

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（上）

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

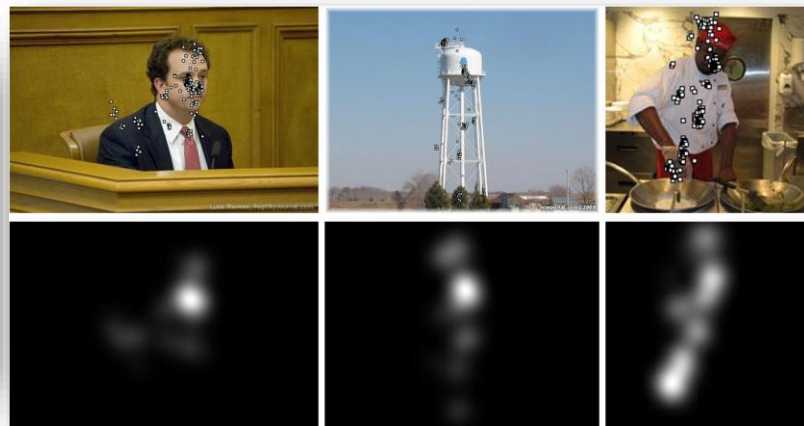
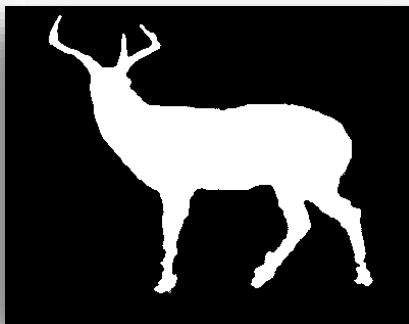
本章结构

- 显著性检测 (Saliency Detection)
- 物体分割 (Object Segmentation)
- 语义分割 (Semantic Segmentation)
 - FCN

显著性检测（Saliency Detection）

2类问题

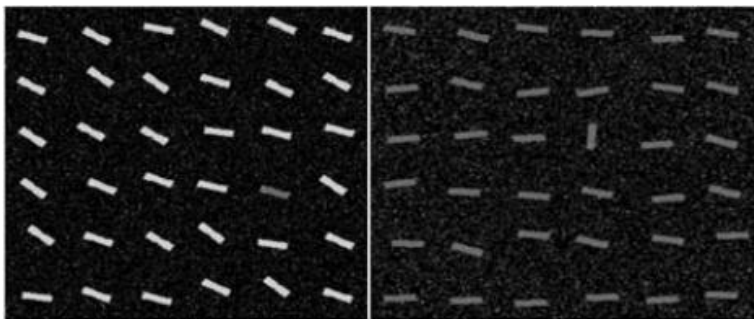
- 显著性物体分割（Salient object segmentation）
 - 最能引起人的视觉注意的物体区域
- 注视点预测（Fixation prediction）
 - 通过对眼动的预测和研究探索人类视觉注意机制



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



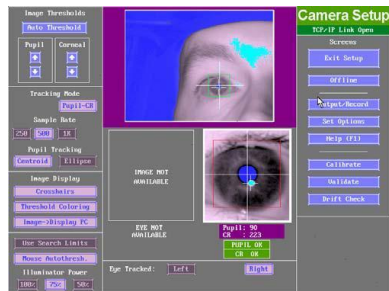
显著性检测 (Saliency Detection)

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

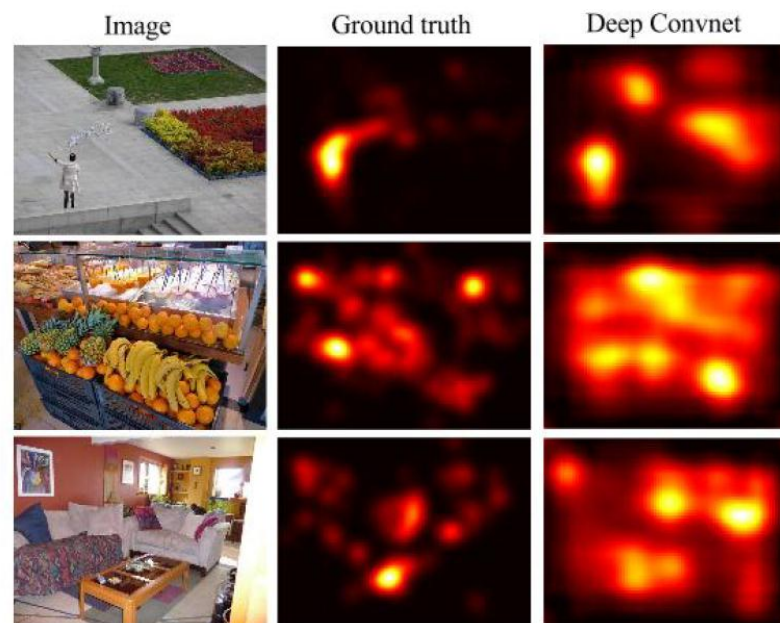
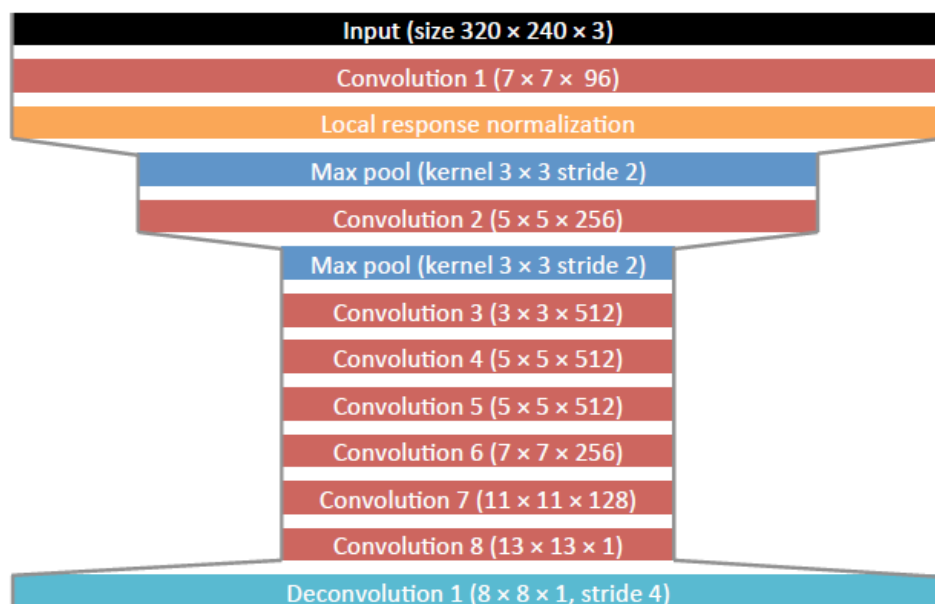


显著物体轮廓

显著性检测 (Saliency Detection)

DNN模型

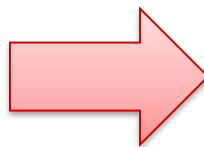
- 由VGG网络修改而成



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

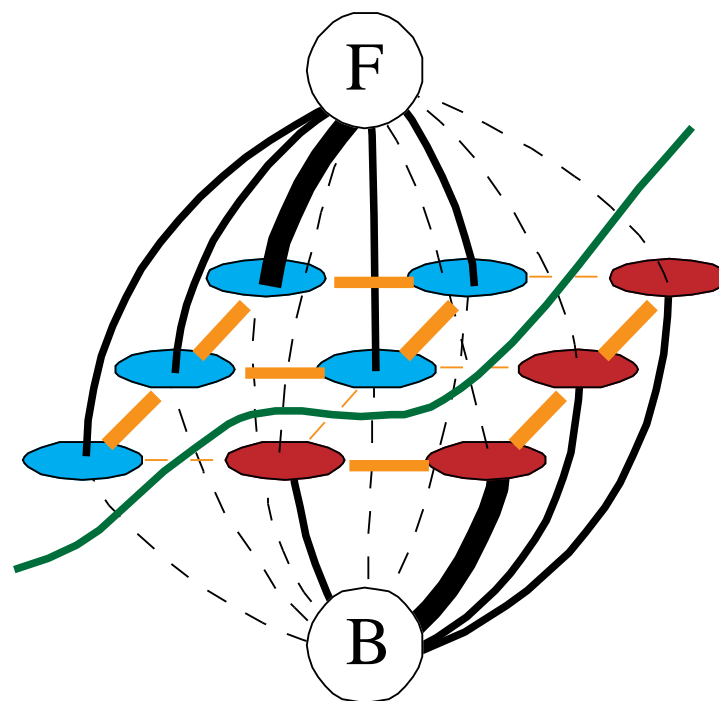


物体分割（Object Segmentation）

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

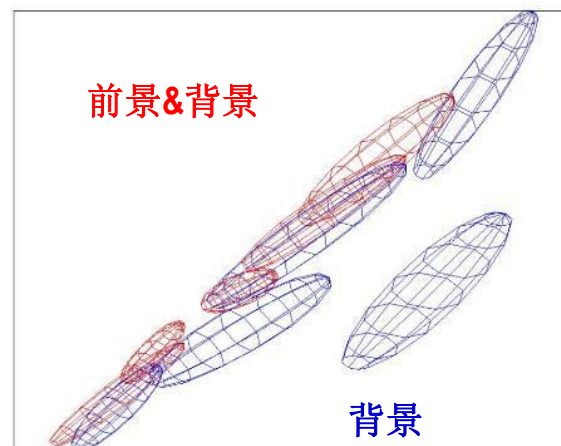
F	F	B
F	F	B
F	B	B



物体分割（Object Segmentation）

GrabCut 分割

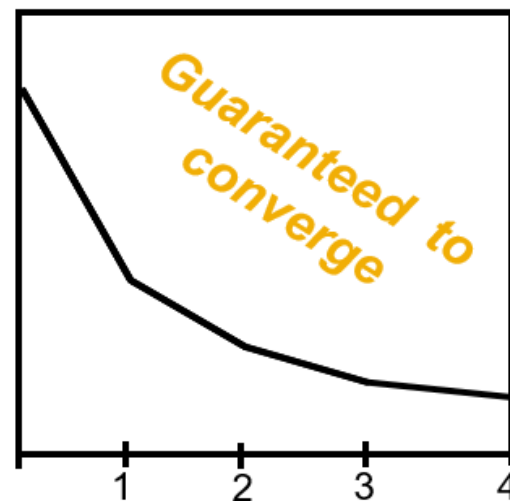
- 前景/背景的颜色模型
 - 高斯混合模型
 - Kmeans 算法获得



物体分割（Object Segmentation）

GrabCut 分割

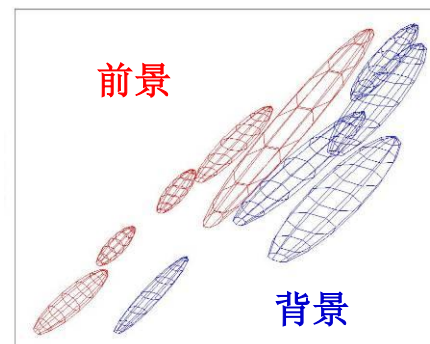
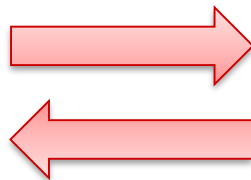
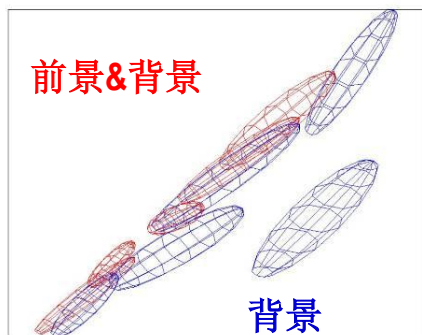
- 迭代进行 Graph Cuts
 - 优化前景和背景的颜色模型
 - 能量随着不断迭代变小
 - 分割结果越来越好



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 



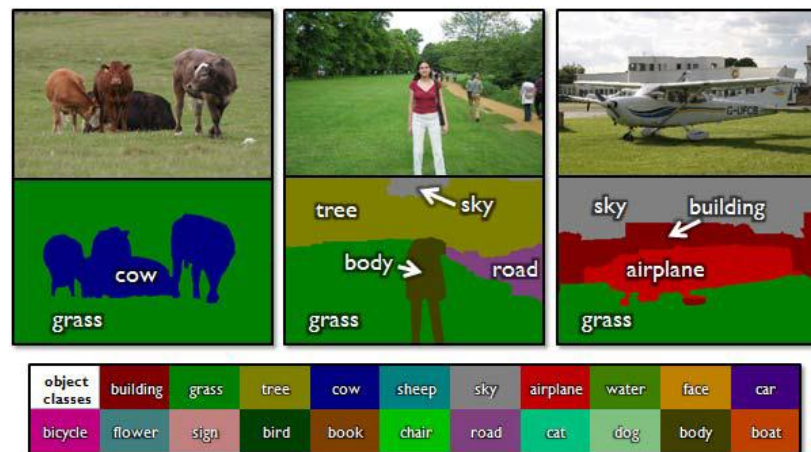
物体分割（Object Segmentation）



语义分割 (Semantic Segmentation)

什么是语义分割

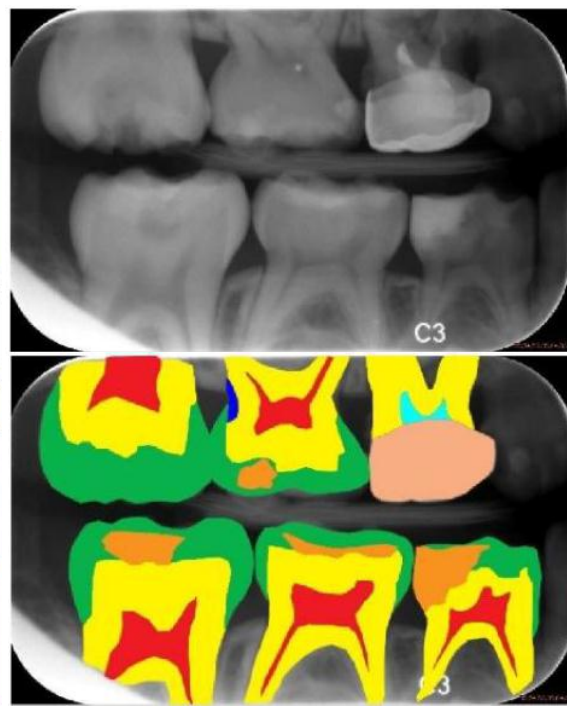
- 目标
 - 从像素水平 (pixel-level) 上, 理解、识别图片的内容
 - 根据语义信息分割
- 输入
 - 图片
- 输出
 - 同尺寸的分割标记 (像素水平)
 - 每个像素会被识别为一个类别 (category)



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 <https://youtu.be/rB1BmBOkKTw>
- 医学X光



语义分割（Semantic Segmentation）

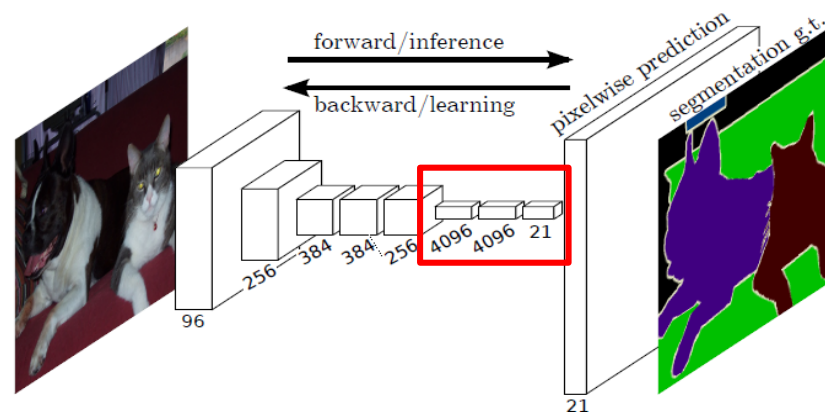
算法研究阶段

- 2015之前：手工特征+图模型（CRF）
- 2015开始：深度神经网络模型
 - 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）

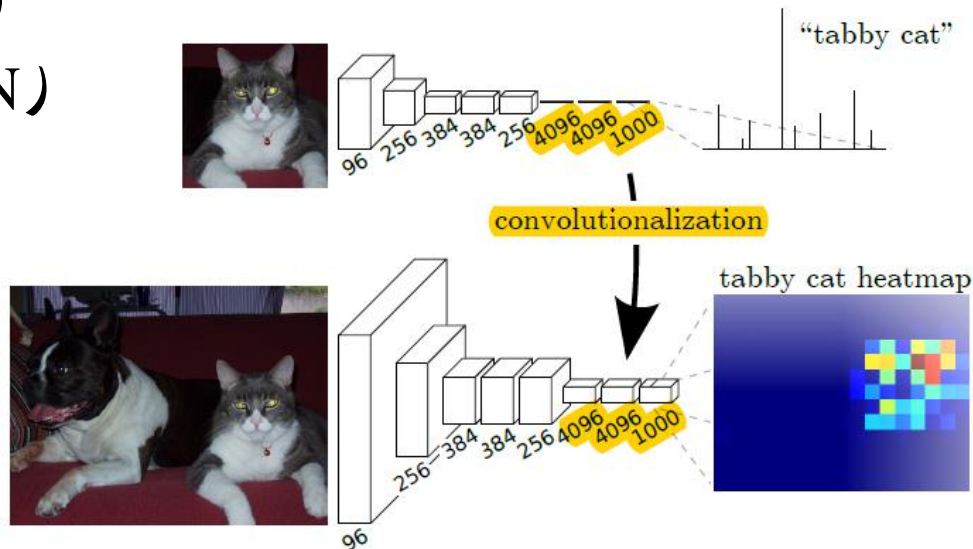
- **卷积化（Convolutionalization）**
 - 将所有全连接层转换成卷积层
 - 适应任意尺寸输入，输出低分辨率分割图片
- **反卷积（Deconvolution）**
 - 将低分辨率图片进行上采样，输出同分辨率分割图片
- **跳层结构（Skip-layer）**
 - 精化分割图片



语义分割（Semantic Segmentation）

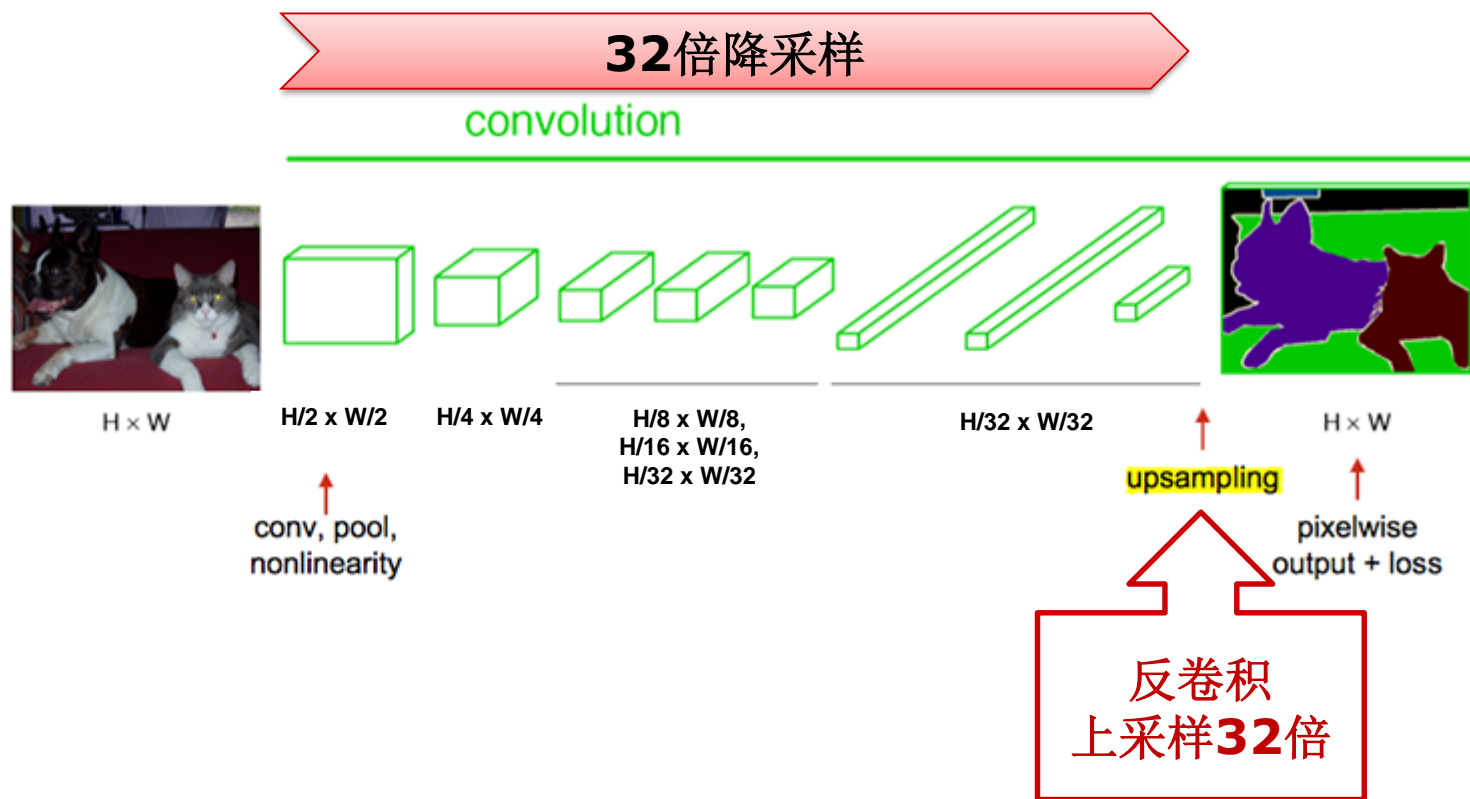
FCN-卷积化（Convolutionalization）

- 基础CNN网络: AlexNet, VGG16, GoogLeNet
- 卷积化后的核尺寸（通道数，宽，高）
 - FC6→ (1x1, 4096)
 - FC7→ (1x1, 4096)
 - FC8→ (1x1, 类别N)
- 分辨率降低32倍
 - 5个卷积层
 - 每层降2倍



语义分割（Semantic Segmentation）

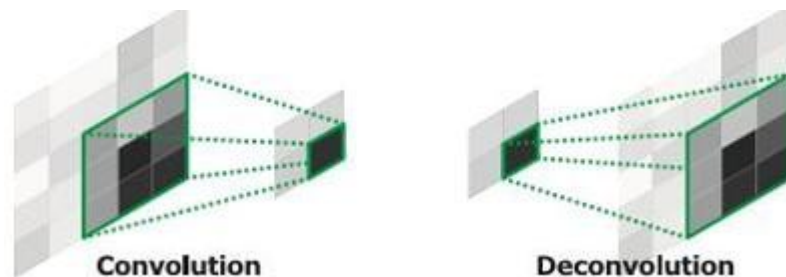
FCN-卷积化的降维问题



语义分割（Semantic Segmentation）

FCN-反卷积（Deconvolution）

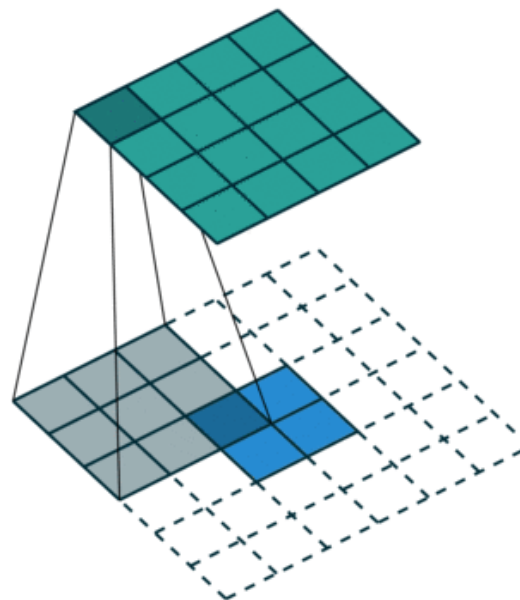
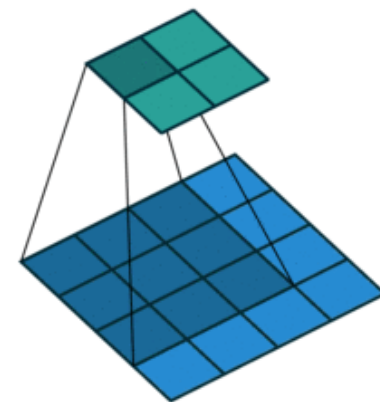
- 一对多操作
- 卷积的逆操作
 - 小数步长 $1/f$
 - 卷积核尺寸不变
- 前向和后向传播
 - 对应于卷积操作的后向和前向传播，优化上做颠倒
 - 反卷积核是卷积核的转置，学习率为0
- 也叫转置卷积（Transposed convolution）
- 可以拟合出双线性插值



语义分割（Semantic Segmentation）

FCN-反卷积（Deconvolution）

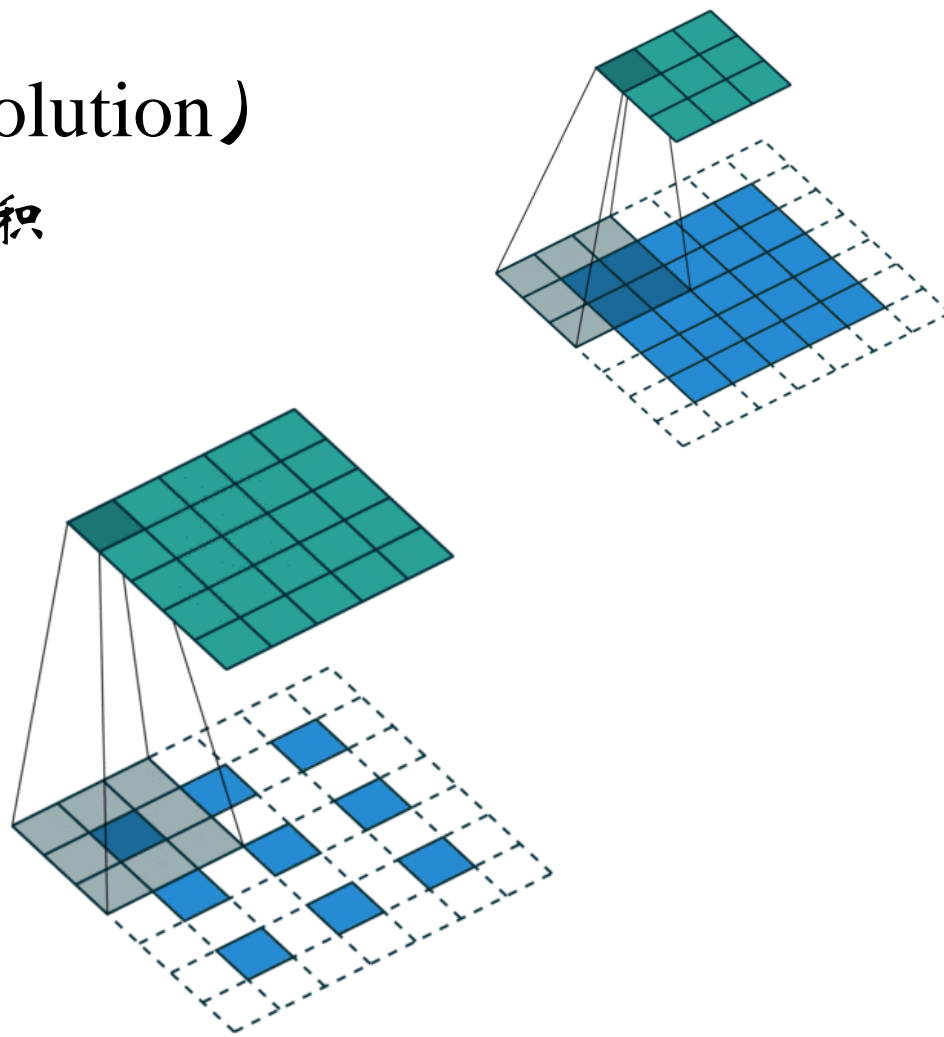
- 外围全补零（Full padding）反卷积
- 输入：2x2
- 输出：4x4
- 参数设置
 - 卷积核尺寸：3x3
 - 步长：1
 - Padding：2
- 被Skip-layer使用



语义分割（Semantic Segmentation）

FCN-反卷积（Deconvolution）

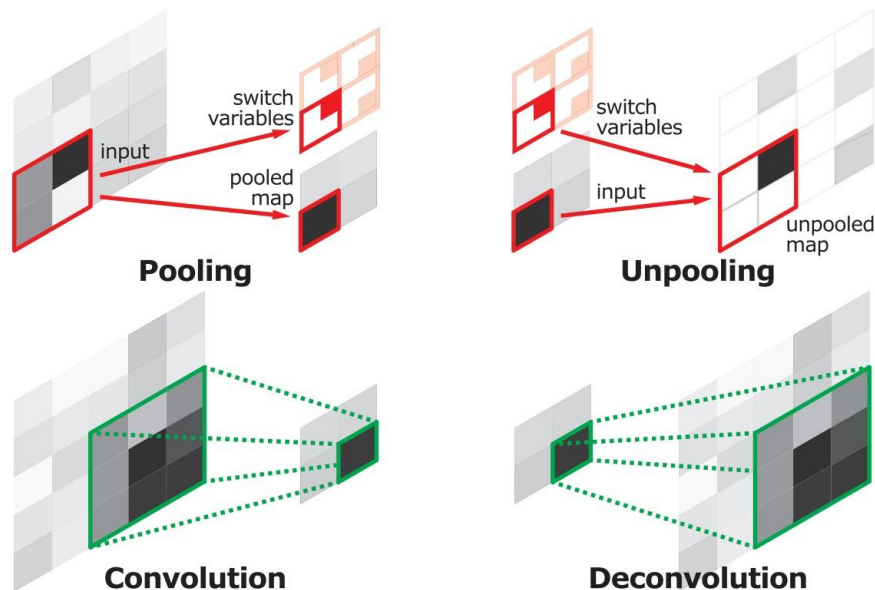
- 插零分数步长反卷积
- 输入：3x3
- 输出：5x5
- 参数设置
 - 卷积核尺寸：3x3
 - 步长：2
 - Padding：1



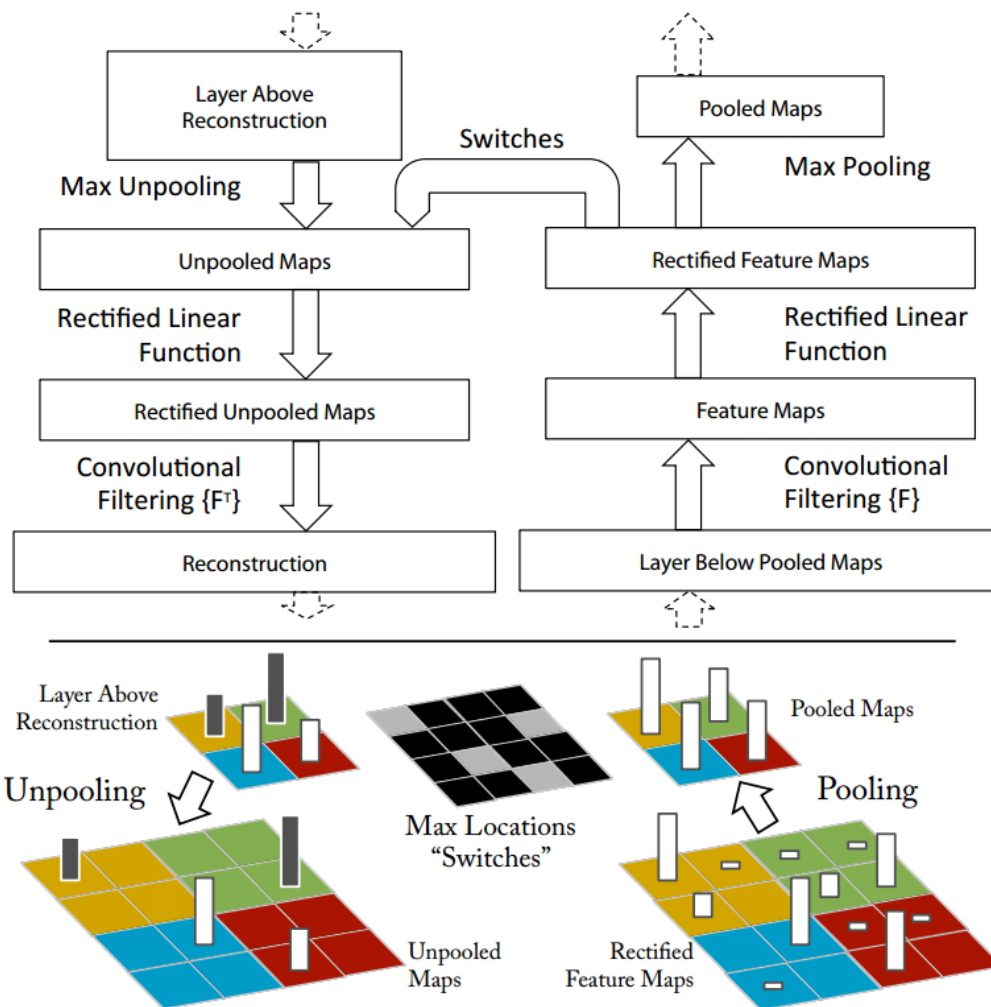
语义分割（Semantic Segmentation）

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）



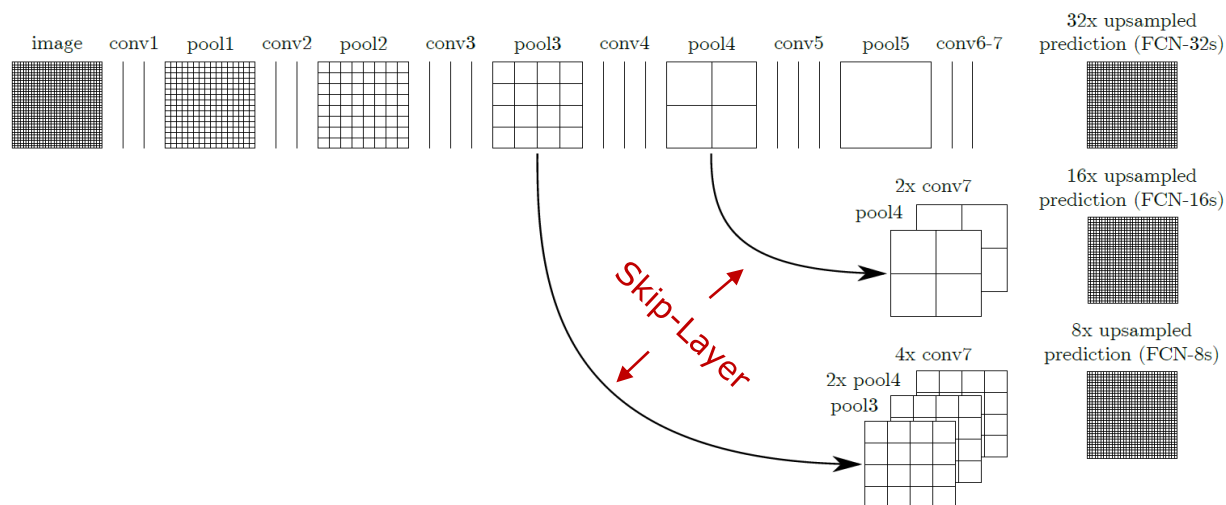
语义分割 (Semantic Segmentation)



语义分割（Semantic Segmentation）

FCN-跳层结构（Skip-layer）

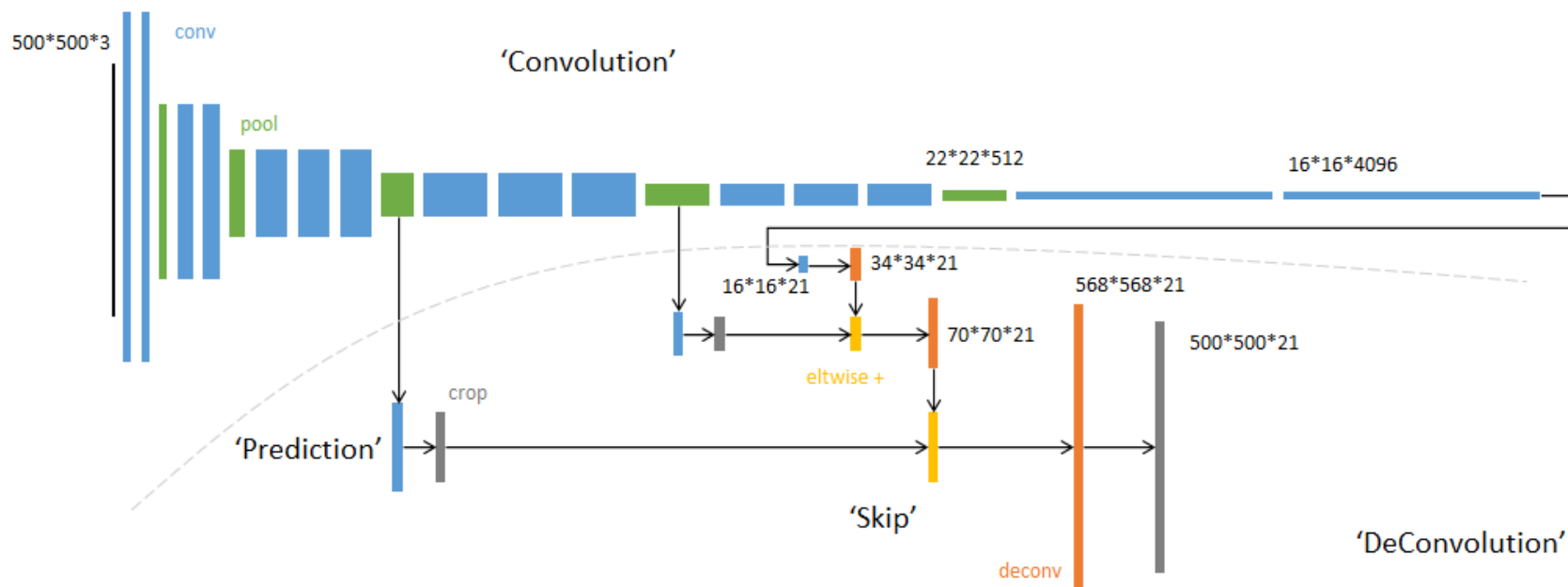
- 原因：直接使用32倍反卷积得到的分割结果粗糙
- 使用前2个卷积层的输出做融合
- 跳层：Pool4和Pool3后会增加一个1x1卷积层做预测
- 较浅网络的结果精细，较深网络的结果鲁棒



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

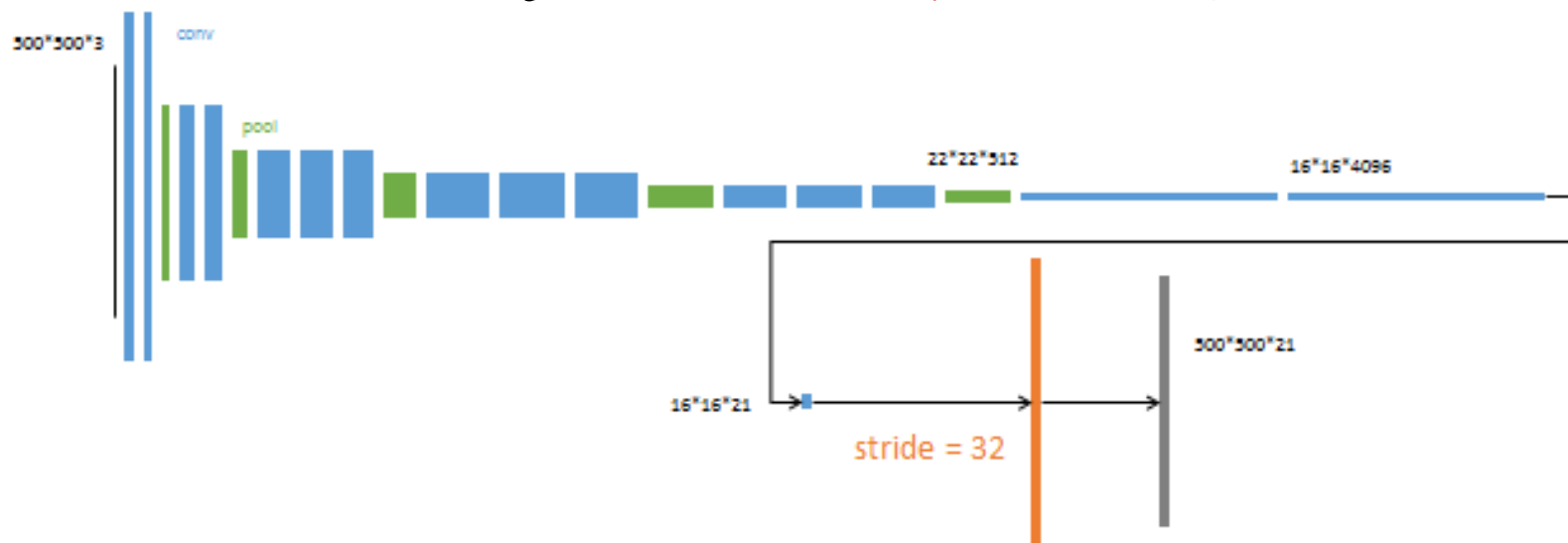
- 第1步
 - 使用AlexNet作为初始网络，保留参数
 - 舍弃最后1个全连接层



语义分割（Semantic Segmentation）

使用AlexNet构建FCN

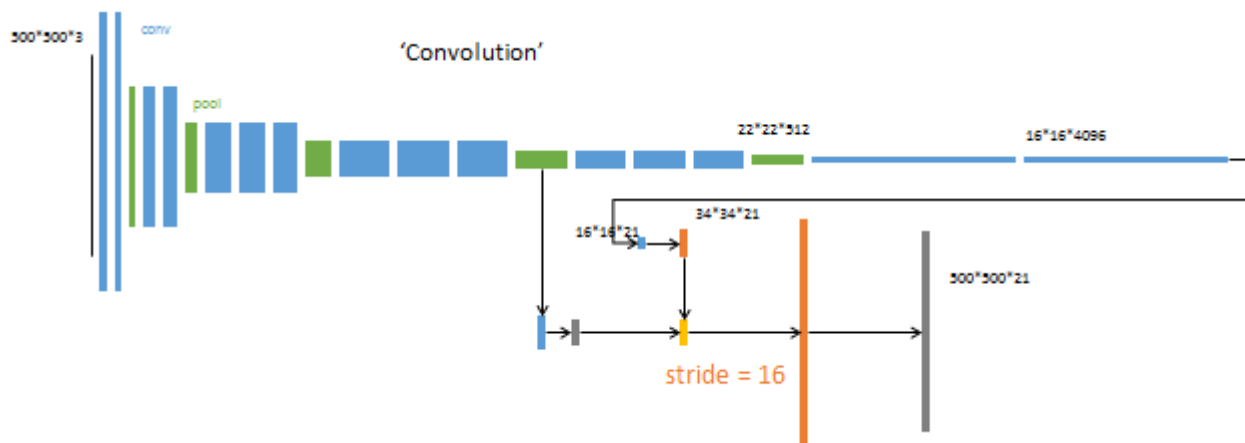
- 第2步（FCN-32s网络）
 - 替换为两个同深度的卷积层（4096,1,1） $\rightarrow 16 \times 16 \times 4096$
 - 追加一个预测卷积层（21,1,1） $\rightarrow 16 \times 16 \times 21$
 - 追加一个步长为32的双线性插值反卷积层 $\rightarrow 500 \times 500 \times 21$



语义分割（Semantic Segmentation）

使用AlexNet构建FCN

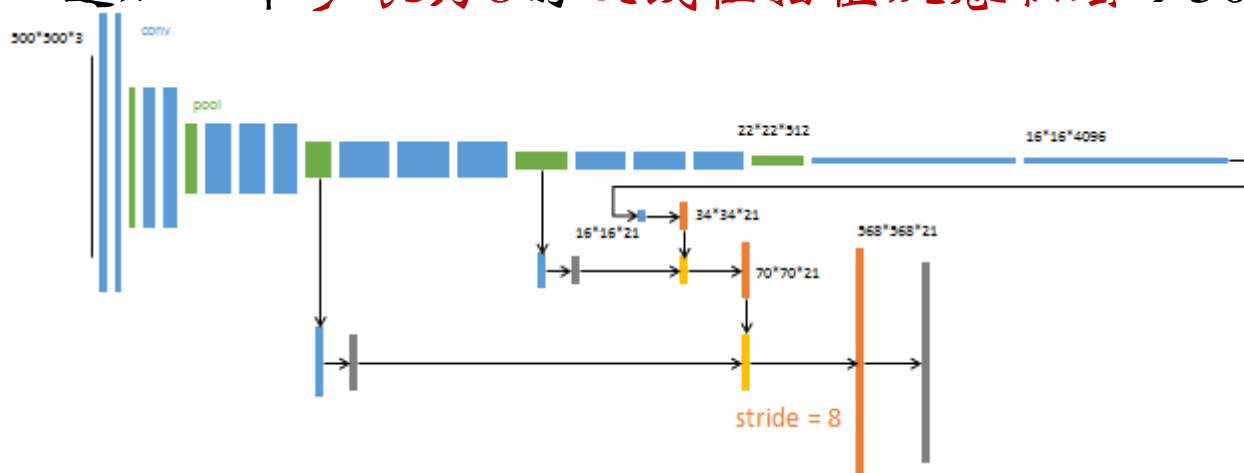
- 第3步（FCN-16s网络）
 - 对最终层Conv7结果2倍上采样 $\rightarrow 34 \times 34 \times 21$
 - 提取Pool4输出，追加预测卷积层（21,1,1） $\rightarrow 34 \times 34 \times 21$
 - 相加融合 $\rightarrow 34 \times 34 \times 21$
 - 追加一个步长为16的双线性插值反卷积层 $\rightarrow 500 \times 500 \times 21$



语义分割（Semantic Segmentation）

使用AlexNet构建FCN

- 第4步（FCN-8s网络）
 - 对上次融合结果2倍上采样 $\rightarrow 70 \times 70 \times 21$
 - 提取Pool3输出，追加预测卷积层 $(21, 1, 1) \rightarrow 70 \times 70 \times 21$
 - 相加融合 $\rightarrow 70 \times 70 \times 21$
 - 追加一个步长为8的双线性插值反卷积层 $\rightarrow 500 \times 500 \times 21$



语义分割 (Semantic Segmentation)

FCN训练

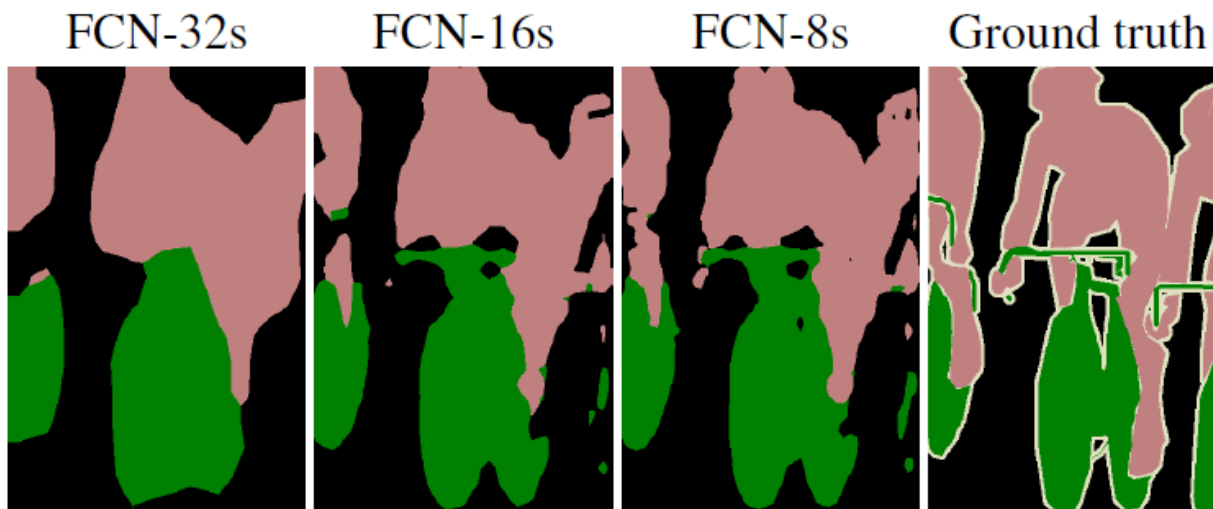
- SGD with momentum (0.9)
 - Learning rate
 - 0.001(AlexNet), 0.0001(VGG16), 0.00001(GoogLeNet)
 - Minibatch: 20
- 初始化
 - 卷积层
 - 前5个卷积层使用初始CNN网络的参数
 - 剩余第6和7卷积层初始化为0
 - 反卷积层
 - 最后一层反卷积层固定为双线性插值, 不做学习
 - 剩余反卷积层初始化为双线性插值, 做学习

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

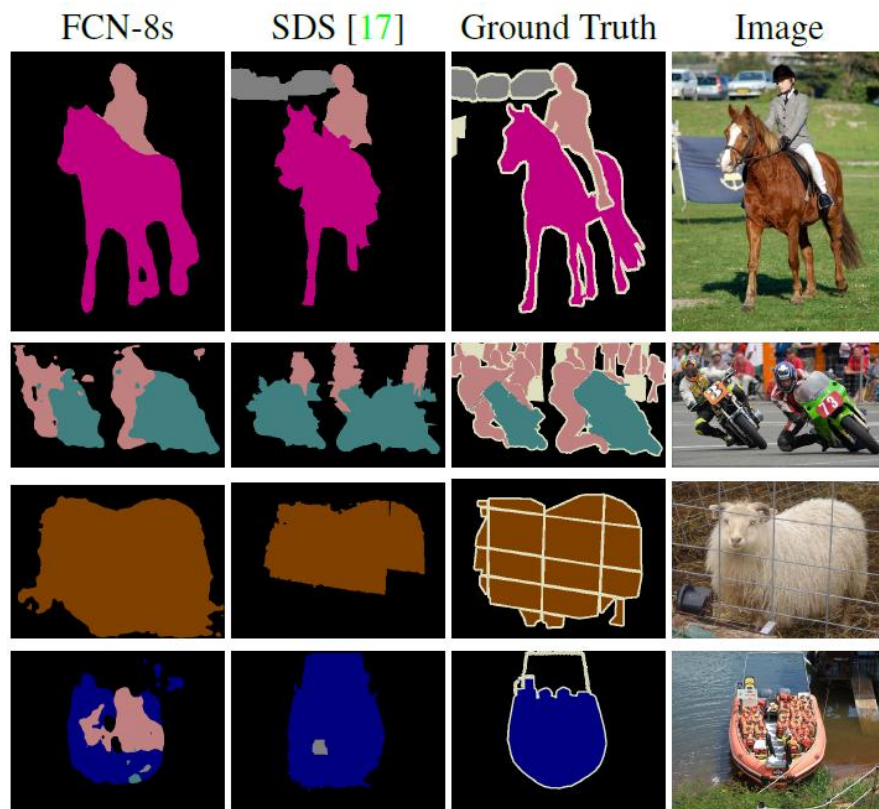
- FCN-VGG16最优

	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

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

