

利用 faster rcnn 训练自然文本图像过程：

首先是配置好 py-faster-rcnn,然后运行 faster-rcnn 自带的 demo 来验证配置成功。接下来就是尝试应用 faster-rcnn 来跑通 ICDAR 挑战 2013 的文本数据集。

数据集的制作过程还是很顺利的(刚开始的时候我直接拿师兄制作好的数据集放在了 voc 相应的文件夹), 后来我按照

voc2007 的格式自己创建一个文件夹 text2 里面分别再建 3 个文件夹 Annotation, ImageSets, JPEGImages 分别存放 xml 文件, txt 文件, 图片文件。ImageSets 下创建一个 Main 文件夹存放 4 种 txt 文件, train.txt,trainval.txt,test.txt,val.txt.各个文件的数据的大小视情况而定。

数据集做好之后接下来就开始修改源码。关于数据集的读写源码主要是/lib/dataset/tools 文件夹下的三个 py 文件。分别为 imdb.py,pascal_voc.py,factory.py。 **imdb.py**：数据库读写类的基类，里面定义了许多数据的操作。

pascal_voc.py：由 imdb 派生的一个数据集子类，rgb 大神这个类进行 pascal_voc 数据集的操作。

factory.py:这是一个工厂类，用类生成 imdb 类并返回数据集

库供网络训练和测试使用

首先修改 pascal_voc.py，模仿这个类，我写了一个自己的类 text2.py。

- 1.修改了相应的类名为 text2。
- 2.将所有的 voc_改成了 text2_,因为我要用是/data/VOCdevkit/的 text2 数据。
- 3.将数据集 data_path 改为 text2 的路径，将默认路径的中的 self.year 直接改成 2007（文件夹名）。
- 4.将 classes 改为 2 类，背景和 text。
- 5.将判断 self.year 是否为 2007 的改为是否为 2003.
- 6 将相应的文件路径由 voc_+self.year 改为 text2.
- 7.将主函数的导入模块改成 text2，然后创建一个 text2 对象，数据集有 trainval2007 改为 trainval2003.

接下就是 factory.py 的修改

- 1.首先导入 dataset.text2 的 text2 模块。
- 2.创建字典 text2_split_2003 用来生成所用数据的名称，从而生成相应的 imdb 类。而 imdb.py 是进行数据读写类的基类，里面封装了许多 db 的基类，不需要做什么修改。

接下里就是网络文件的修改了

- 1.由于我们是文本检测，所以我们将我们的 data 层的 num_classes 由 21 改成了 2，text+背景。
- 2.把 cls_score 层把 num_classes 由 21 变成了 2
- 3.把 bbox_pred 层的 num_output 从原来的 84 改成 8，种类个数乘以 4.
- 4.超参数的话在 solver.prototxt 文件中修改。

最后一步就是下载预训练模型 imagenet_models.脚本上给的网站好像下不了，正好电脑上有个 RFCN 的工程 you 相应模型我就拿来使用了，我使用了里面的 resnet101model.caffemodel 来初始化训练网络。

准备工作做好了，开始训练了。faster-rcnn 有两种训练方式，一种是 faster_rcnn_alt_opt.sh 一种是 faster_rcnn_end2end.sh.为了便于调试我选择了端到端的训练方式。