[[1]](#footnote-1)

**基于deeplab v3+神经网络模型的图像分割研究**

经润东1

摘 要: 利用deeplab v3+开源代码训练PASCAL VOC 2012数据集,研究其权重设置以及在voc2012的基础上添加自建的数据集,然后完成训练.其中在各大搜索引擎中找到的猪作为添加数据集的数据

关键词: 计算机视觉;深度学习; DeepLabv3 架构;图像分类

**Image segmentation based on deeplab v3 neural network model is studied**

Rundong1

**Abstract:**  using deeplab v3+ PASCAL VOC 2012 + open source code training data set, and study its weight set and on the basis of voc2012 add self-built data sets, and then complete the training. One in the major search engines to find pigs as adding data sets of data

**Key Words:** Computer vision;Deep learning;DeepLabv3 architecture;Image classification

1. 引 言

语义分割是计算机视觉中的基本任务，在语义分割中我们需要将视觉输入分为不同的语义可解释类别，「语义的可解释性」即分类类别在真实世界中是有意义的。例如，我们可能需要区分图像中属于汽车的所有像素，并把这些像素涂成蓝色。

虽然像聚类这样的无监督方法可以用于分割，但其结果不一定是有语义的。这些方法无法对它们训练的类进行细分，但是在搜索区域界限方面更加擅长。

与图像分类或目标检测相比，语义分割使我们对图像有更加细致的了解。这种了解在诸如自动驾驶、机器人以及图像搜索引擎等许多领域都是非常重要的。因此，本文讨论的主题是使用深度学习方法进行有监督的语义分割。

. 目前有一些常用于训练语义分割模型的数据集：Pascal VOC 2012：有 20 类目标，这些目标包括人类、机动车类以及其他类，可用于目标类别或背景的分割;Cityscapes：50 个城市的城市场景语义理解数据集;Pascal Context：有 400 多类的室内和室外场景;Stanford Background Dataset：至少有一个前景物体的一组户外场景。本文选择Pascal VOC 2012作为训练数据集,其中用于语义分割的图片共有2913张,同时本文从搜索引擎中选取了209张猪的图片,在对图像进行预处理以及标注之后添加到Pascal VOC 2012中,由于训练集中包含有21个类,训练完成的神经网络有时会出现对相似类的误判,不过deeplab v3+对于图像的边缘具有很好的判断效果.

deeplab v3+是Google 团队在2018年春节前夕在 arXiv 发布的语义分割模型,在如今deeplab v3+仍处于业内顶尖水平.

1. 卷积神经网络

卷积神经网络( convolutional neural networks， CNN) 作为深度学习一种经典模型，是受到生物视 觉认知启发( 局部视野) 而专门设计的一种针对二 维结构图像的多层感知器。近年来，随着科技的进 步，卷积神经网络模型在遥感图像分类的精度已大 幅度提高。 卷积神经网络由卷积层、非线性映射层、池化层 ( 有的称为下采样层) 和全连接层组成。数值计算 设备的不断改进促使卷积神经网络在计算机视觉等 领域取得快速发展，同时大数据环境也推进了卷积 神经网络结构的不断优化。采用卷积神经网络算法 对高分辨率遥感图像进行分类，避免了特征提取和分类过程中数据重建的复杂度，相比传统的非监督 分类方法，基于卷积神经网络的遥感图像分类算法 在分类精度方面有所提高;同时，卷积神经网络具有 有效减少训练中输入的参数数目，并利用卷积层和 池化( 下采样) 层分别降低噪声影响、聚合遥感图像 不同位置特征的特点，在降低特征维度的同时，特征 的尺度不变，还不容易出现过拟合的现象。

1. deeplab v3+
   1. ResNet模型

ＲesNet( residual neural network) 即深度残差网 络模型，最早是由微软亚洲研究院何凯明等人提出， 在大赛中取得冠军，并且错误率仅为 3. 57%，相比 AlexNet、VGGNet 以及 GoogleNet 等 经 典 模 型， ＲesNet 在遥感图像分类效果上非常突出，其结构可 以极快地加速神经网络的训练，模型在遥感图像分 类的准确率也有比较大的提升。 同时， ＲesNet 在网络结构优化上优势明显，传 统的卷积神经网络在处理遥感图像分类过程中传递 特征信息或多或少会存在信息丢失和损耗等问题， 同时还可能出现梯度消失或者爆炸等问题，因此很 多网络无法参与训练。ＲesNet 在一定程度上解决 了这个问题，通过直接绕道输入信息并输出，有效地 保护了信息的完整性，即使网络层逐渐加深，仍然能 够保持较好的分类效果; 且相较于传统的网络模型 结构，整个模型只需要学习输入、输出差别的这一部 分，有效简化了学习目标和难度.

* 1. DeepLabv3+ 架构

Deeplabv3+ 架构是结合了深度卷积神经网络 ( deep convolutional neural networks， DCNNs) 和概率 图模型( dense conditional random fields， DenseCＲFs) 的方法，该方法设计了串行和并行的带孔卷积模块， 采用多种不同的 atrous 比率来获取多尺度的内容信 息; 该架构中的 Atrous Spatial Pyramid Pooling ( ASPP) 模块通过挖掘不同尺度的卷积特征以及编 码了全局内容信息的图像层特征来提升分类效果。 Atrous convolution，即带孔卷积，为了解决局部 图像空间信息变换的不变性阻碍密集预测任务的问 题，使用带孔卷积来进行预训练，它主要是通过删除 下采样所提取密集的要素网络最后几层以及上采样 相对应的滤波器核，相当于在不同的滤波器权重之 间插入 holes，来提取更紧凑的特征，该架构可以控 制 DCNNs 计算特征响应的分辨率，而无需学习新的 额外参数。 ASPP，采用 4 个并行的不同 atrous 比率的带孔

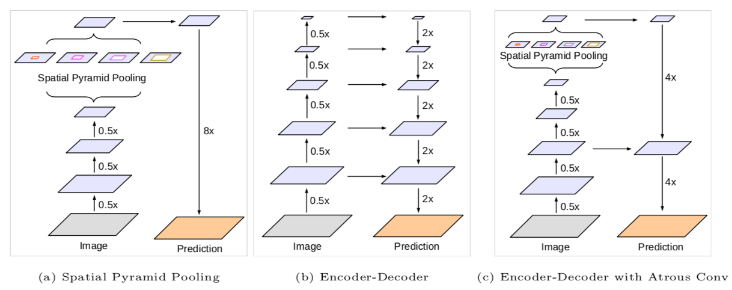
卷积对特征映射进行处理，它的灵感来自空间金字 塔池化，它表明对不同尺度的特征进行重新采样，可 以准确有效地对任意尺度的区域进行分类。 Deeplabv3plus 以 Xception 网络作为骨架网络， 在 ASPP 和解码模块使用深度可分离卷积提高编码 器－解码器网络的运行速率和稳定性，采用空间金 字塔模块和 encoder－decoder 结构，Xception 网络模 型中引入了 ASPP 模块，可以在多尺度上捕获信息， 实现遥感图像分类。Deeplabv3+ 架构原理见图 1。

图 1 DeepLabv3+ 原理示意图

通常网络中分类准确率和位置准确性是一对矛盾的存在，有一些工作尝试在保证分类准确率的条件下解决位置问题：

* 利用卷积网络中多个层次的信息
* 利用超像素，把位置问题交给底层次的分割处理（此处低层次分割是无语义分割）

一般利用到条件随机场（CRFs）来处理分割中不光滑问题，它只考虑到目标像素点的附近点，是一个短距离的CRFs。由于网络中得到的结果已经比较光滑了，更希望的是修复一些小的结构，所以用到了全连接的CRF模型。它的能量函数：

目标就是要最小化能量函数，先看第一项：



这一项保证了分类的准确率，准确率越高P(x)越接近1（0到1），此项值越小。

再看第二项：



此处有：



说明只考虑标签不相同的两个像素点，因为是全连接所以这两个点为**模型中任意两点.** 

为其右边式子的权重，右边的它是一个高斯核函数：

此高斯核函数具体为：



此核函数主要由两个像素点的位置和颜色决定，位置为主，颜色为辅。右下角带有α、β、γ的参数和w1、w2参数，试验时由w2=3和γ=3决定，搜索最优.

* 1. 代码实现
     1. 数据处理

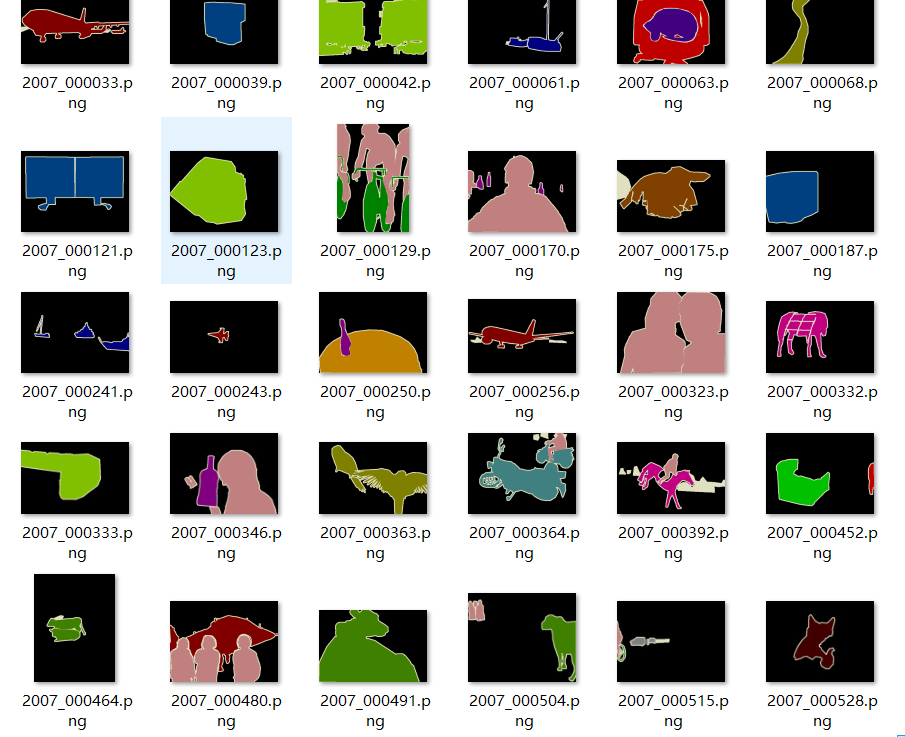
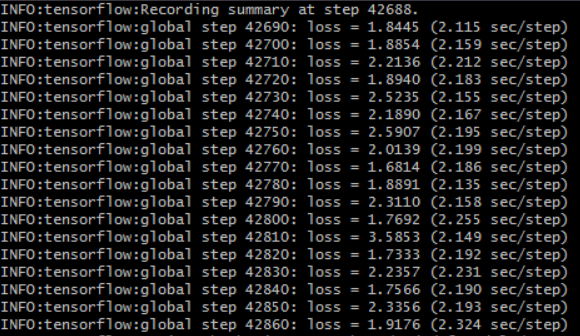
使用labelme对图片进行标注,导出相应的json文件, 调用labelme工程文件下的labelme2voc.py生成VOC格式的标签数据图2 Pascal VOC 2012数据集



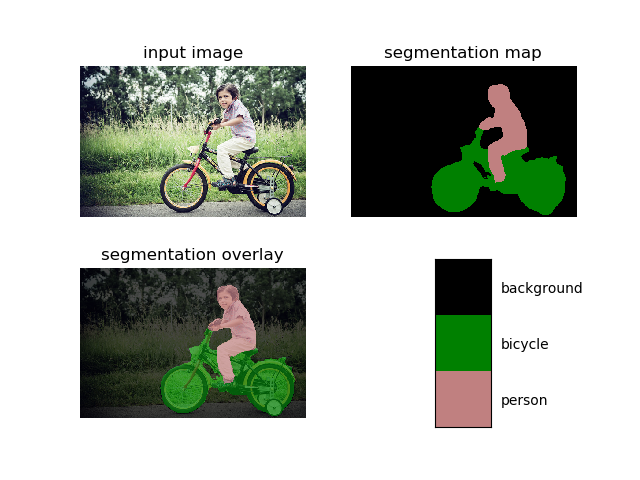
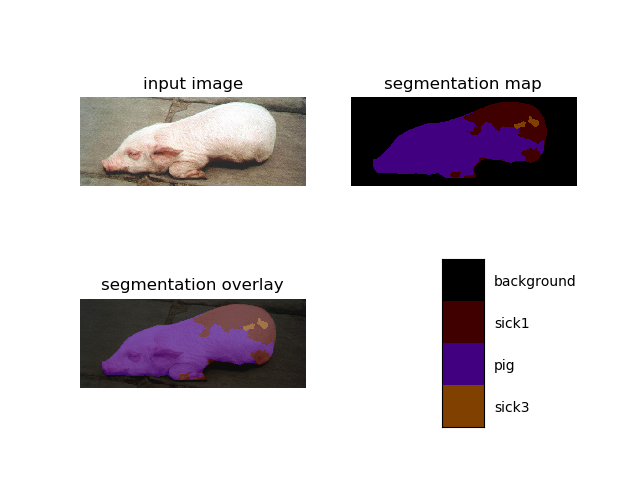
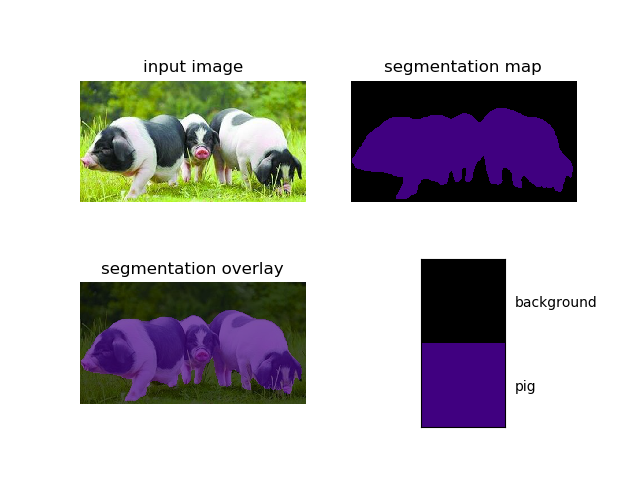
图3 自建数据集

移除 ground-truth 中的 colormap,生成对应的TFRecord文件,注册自己的数据集,其中还需要添加数据集的相关配置，在get\_dataset\_colormap.py文件中修改, 写一个用来返回一个np.ndarray对象的函数, 最后修改create\_label\_colormap函数

* + 1. 训练模型

以local\_test.sh中的代码为例,修改其中部分参数大小和名称,以匹配我们数据集

漫长的训练完成之后得到checkpoint文件, 而进行预测时只需模型和权重，不需要 metadata，故可使用官方提供的脚本生成推导图。运行后导出,以下为预测结果



1. 参考文献
2. DeepLabv1: Semantic image segmentation with deep convolutional nets and fully connected CRFs ICLR 2015 (International Conference on Learning Representations)
3. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation
4. GC.C,J.AK.Markov random field.texture models［J］ ．.IEEE Transactions on Pattern Analysis and Machine Intelligence.

1. [↑](#footnote-ref-1)