Fully Convolutional Networks for Semantic Segmentation
基于全卷积网络的语义分割

摘要

- 卷积网络
  - 是强大的视觉模型
  - 可以产生多层次的特征
- 提出了
  - 全卷积网络FCN
    - **端到端，像素到像素**地训练
    - 重点
      - 在**语义分割**领域上超越了最先进的技术
      - **接受任意大小的输入**，经过推理和学习，**产生相应大小的输出**
    - 内容
      - 定义并描述全卷积网络，将**分类**转移到**密集预测（像素到像素）**
      - **将现有的分类网络调整为FCN**
        - 现有分类模型
          - AlexNet
          - VGG net
          - GoogLeNet
        - 采用数据集
          - PASCAL VOC
          - NYUDv2
          - SIFT Flow
      - 通过**微调（fine-tuning）**将学习到的参数转移到**分割**任务上（**迁移学习？**）
      - **定义一个新颖的“跳跃”结构**
        ->解决局部信息随着网络加深而丢失的问题（将浅层信息备份）
        将来自**深的、粗糙**（图像小，物体的空间信息比较丰富）的层的**语义信息** 和 来自**浅的、细致**（图像大，物体的几何信息比较丰富）的层的**表征信息**结合

介绍

- 卷积网络
  - 每个像素都被标记为属于 包含它的目标或区域 的类别
    - 已经实现的**语义分割**
    - **从粗糙到细致的推理的改进**
      对每个像素进行预测
- **端到端、像素到像素**的全卷积网络FCN
  - **有监督的预训练（supervised pre-training）**
  - **像素预测（pixelwise prediction）**

相关工作

- 深度分类网络
  - 迁移学习
    - 视觉识别
    - 检测
    - 实例分割和语义分割
- FCNs
  - 具有全卷积推理和学习的检测
    - Matan首次将卷积网络LeNet扩展为**任意大小的输入**（数字串的识别）
    - Wolf&Platt将卷积网络扩展为**2维图像的输出**
  - 具有全卷积推理
    - Ning粗糙的**多类别分割**
    - Sermanet滑动窗口检测
    - Pinheiro&Collobert语义分割
    - Eigen图像恢复
  - 具有全卷积训练
    - Tompson姿势估计
- 使用卷积网络进行**密集预测**
  - 历史工作
    - Ning语义分割
    - Ciresan边界预测
    - Eigen图像恢复和深度估计
  - 历史例子的特点
    - 限制容量和接受野
    - patchwise拼凑式训练
    - 输入移位-输出交错
    - ……
- 在深度分类架构上扩展
  - 类似的历史方法
    - 在一个混合模型中，采用了深度分类网络来做语义分割
    - 用采样边界盒/区域来微调一个R-CNN
      不是端到端学习
  - **本文**
    - 使用图像分类作为有监督的预训练
    - 通过 微调全卷积 简单高效地学习输入的整个图像
      端到端的

FCNs

- 使用分类器（图像级别）进行密集预测（像素级别）
  - 经典的识别网络中的全连接层 要求**固定大小的输入**，且产生**非空间（一维）的输出**
    - LeNet
    - AlexNet
  - 从分类器到全卷积
    将最后的**全连接层换成卷积层**得到FCN，
    可以输入任意大小的图像并产生相应大小的输出
    - **产生问题**：
      输出的维度会因为二次采样不断降低，图像越来越小
      需要将图像进行不断放大到原图像的大小
- 粗糙的输出->密集预测 的解决方法
  - 移动和缝合shift-and-stitch（没有采用）
    - 不插值
    - 输入移位+输出交错
  - 网络内向上采样upsampling（即反卷积deconvolution）
    - **插值**
      简单的双线性插值仅依赖于输入和输出神经元的相对位置
    - **反卷积操作简单**
      只要将卷积的向前和向后传递做反向处理
    - 向上采样应用在网络内**端到端**的学习（从像素到像素的损失开始反向传播）
    - 反卷积的过滤器不需要固定，可以通过学习得到
- patchwise拼凑式（小块）训练 vs 整个图像的全卷积训练
  - （采样）patchwise训练（没有采用）
    - 减少图像的冗余信息
    - 常用来解决图像的空间相关性
  - （整个图像）全卷积训练（采用，figure5）
    - 对loss加权->解决类不平衡性
    - 对loss采样->解决输入图像的空间相关性

分割架构

- 从分类器到密集FCN
  - 丢弃每个网络的分类器层
  - 将全连接层替换为卷积层
- 改进预测输出的精度
  - 层融合——建立连接将最后的预测层与较低层，以合适的步长结合（figure3）
    - 对较浅层（比如pool4，1/16）添加1x1的卷积层 -> 产生类别预测（1/16）
    - 对较深层（比如conv7，1/32）添加2x的向上采样层 -> 产生类别预测（1/16）
    - 将两个类别预测（1/16）相加（融合）后，向上采样/反卷积（步长16）
      -> 产生跟原图像一样大小的预测图（FCN-16s）
    - 实验表示，继续到FCN-8s（与pool3融合）会使结果有所改进，
      但不需要继续与更浅层（pool2、pool1）融合，改进不大
  - 其他方法
    - 减小池化层的步长 -> 需要增大卷积核
    - shift-and-stitch -> 改进不如层融合
- 实验框架
  - 优化
    - SGD
  - fine-tuning微调
    - 通过整个网络的反向传播对所有层进行微调
  - patch sampling小块采样（不必要）
    - 对过多的图像需要花费更多时间收敛
  - 类别平衡（不必要）
    - 类别不平衡——有3/4的部分是背景
    - 全卷积训练可以通过对loss进行加权和采样来平衡
  - 密集预测（像素到像素）——通过向上采样
    - **最后一层**的反卷积过滤器固定为 双线性插值
    - **中间**的向上采样层，初始化为双线性向上采样，之后学习得到
  - 在每个方向预测为32像素（没明显改进）
  - 使用更多的训练数据（有改进）

结果（figure6）

- 使用PASCAL数据集效果最好
  - 可以恢复精细的结构
  - 可以分离紧密相交的物体
  - 受遮挡物的影响小