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

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

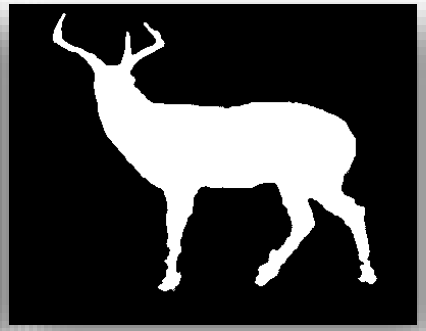
本章结构

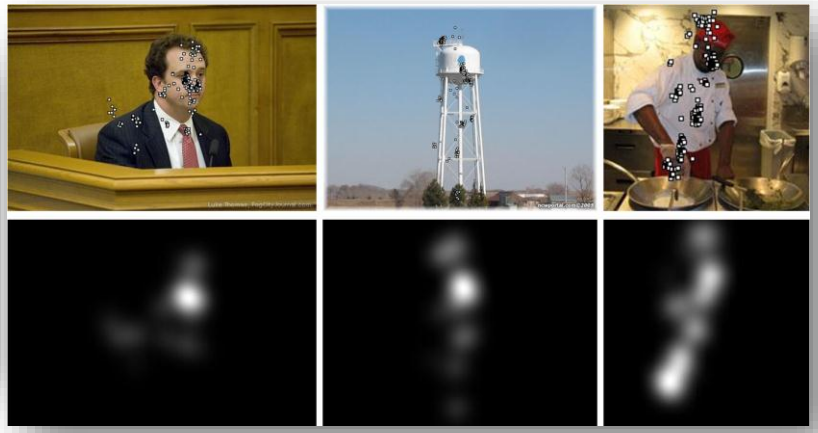
- 显著性检测（Saliency Detection）
- 物体分割（Object Segmentation）
- 语义分割（Semantic Segmentation）
- 三大数据集介绍（Pascal VOC，MSCOCO，Cityscapes）
- 应用案例：
  - 自动驾驶场景图片的语义分割-全卷积网络DeepLab

显著性检测（Saliency Detection）

2类问题

- 显著性物体分割（Salient object segmentation）
  - 最能引起人的视觉注意的物体区域
- 注视点预测 (Fixation prediction)
  - 通过对眼动的预测和研究探索人类视觉注意机制

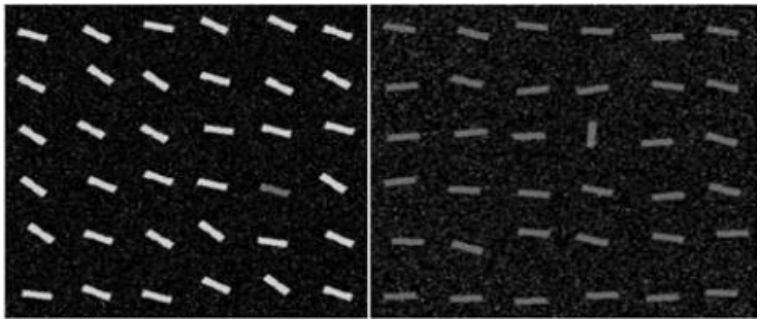




显著性检测（Saliency Detection）

两种策略的视觉注意机制

- 自底而上基于数据驱动的注意机制
  - 从数据出发
  - 与周边有较强对比度或差异
  - 颜色、亮度、边缘等特征
- 自上而下基于任务驱动的目标的注意机制
  - 从认知因素出发，如知识、预期、兴趣等

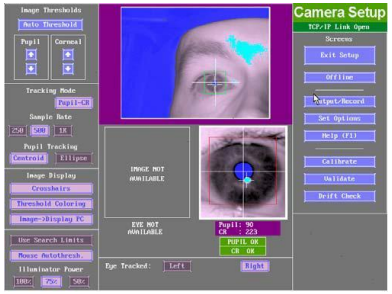


显著性检测（Saliency Detection）

Pascal VOC数据集

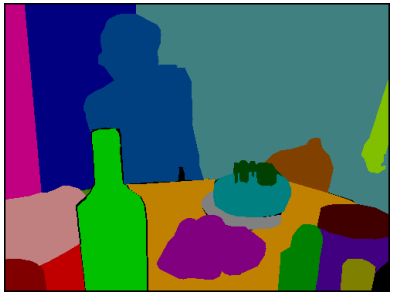
- 显著物体标注
- 眼动数据

原始图像（PASCAL VOC）



眼动追踪实验

眼动数据



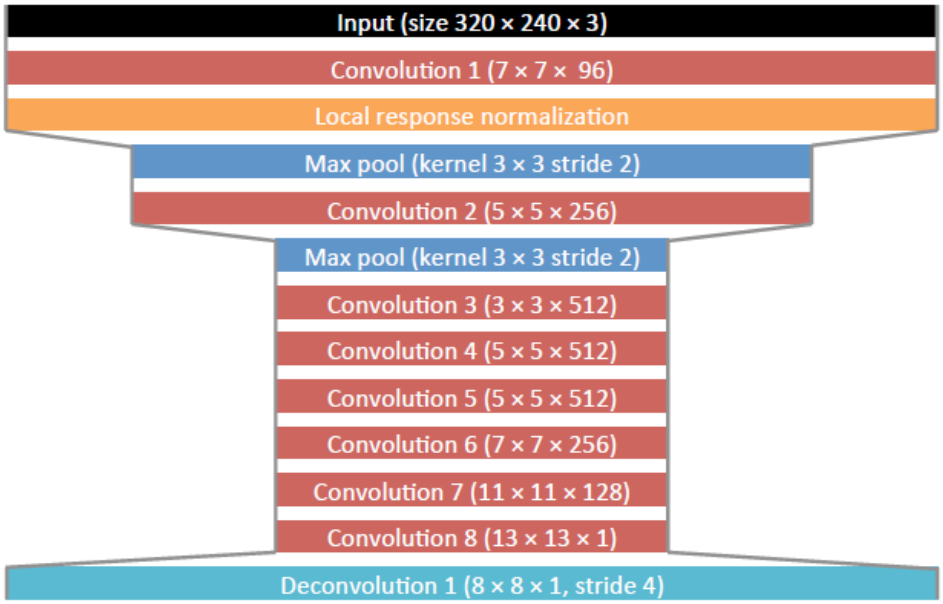
人工图像标注

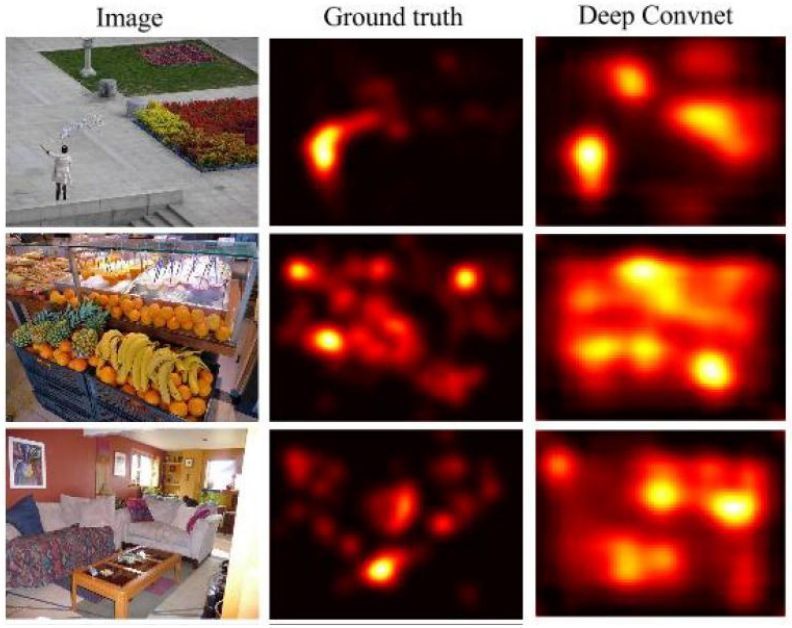
显著物体轮廓

显著性检测（Saliency Detection）

DNN模型

- 由VGG网络修改而成

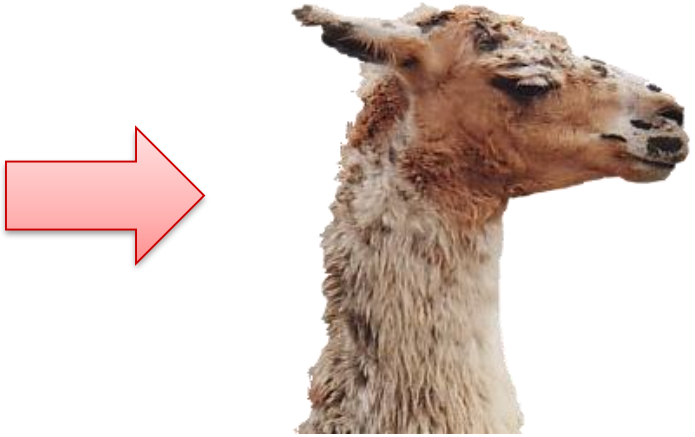




物体分割（Object Segmentation）

前景背景分割

- 前景一般包含物体
- 需要交互提供初始标记

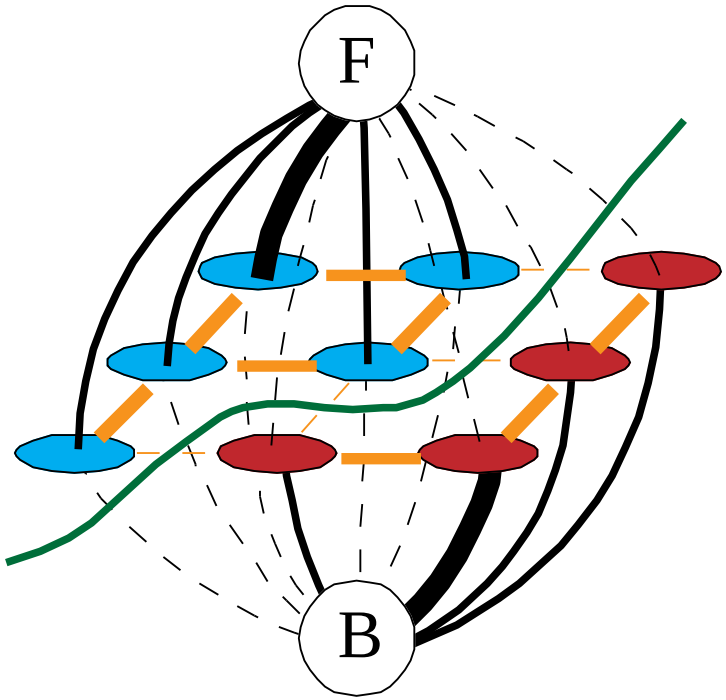


物体分割（Object Segmentation）

F	F	B
F	F	B
F	B	B

Graph Cuts分割

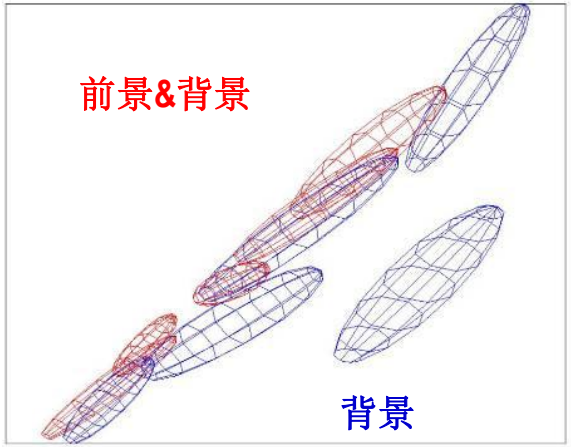
- 基于图论的分割方法
- 分割模型
  - 每个像素是一个节点
  - 加2个节点F/B
  - 边
    - 像素跟F/B的连接
    - 相邻像素的连接
  - 最小割最大流算法优化



物体分割（Object Segmentation）

GrabCut分割

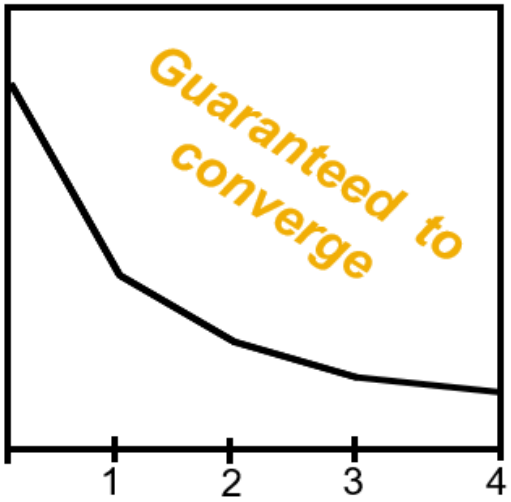
- 前景/背景的颜色模型
  - 高斯混合模型
  - Kmeans算法获得



物体分割（Object Segmentation）

GrabCut分割

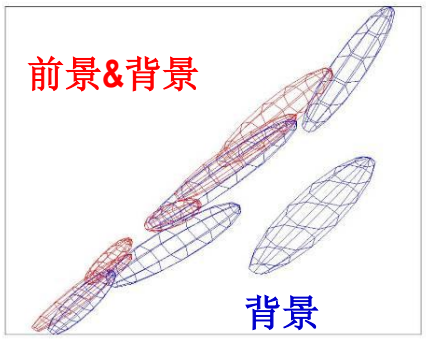
- 迭代进行Graph Cuts
  - 优化前景和背景的颜色模型
  - 能量随着不断迭代变小
  - 分割结果越来越好

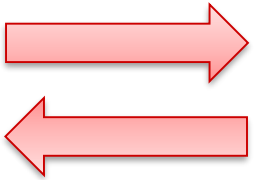


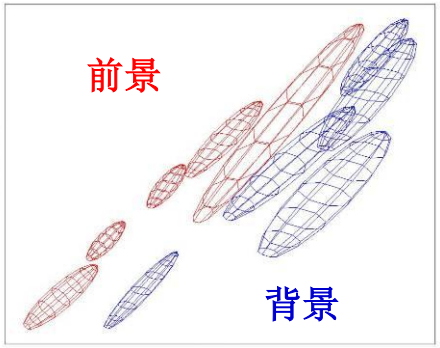
物体分割（Object Segmentation）

GrabCut分割

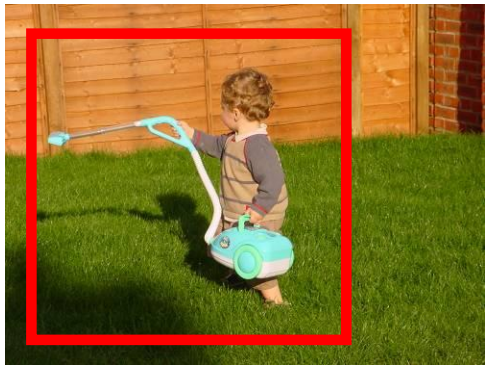
- 算法流程
  - 使用标记初始化颜色模型（K=5）
  - 执行Graph Cuts







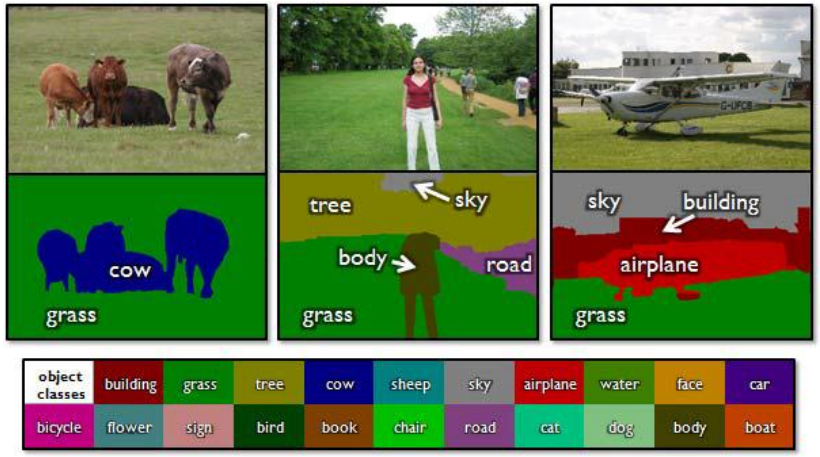
物体分割（Object Segmentation）



语义分割（Semantic Segmentation）

什么是语义分割

- 目标
  - 从像素水平（pixel-level）上，理解、识别图片的内容
  - 根据语义信息分割
- 输入
  - 图片
- 输出
  - 同尺寸的分割标记（像素水平）
  - 每个像素会被识别为一个类别（category）

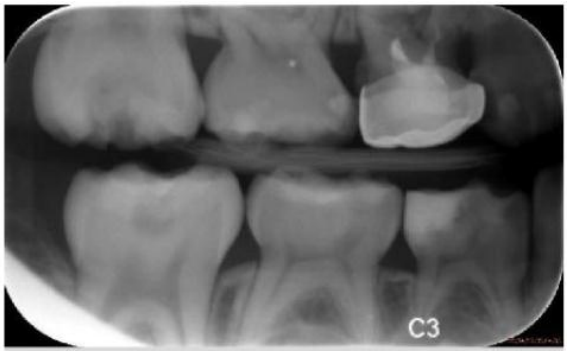


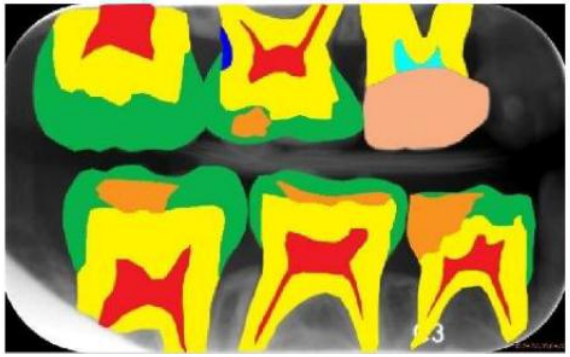
语义分割（Semantic Segmentation）

语义分割的用处

- 机器人视觉和场景理解
- 辅助/自动驾驶
- 医学X光

辅助/自





语义分割（Semantic Segmentation）

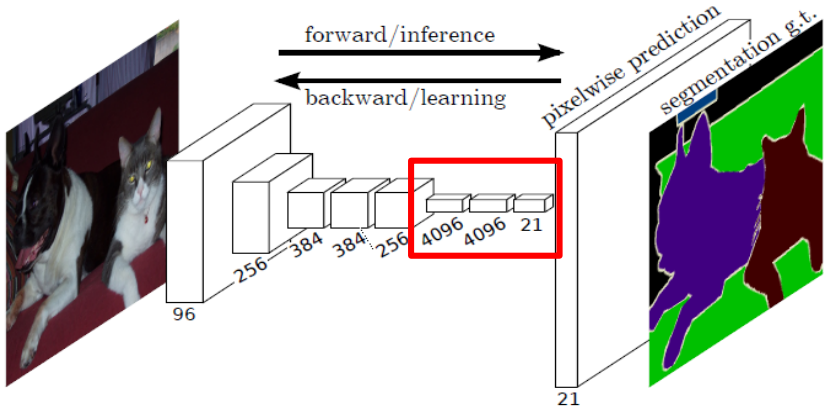
算法研究阶段

- 2015之前：手工特征+图模型（CRF）
- 2015开始：深度神经网模型
  - 思路：改进CNN，并使用预训练CNN层的参数
  - 传统CNN的问题
    - 后半段网络无空间信息
    - 输入图片尺寸固定
  - 全卷积网络（Fully Convolutional Networks）
    - 所有层都是卷积层
    - 解决降采样后的低分辨率问题

语义分割（Semantic Segmentation）

全卷积网络（Fully Convolutional Networks-FCN）

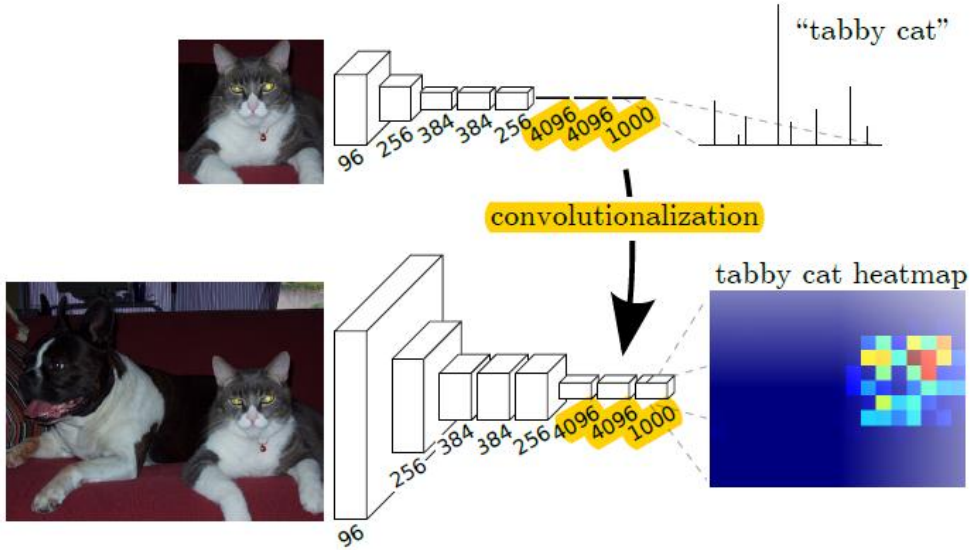
- 卷积化（Convolutionalization）
  - 将所有全连接层转换成卷积层
  - 适应任意尺寸输入，输出低分辨率分割图片
- 反卷积（Deconvolution）
  - 将低分辨率图片进行上采样，输出同分辨率分割图片
- 跳层结构（Skip-layer）
  - 精化分割图片



语义分割（Semantic Segmentation）

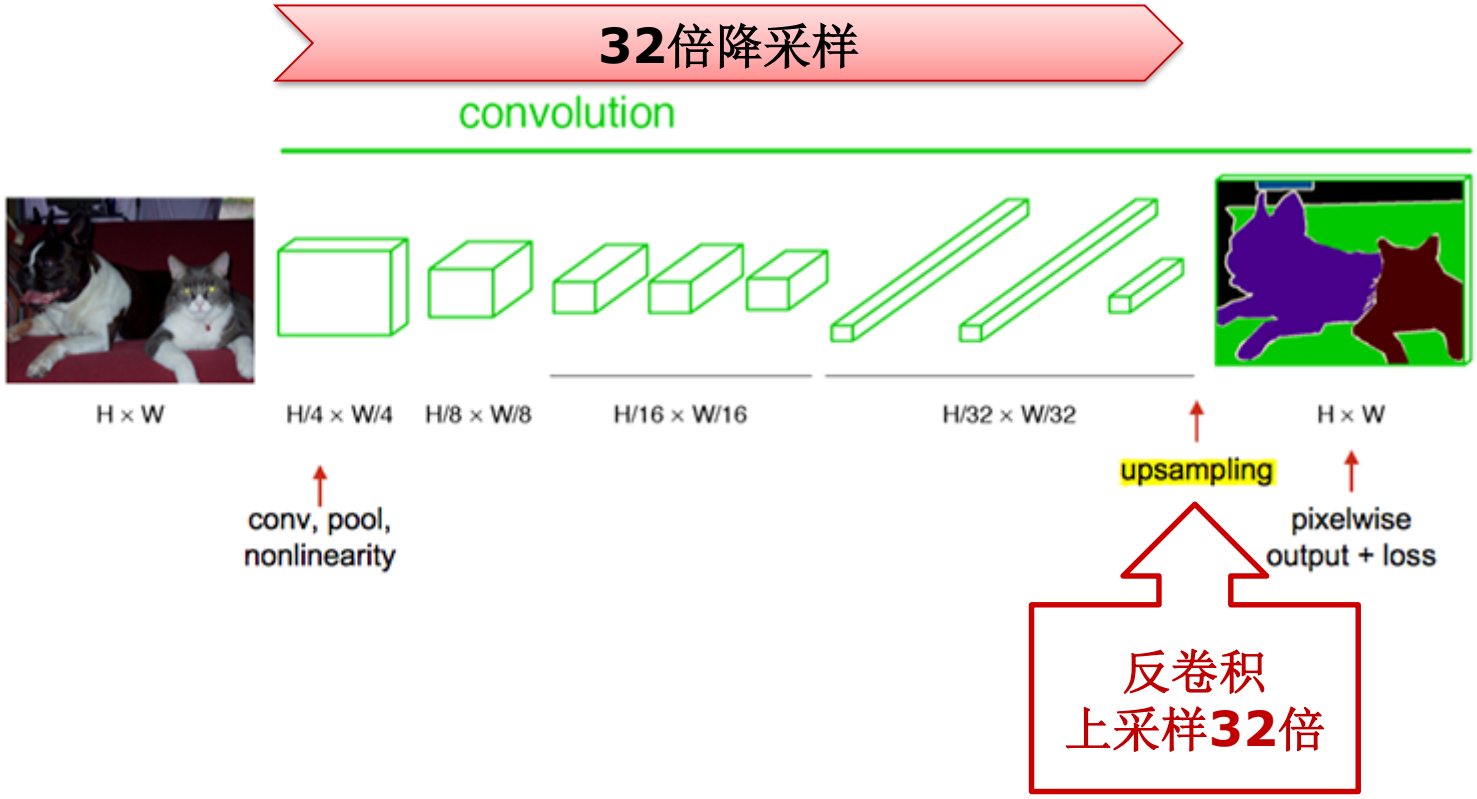
FCN-卷积化（Convolutionalization）

- 基础CNN网络: AlexNet, VGG16, GoogLeNet
- 卷积化后的核尺寸（通道数，宽，高）
  - （4096，1，1）
  - （4096，1，1）
  - （1000/21，1，1）
- 分辨率降低32倍
  - 5个卷积层
  - 每层降2倍



语义分割（Semantic Segmentation）

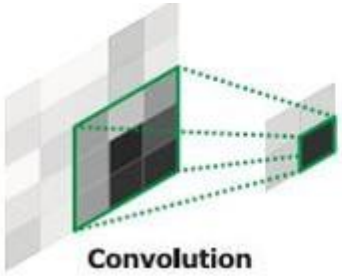
FCN-卷积化的降维问题

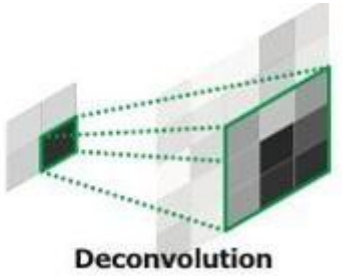


语义分割（Semantic Segmentation）

FCN-反卷积（Deconvolution）

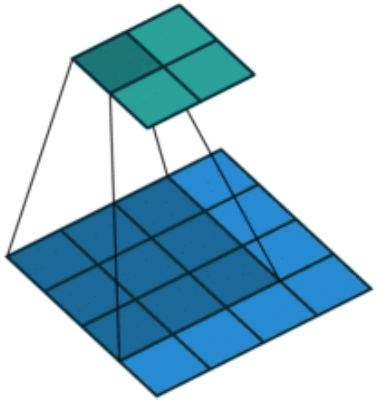
- 一对多操作
- 卷积的逆操作
  - 小数步长1/f
  - 卷积核尺寸不变
- 前向和后向传播
  - 对应于卷积操作的后向和前向传播，优化上做颠倒
  - 反卷积核是卷积核的转置，学习率为0
- 也叫转置卷积（Transposed convolution）
- 可以拟合出双线性插值



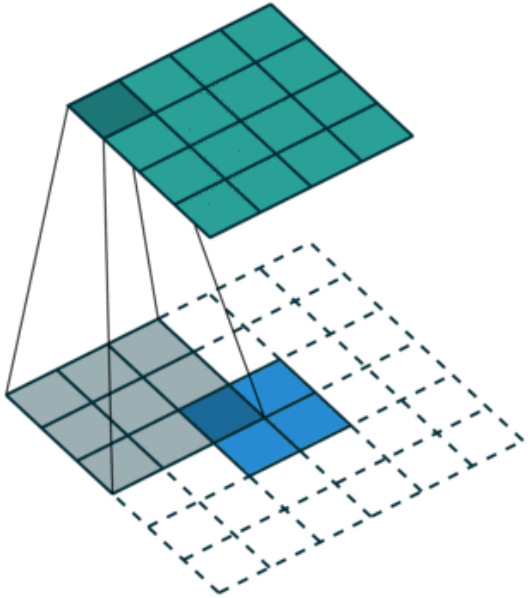


语义分割（Semantic Segmentation）

FCN-反卷积（Deconvolution）



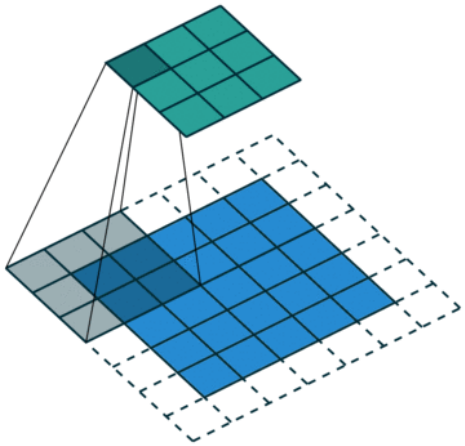
- 外围全补零（Full padding）反卷积
- 输入：2x2
- 输出：4x4
- 参数设置
  - 卷积核尺寸：3x3
  - 步长：1
  - Padding：2
- 被Skip-layer使用

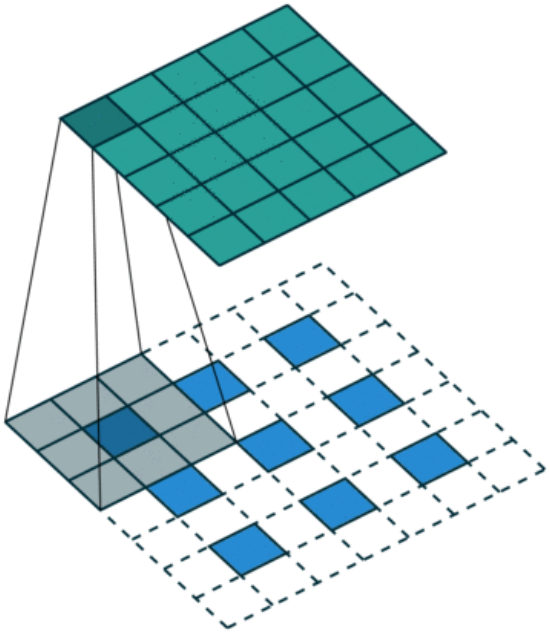


语义分割（Semantic Segmentation）

FCN-反卷积（Deconvolution）

- 插零分数步长反卷积
- 输入：3x3
- 输出：5x5
- 参数设置
  - 卷积核尺寸：3x3
  - 步长：2
  - Padding：1

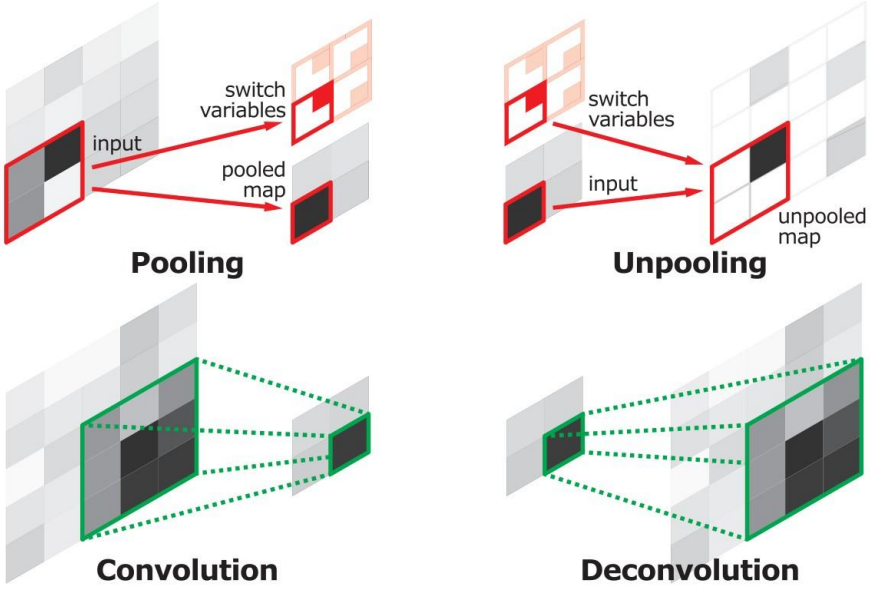




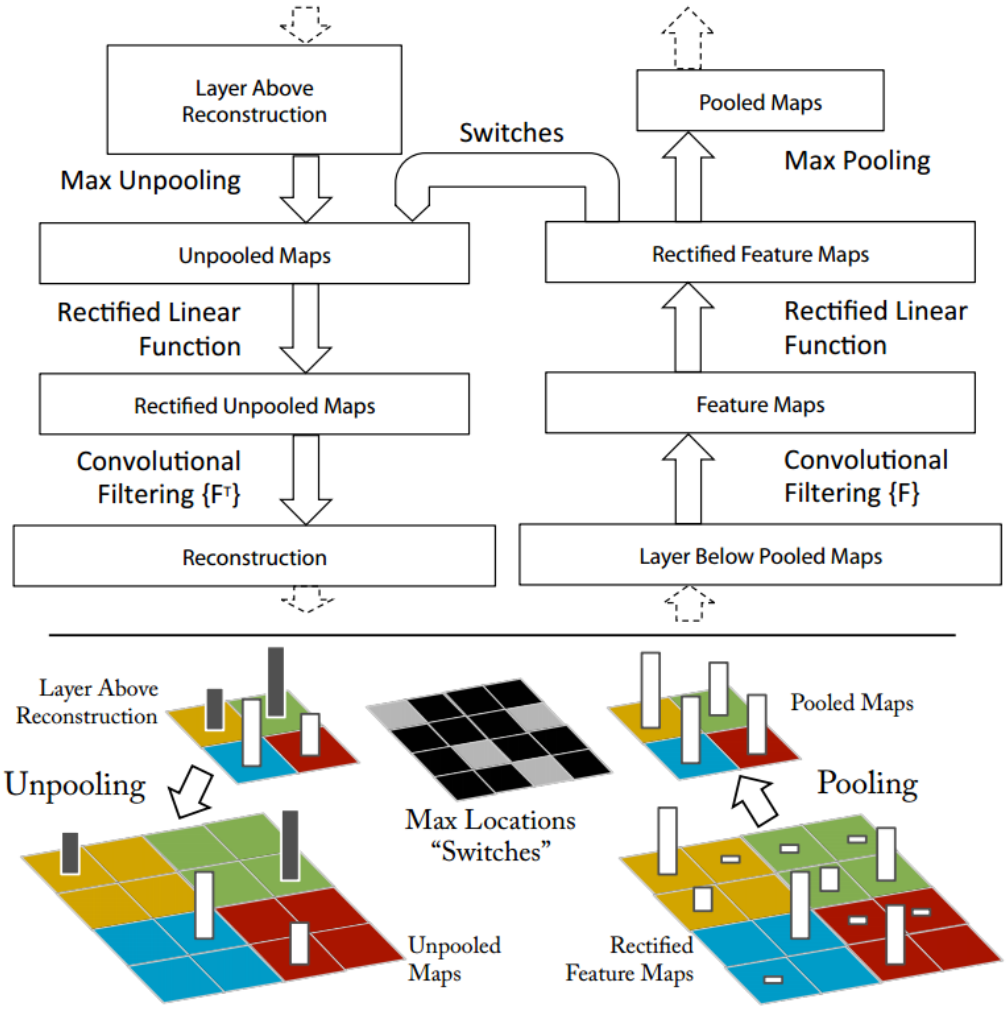
语义分割（Semantic Segmentation）

FCN-反卷积（Deconvolution）

- 反池化操作（Unpooling）
  - 记录池化时的位置
  - 将输入特征按记录位置摆放回去（近似）



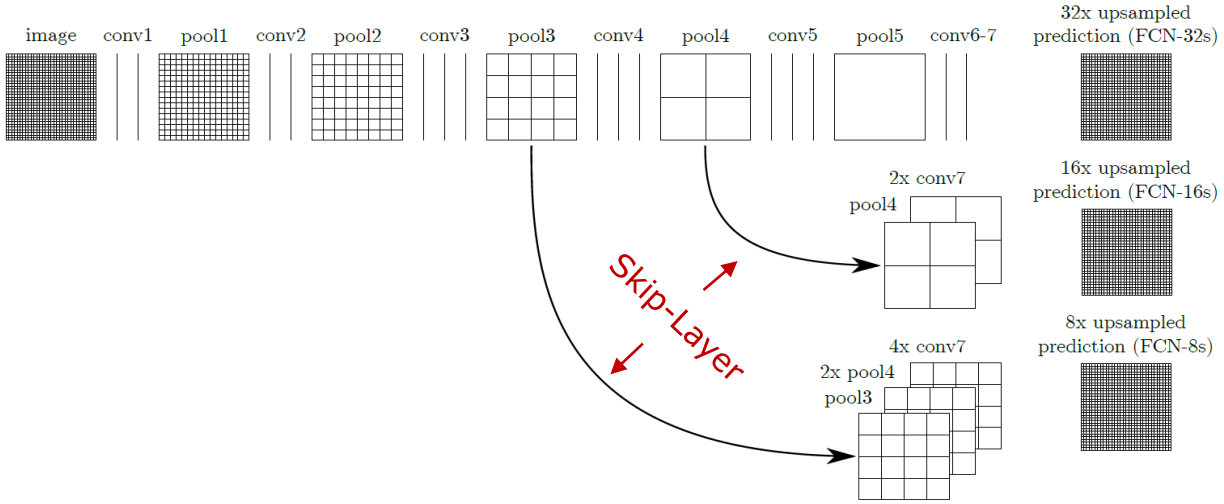
语义分割（Semantic Segmentation）



语义分割（Semantic Segmentation）

FCN-跳层结构（Skip-layer）

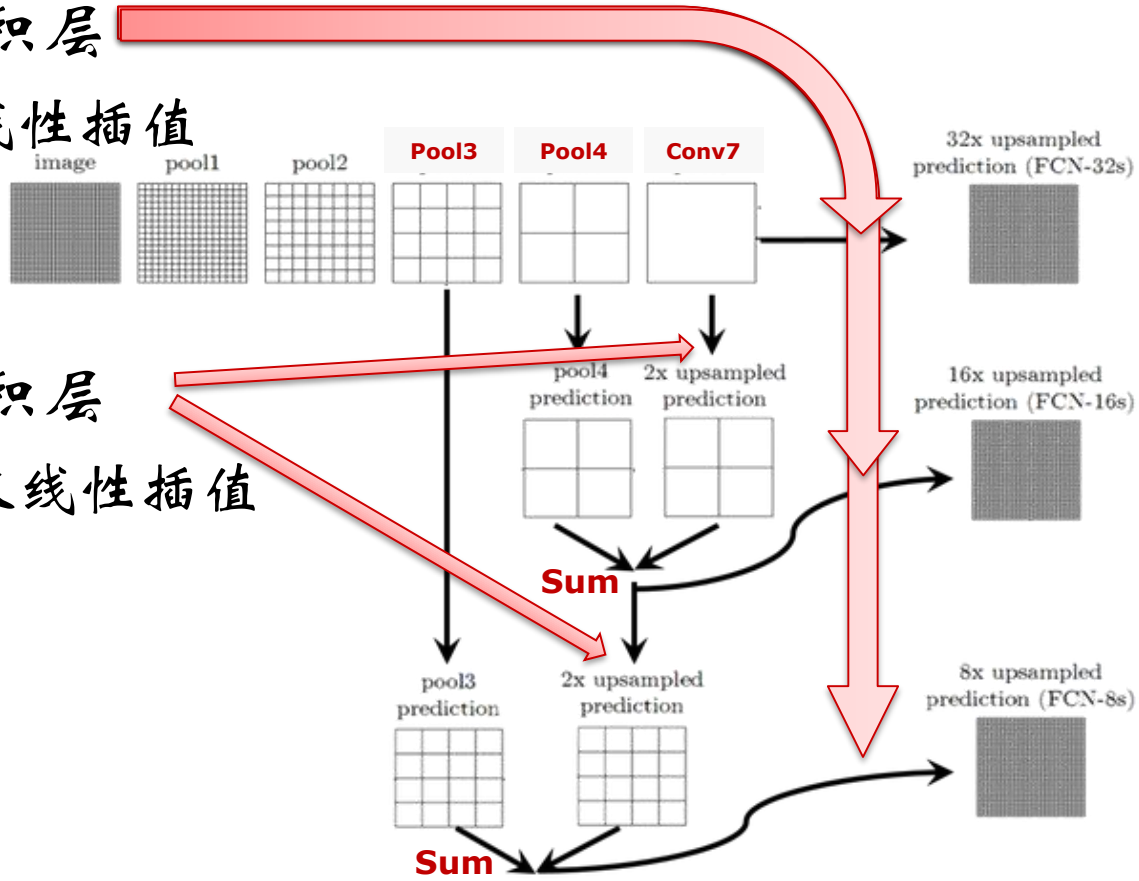
- 原因：直接使用32倍反卷积得到的分割结果粗糙
- 使用前2个卷积层的输出做融合
  - 跳层：Pool4和Pool3后会增加一个1x1卷积层做预测
  - 较浅网络的结果精细，较深网络的结果鲁棒



语义分割（Semantic Segmentation）

FCN-跳层结构（Skip-layer）

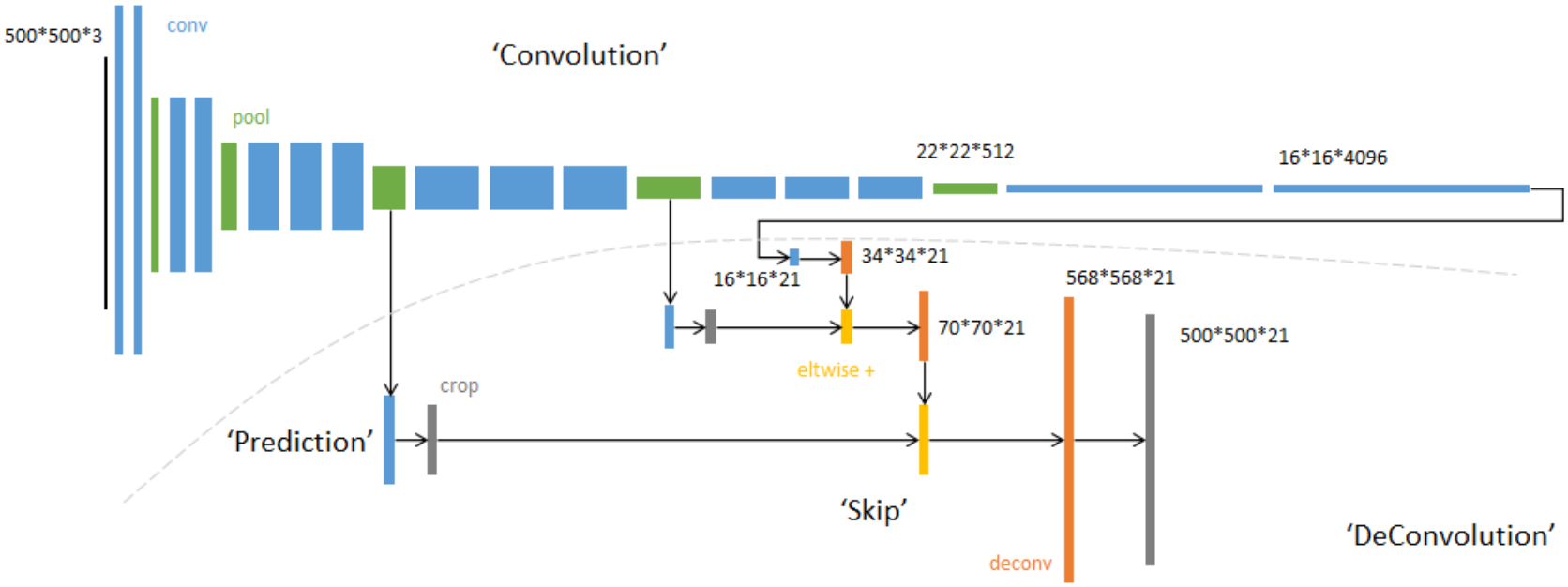
- 最后的反卷积层
  - 固定为双线性插值
  - 不学习
- 中间的反卷积层
  - 初始化为双线性插值
  - 需要学习



语义分割（Semantic Segmentation）

FCN构架图例

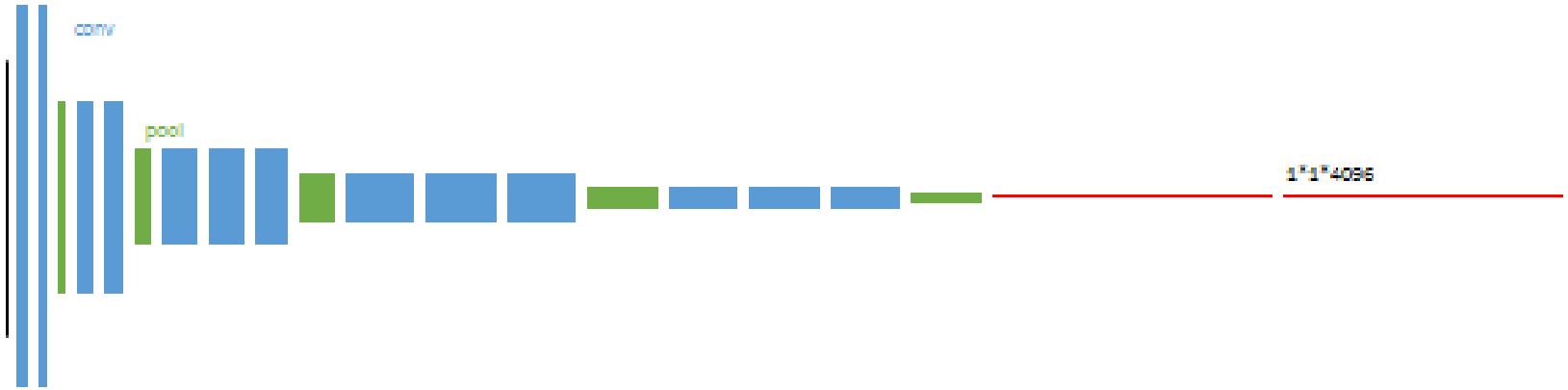
蓝色	绿色	黄色	橙色	灰色
卷积层	池化层	逐像素相加	反卷积层	剪裁层
	获取特征	融合结果	上采样	统一尺寸



语义分割（Semantic Segmentation）

使用AlexNet构建FCN

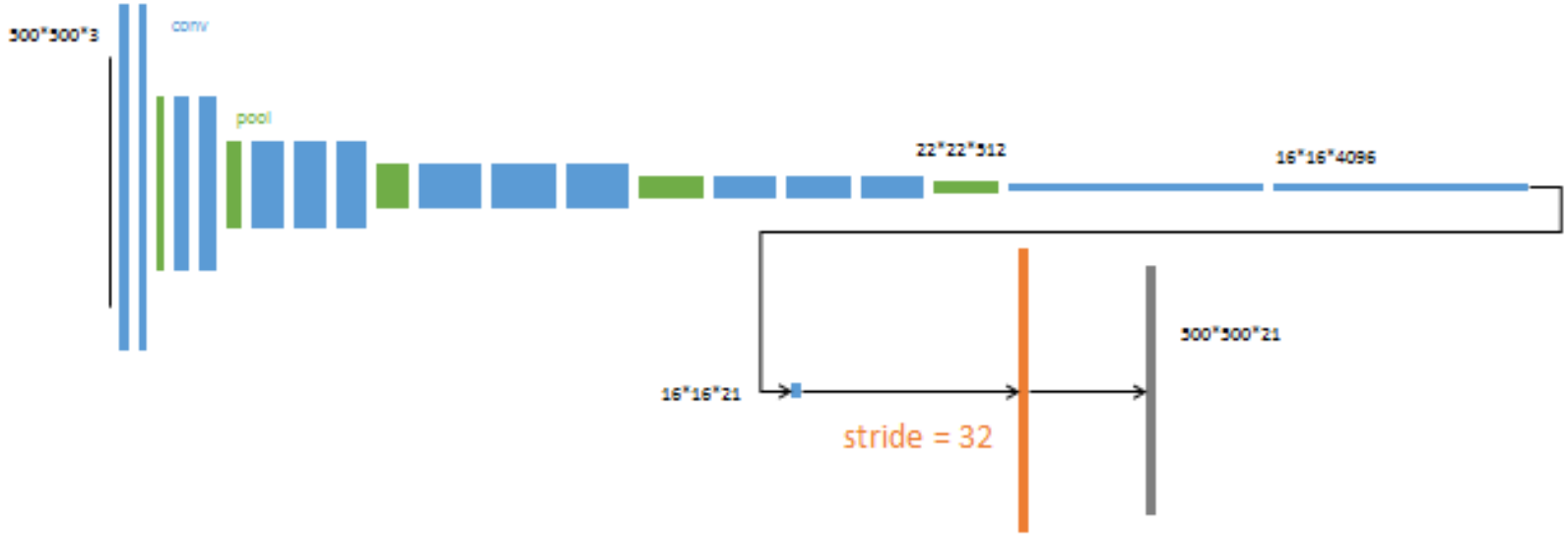
- 第1步
  - 使用AlexNet作为初始网络，保留参数
  - 舍弃2个全连接层



语义分割（Semantic Segmentation）

使用AlexNet构建FCN

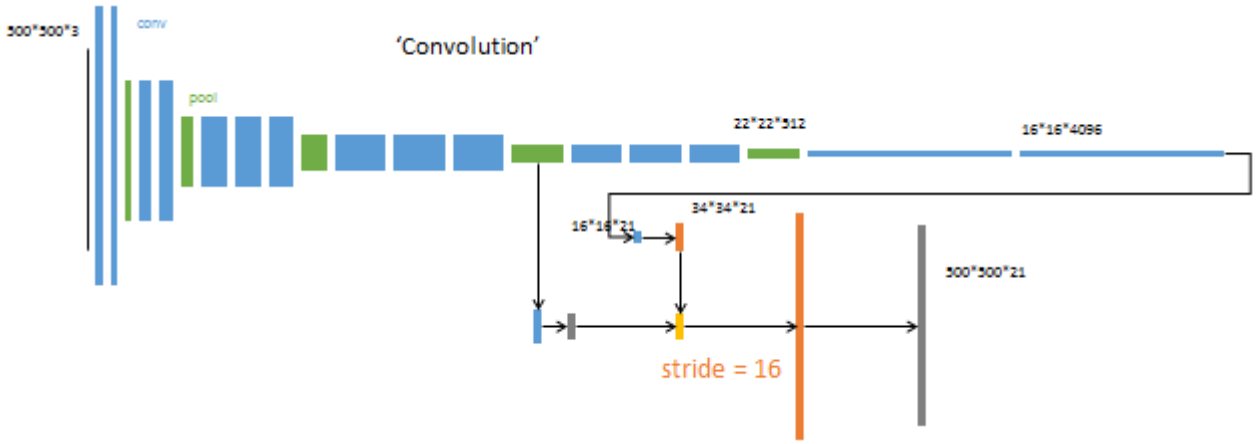
- 第2步（FCN-32s网络）
  - 替换为两个同深度的卷积层（4096,1,1）→16x16x4096
  - 追加一个预测卷积层（21,1,1）→16x16x21
  - 追加一个步长为32的双线性插值反卷积层→500x500x21



语义分割（Semantic Segmentation）

使用AlexNet构建FCN

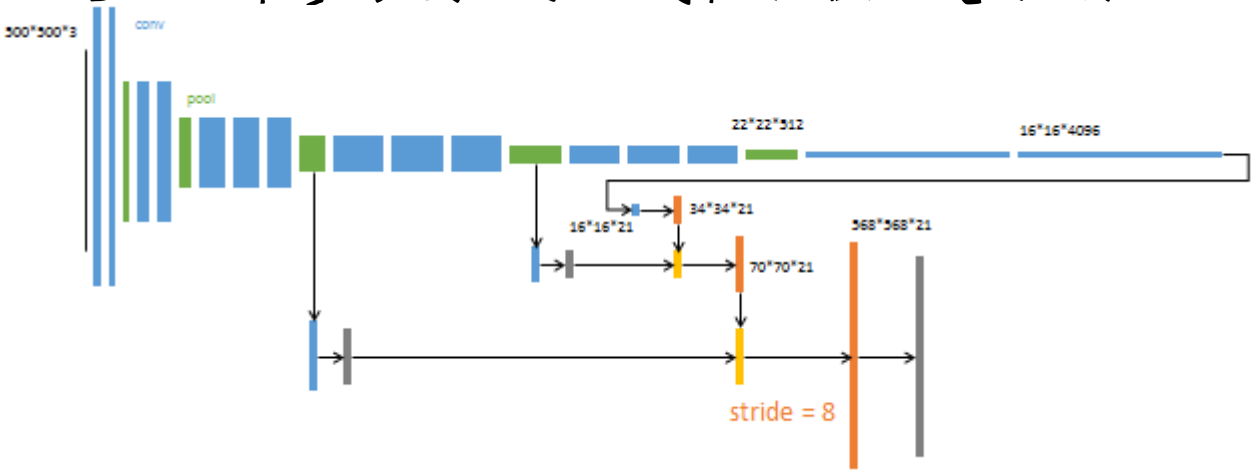
- 第3步（FCN-16s网络）
  - 对最终层Conv7结果2倍上采样→34x34x21
  - 提取Pool4输出，追加预测卷积层（21,1,1）→34x34x21
  - 相加融合→34x34x21
  - 追加一个步长为16的双线性插值反卷积层→500x500x21



语义分割（Semantic Segmentation）

使用AlexNet构建FCN

- 第3步（FCN-8s网络）
  - 对上次融合结果2倍上采样→70x70x21
  - 提取Pool3输出，追加预测卷积层（21,1,1）→70x70x21
  - 相加融合→70x70x21
  - 追加一个步长为8的双线性插值反卷积层→500x500x21



语义分割（Semantic Segmentation）

FCN训练

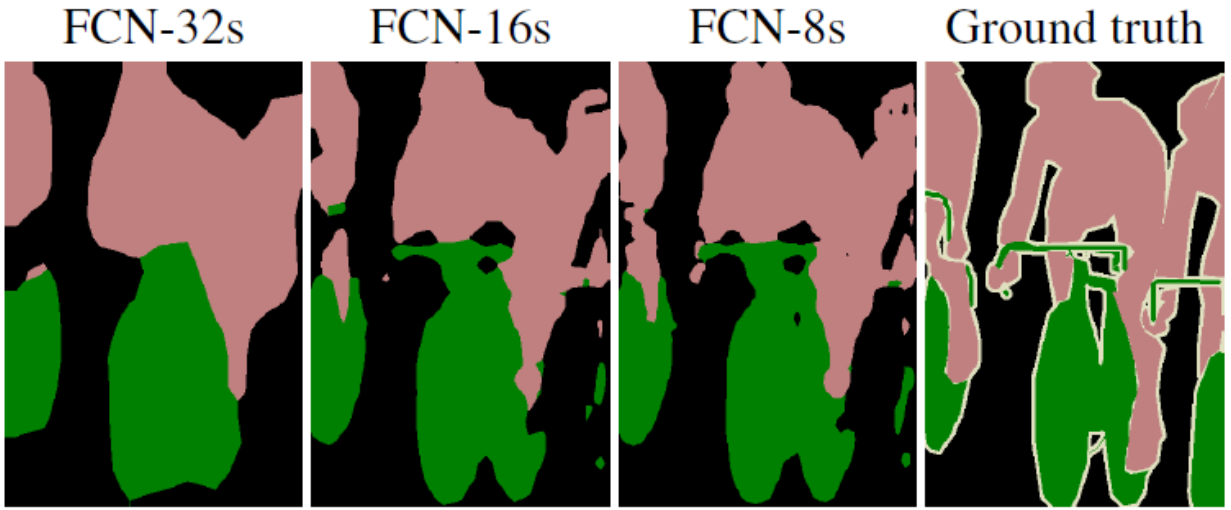
- SGD with momentum (0.9)
  - Learning rate
    - 0.001(AlexNet)，0.0001(VGG16)，0.00001(GoogLeNet)
  - Minibatch：20
- 初始化
  - 卷积层
    - 前5个卷积层使用初始CNN网络的参数
    - 剩余第6和7卷积层初始化为0
  - 反卷积层
    - 最后一层反卷积层固定为双线性插值不做学习
    - 剩余反卷积层初始化为双线性插值

语义分割（Semantic Segmentation）

FCN的跳层结构性能

- FCN-8s最优

	pixel acc.	mean acc.	mean IU	f.w. IU
FCN-32s-fixed	83.0	59.7	45.4	72.0
FCN-32s	89.1	73.3	59.4	81.4
FCN-16s	90.0	75.7	62.4	83.0
FCN-8s	**90.3**	**75.9**	**62.7**	**83.2**



语义分割（Semantic Segmentation）

FCN的基础网络性能

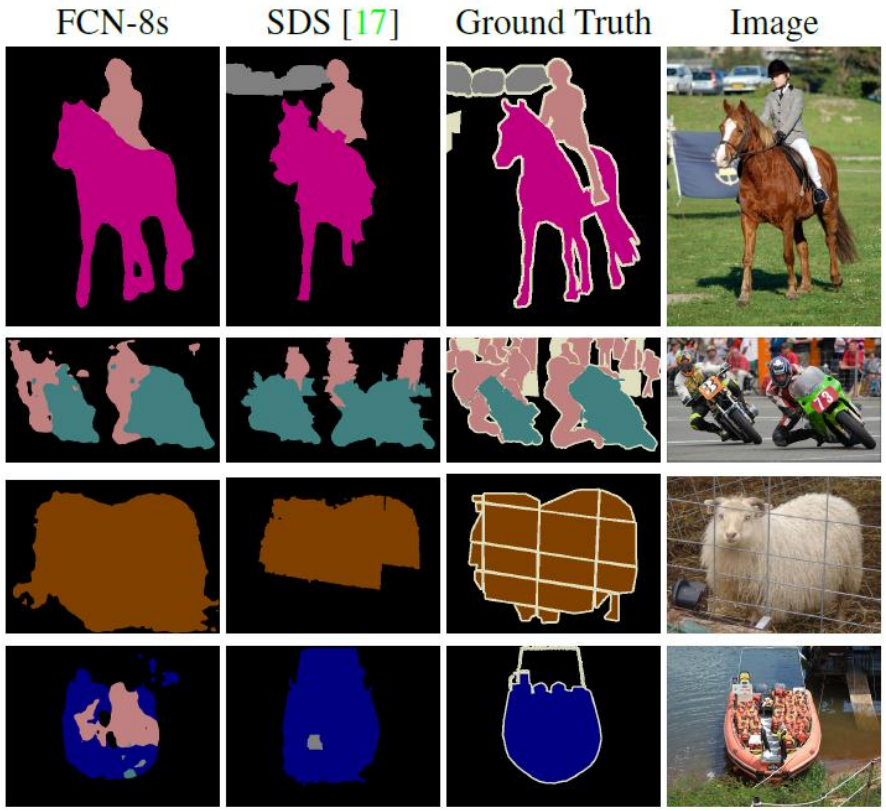
- FCN-VGG16最优

	FCN-AlexNet	FCN-VGG16	FCN-GoogLeNet[4]
mean IU	39.8	**56.0**	42.5
forward time	50 ms	210 ms	59 ms
conv. layers	8	16	22
parameters	57M	134M	6M
rf size	355	404	907
max stride	32	32	32

语义分割（Semantic Segmentation）

FCN-8s的Pascal VOC竞赛结果

- 边缘准确性比较差
  - 第1个卷积层大量补零
  - 之后做裁剪
  - 保证输出分辨率
  - 带来噪声



	mean IU VOC2011 test	mean IU VOC2012 test	inference time
R-CNN [12]	47.9	-	-
SDS [17]	52.6	51.6	~ 50 s
FCN-8s	**62.7**	**62.2**	~ **175 ms**

语义分割（Semantic Segmentation）

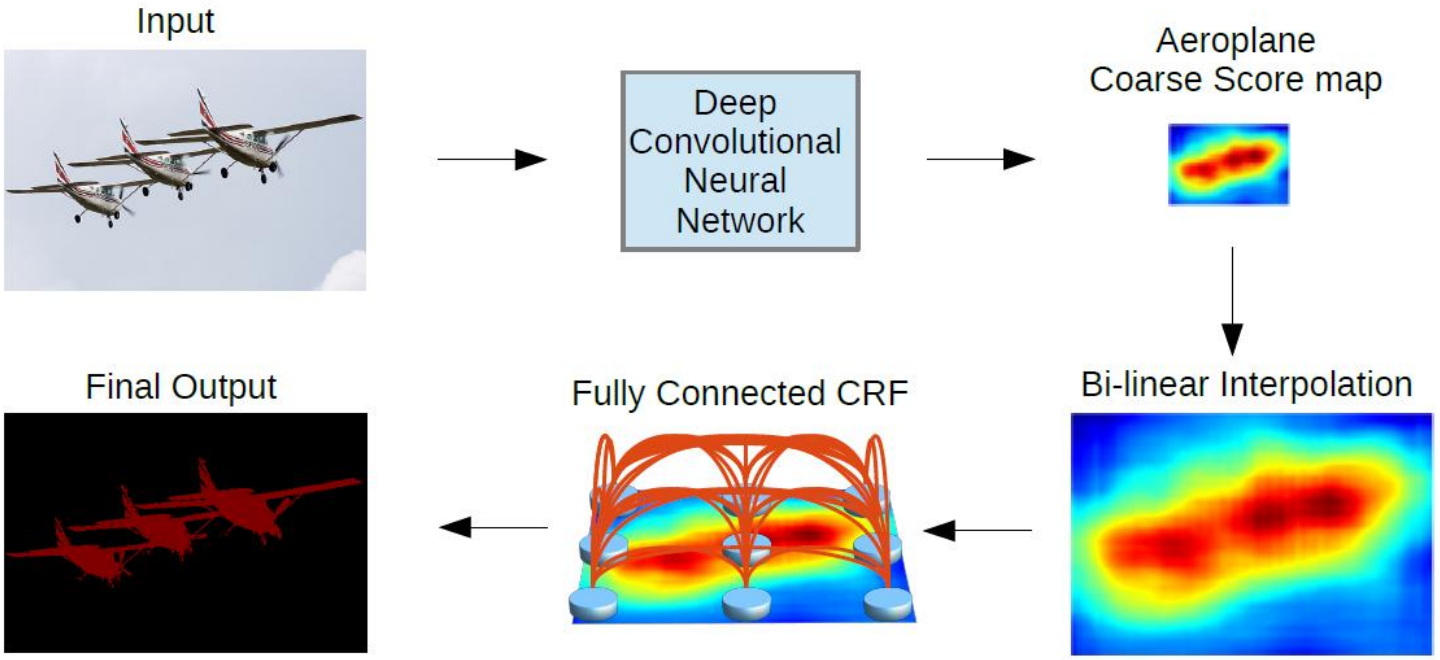
DeepLab全卷积网络

- 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）

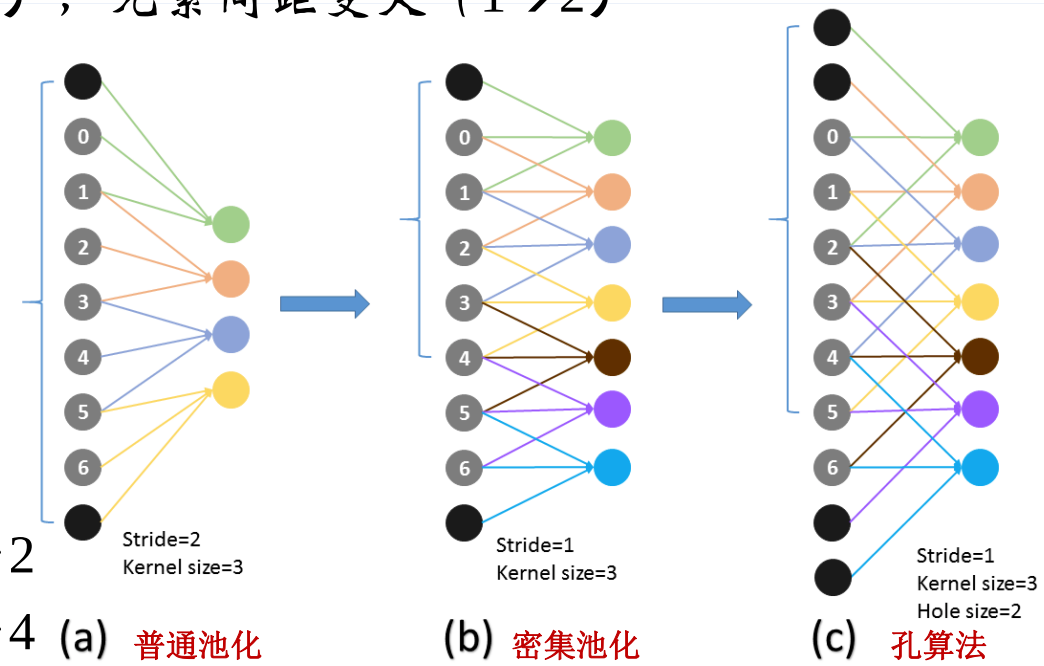
DeepLab-DCNN

- 孔（Hole）算法
  - 解决原始FCN网络的输出不密集问题（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
    - 保证了网络最终的密集输出（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（3x3），元素间距变大（1→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

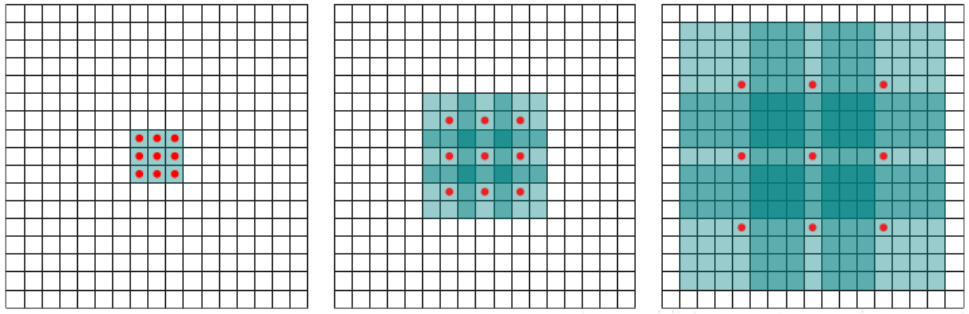


语义分割（Semantic Segmentation）

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→Rate
    - Rate越大，感受野越大

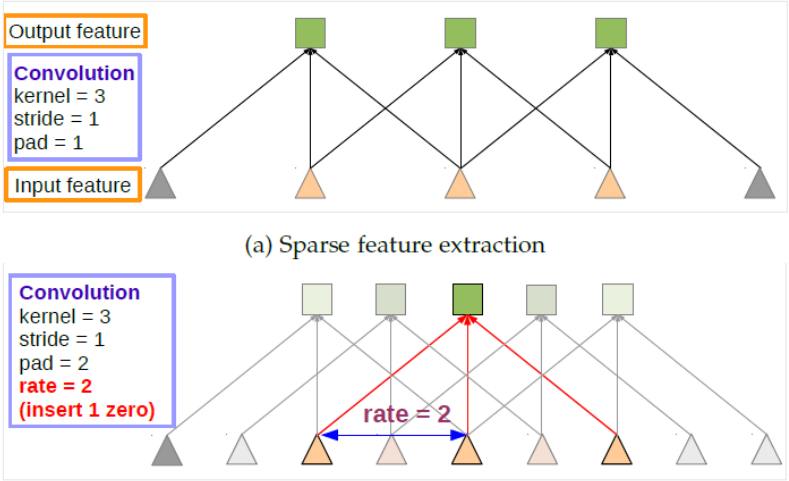
卷积核尺寸3x3



Rate = 1
无插零

Rate = 2
插1个零

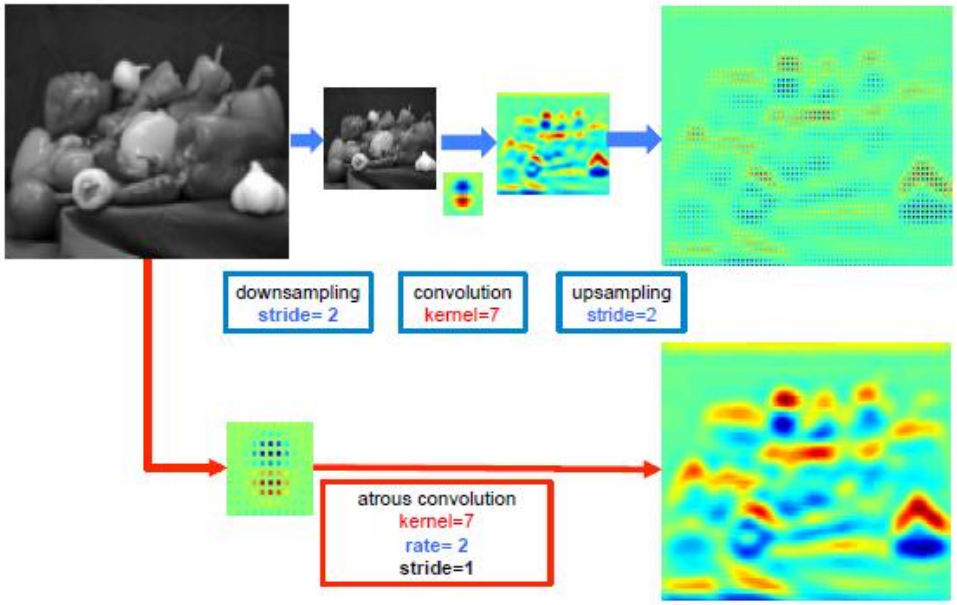
Rate = 4
插3个零



语义分割（Semantic Segmentation）

DeepLab-DCNN

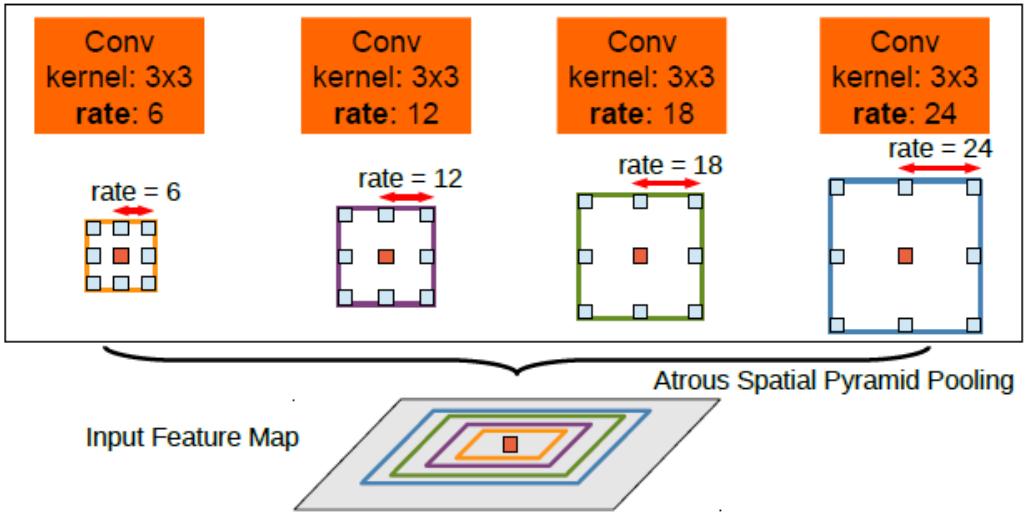
- 膨胀卷积效果
  - 稀疏特征提取：x2降采样→7x7卷积→x2上采样
  - 稠密特征提取：7x7膨胀卷积
- 优势
  - 参数&计算量一样
  - 灵活控制分辨率



语义分割（Semantic Segmentation）

DeepLab-DCNN

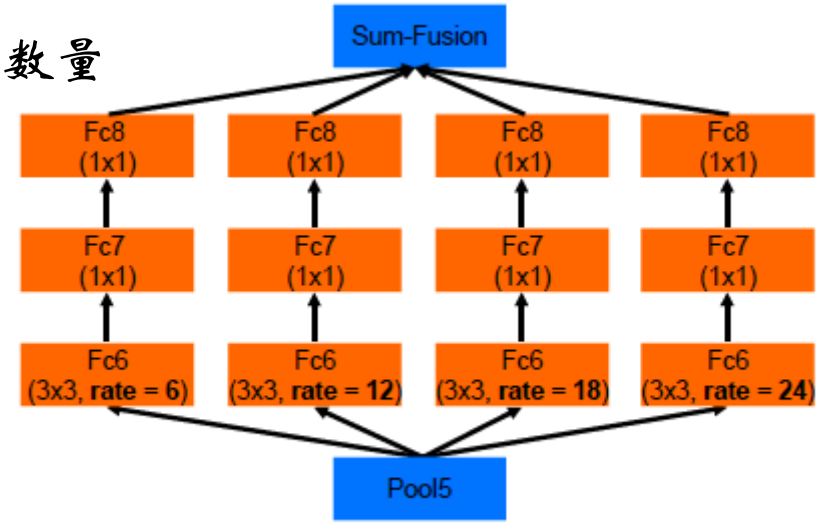
- Atrous空间金字塔池化（Atrous Spatial Pyramid Pooling）
  - 不同感受野（rate）捕捉不同尺度上的特征
  - 在Conv6层引入4个并行膨胀卷积
    - Rate: 6, 12, 18, 24



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→Fc7→Fc8
    - 深度：4096→2014→类别数量
    - 卷积核：3x3→1x1→1x1
  - 融合：概率相加



(b) DeepLab-ASPP

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
  - FCN网络输出结果的8倍双线性插值
  - 上一轮迭代结果
- 能量计算基于图片RGB像素值

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

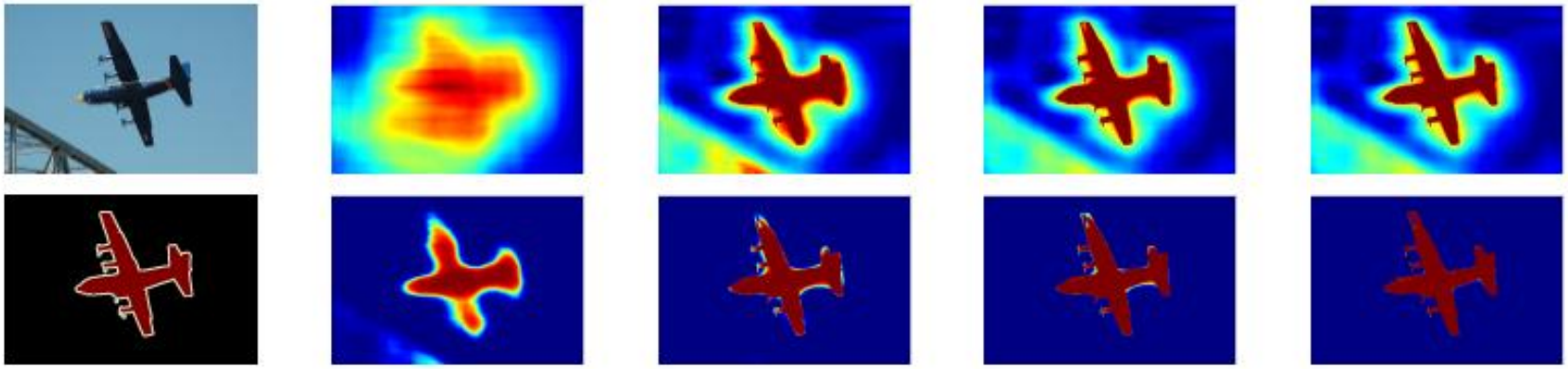
$$\theta_i(x_i) = -\log P(x_i)$$

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

语义分割（Semantic Segmentation）

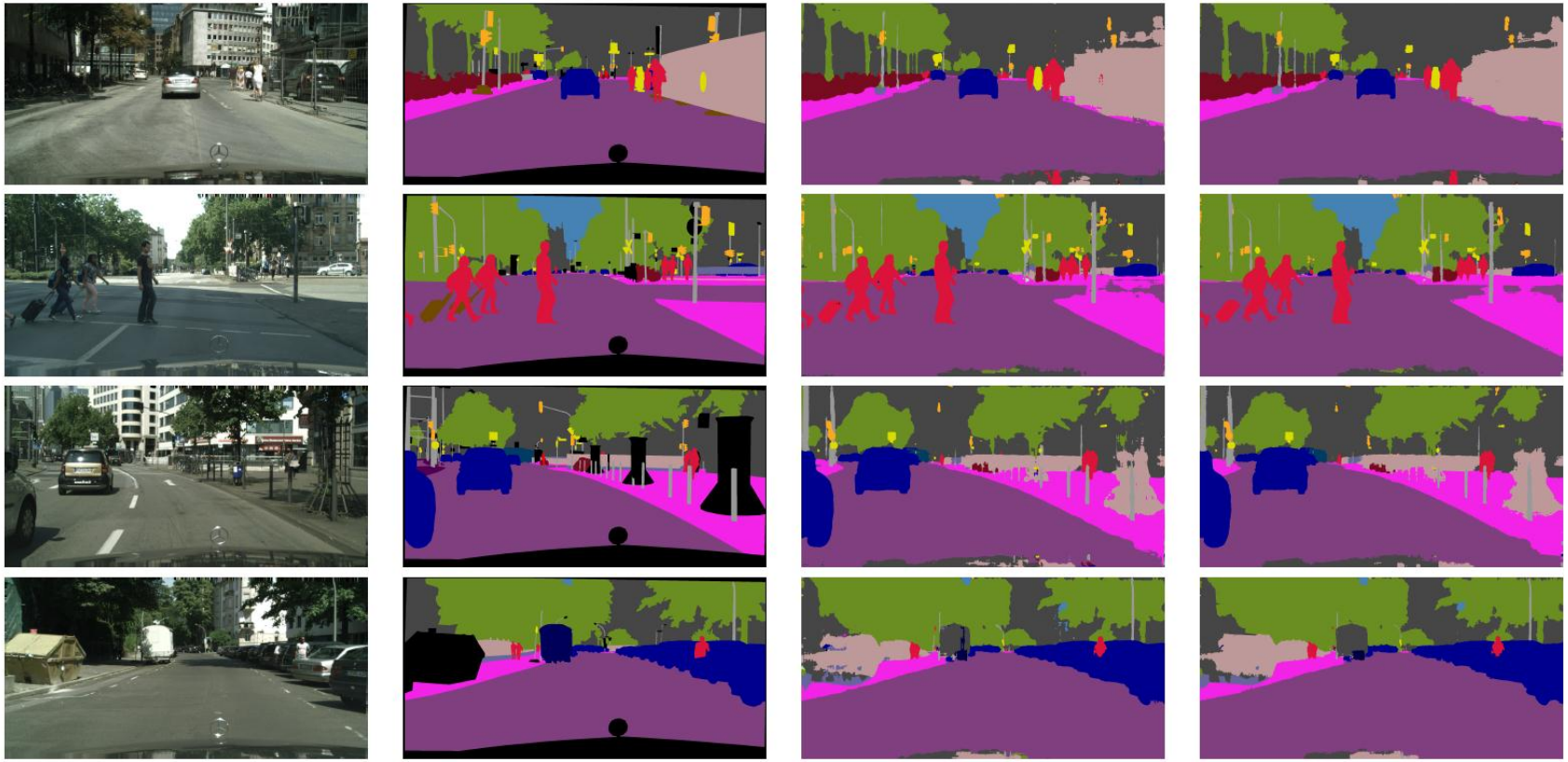
DeepLab-全连接CRF

- 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



(a) Image　(b) G.T.　(c) Before CRF　(d) After CRF

语义分割（Semantic Segmentation）

Cityscapes数据集性能

- ResNet-101 优于 VGG16

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

语义分割数据集

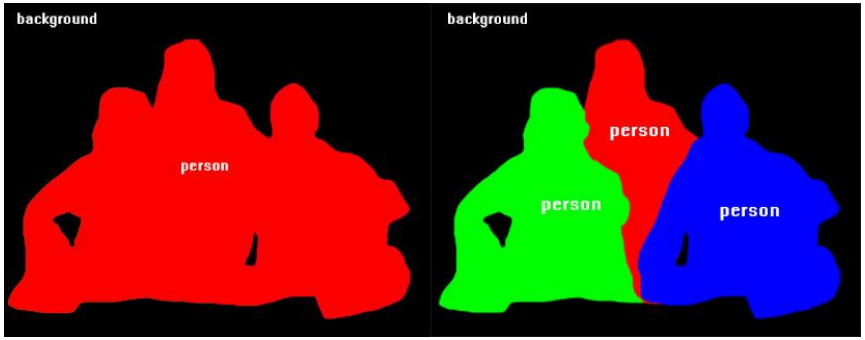
Pascal VOC - 2012

- 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
- 演示内容
  - 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/h5chJQt

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