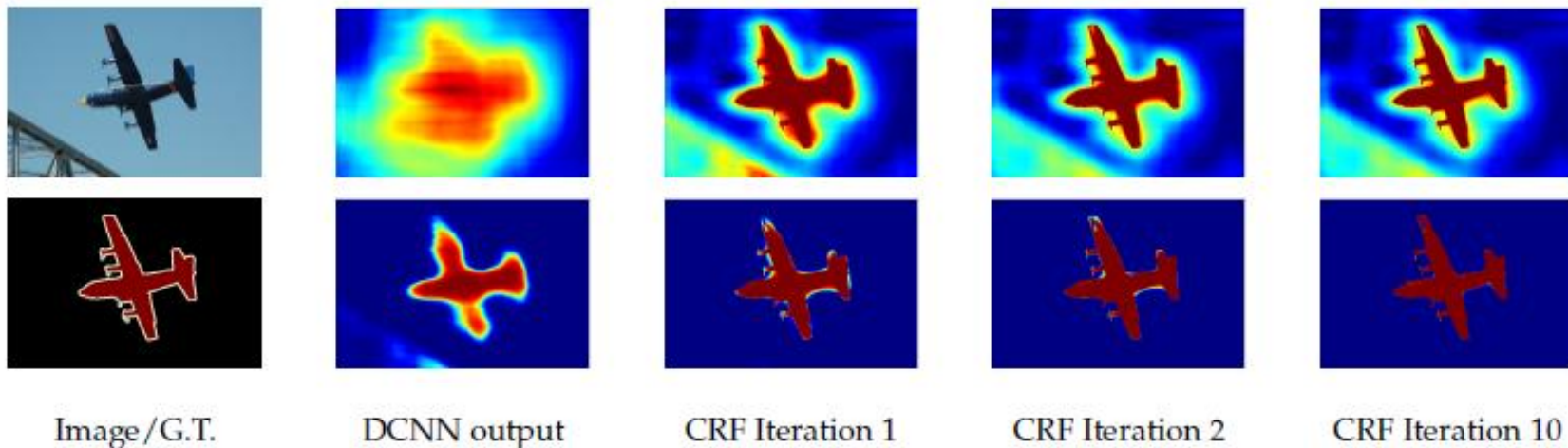
$$\theta_i(x_i) = -\log P(x_i)$$

$$\theta_{ij}(x_i, x_j) = \mu(x_i, x_j) \left[w_1 \exp \left(-\frac{\|p_i - p_j\|^2}{2\sigma_\alpha^2} - \frac{\|I_i - I_j\|^2}{2\sigma_\beta^2} \right) + w_2 \exp \left(-\frac{\|p_i - p_j\|^2}{2\sigma_\gamma^2} \right) \right] \quad (3)$$

语义分割（Semantic Segmentation）

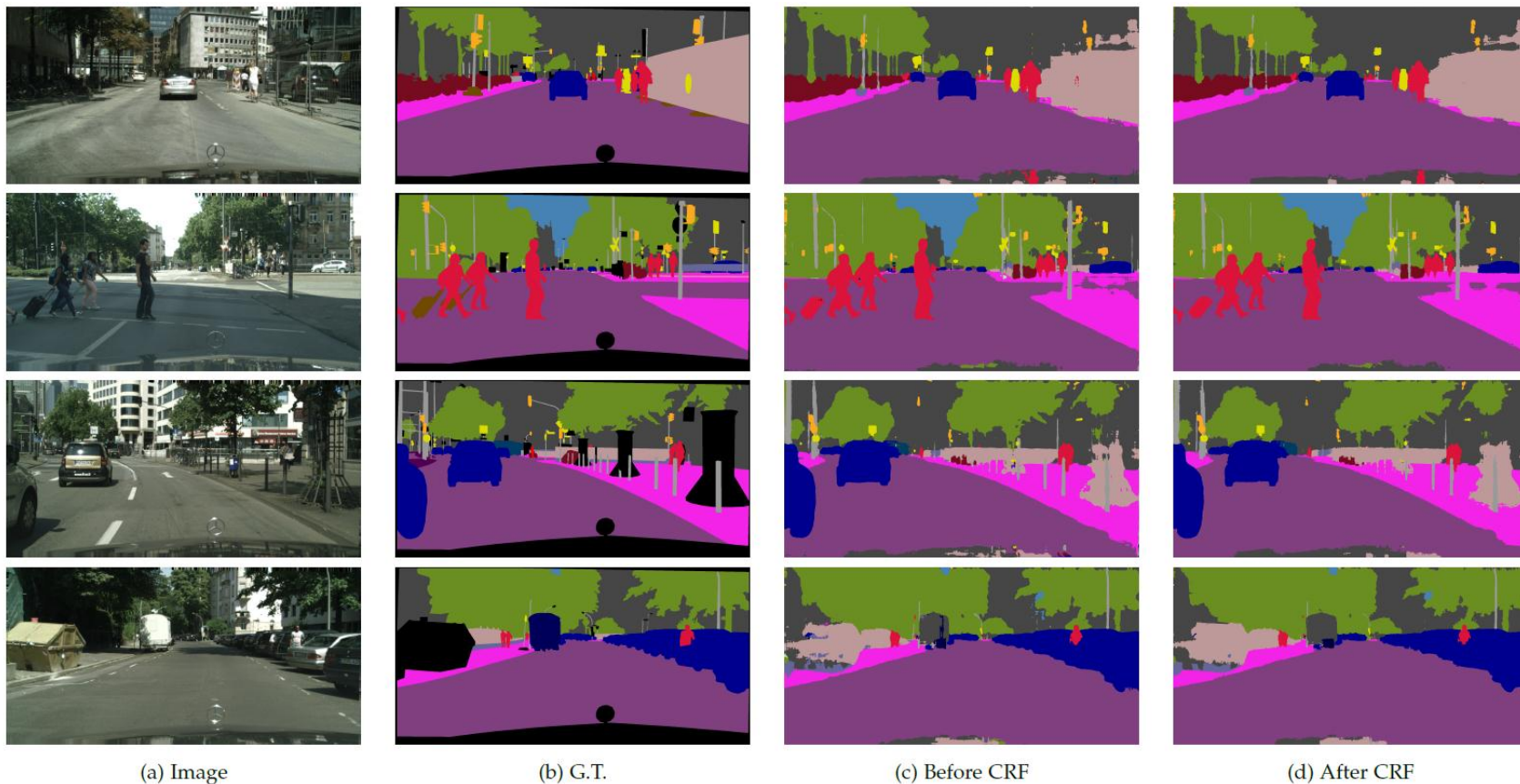
DeepLab-全连接CRF

- 第一行：飞机类别的分值（softmax之前）
- 第二行：飞机类别的概率值（softmax之后）



语义分割 (Semantic Segmentation)

Cityscapes数据集分割效果



语义分割（Semantic Segmentation）

Cityscapes数据集性能

- ResNet-101 优于 VGG16

Full	Aug	LargeFOV	ASPP	CRF	mIOU
<i>VGG-16</i>					
		✓			62.97
		✓		✓	64.18
✓		✓			64.89
✓		✓		✓	65.94
<i>ResNet-101</i>					
✓					66.6
✓		✓			69.2
✓			✓		70.4
✓	✓		✓		71.0
✓	✓		✓	✓	71.4

语义分割数据集

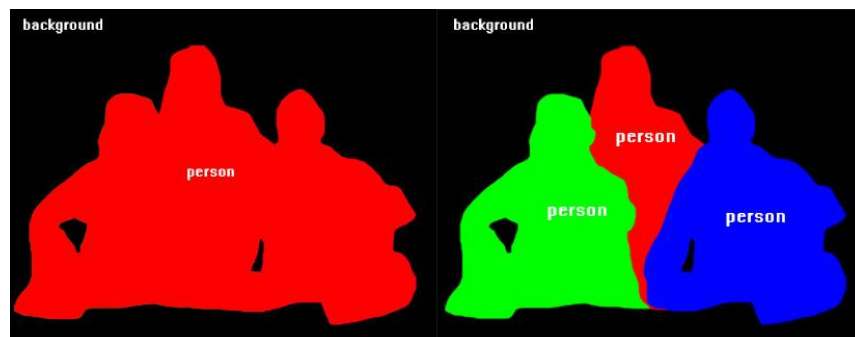
Pascal VOC - 2012

- 20个物体类别
 - 人类
 - 动物（鸟、猫、牛、狗、马、羊）
 - 交通工具（飞机、自行车、船、公共汽车、小轿车、摩托车、火车）
 - 室内（瓶子、椅子、餐桌、盆栽植物、沙发、电视）
- 像素级标签9,993张图片

语义分割数据集

MSCOCO

- 80个类别
- COCO-stuff扩展集：172类别
 - Object: 80
 - Stuff: 91
 - Unknown: 1
- 主要用于：
 - 实例级别的分割 (Instance-level)
 - 图片描述 (Image Captioning)
- <http://mscoco.org/>



语义分割数据集

Cityscapes

- 30个类别
- 标注：
 - 5,000张像素标注 (pixel level)
 - 20,000张多边形标注 (instance level)
- 辅助/自动驾驶中的语义场景理解
- 采集于50个城市
- <https://www.cityscapes-dataset.com>

演示环节

- Github
 - <https://github.com/349zzjau>
- 百度网盘
 - <http://pan.baidu.com/s/1gfpCCwj>
- 代码演示
 - DeepLab

疑问

□ 问题答疑：<http://www.xxwenda.com/>

■ 可邀请老师或者其他人回答问题

Q & A

小象账号：349zzjau

课程名：基于深度学习的计算机视觉

课后调查问卷<http://cn.mikecrm.com/XGIQYBp>

联系我们

小象学院：互联网新技术在线教育领航者

- 微信公众号：小象
- 新浪微博：ChinaHadoop

