**实训周报**

15331416 赵寒旭

**1. 学习内容**

**1）论文阅读**

（1）论文名字

Multi-label Image Recognition by Recurrently Discovering Attentional Regions

（2）论文内容简述

论文提出了解决**多标签图片识别**问题的新模型。

模型结构：CNN+ST+LSTM

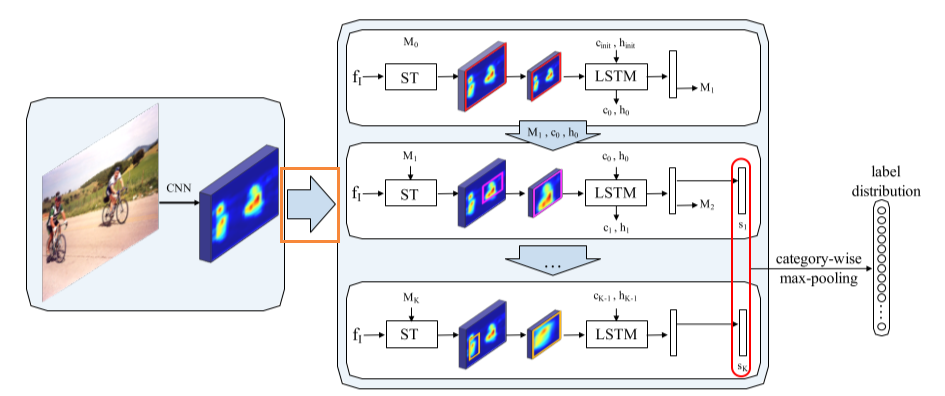
① 图片经过深度卷积神经网络（VGG-16）提取最后一层特征的响应图。

② ST在每次循环中定位一个Attention Region

③ LSTM预测该区域的类别分数，同时更新ST的参数

④ 结合所有循环产生的类别分数，得到最终的类别分布

**2）cifar数据集上ten crop实验**



本周任务是在AlexNet网络下加入tencrop机制，在cifar数据集上进行实验，观察其对模型性能是否有提升。

因为cifar数据集图片尺寸很小，仅32\*32，取最后一层卷积层的输出不足以进行ten crop。

目前我对卷积网络做了一定修改，使得最后一层卷积层输出feature map为256\*5\*5。

个人理解：我现在所做的ten crop部分即结构中橙色框出的箭头部分。

拟将AlexNet最后一层输出的feature map剪切翻转为10个256\*3\*3的feature map，依次放入一个只有全连接层的网络FullNet输出分类分数，将得到的10个向量做平均，得到原256\*5\*5 feature map最终的分类分数outputs。

这里的FullNet暂代了ST和LSTM的分类功能。

反向传播时，计算outputs与原图targets对应的loss，用于更新FullNet网络参数。

**2. 遇到的困难以及解决方案，或者学习收获**

**1）遇到的困难以及解决方案**

（1）预训练的卷积网络

我本来是想先训练一个新的最后输出256\*5\*5的完整网络，保存参数用于输出feature map。

因为这个网络没有输出分类结果，无法做反向传播更新网络参数，它应该时一个已经训练过的，并且在最后的网络中仅作为生成feature map部分不参与反向传播。

但是分给我们组的2，3号卡当时看的时候都被占满了，跑不起来，就没有训练。

（2）tencrop

先把网络输出的Variable转成numpy，剪切矩阵并做翻转，把十个结果作为FullNet的输入，最后把网络的十个输出结果取平均计算loss，反向传播只传回FullNet。

**2）学习收获**

基本了解了论文算法。

大致确定任务做法。

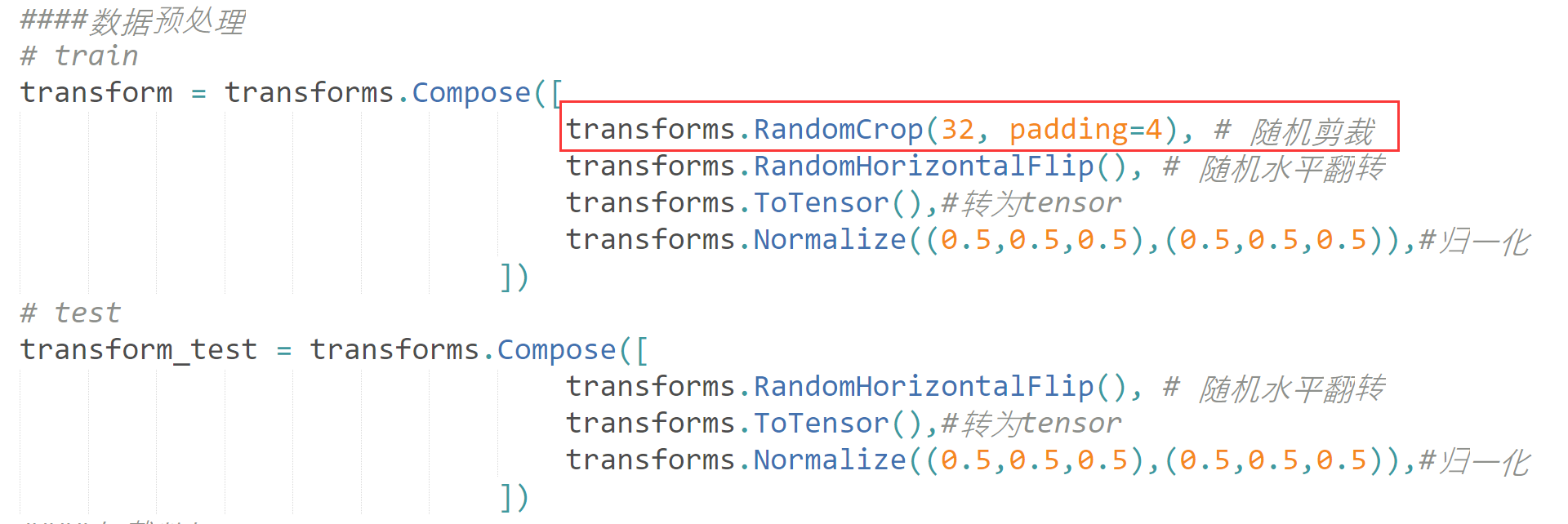
**3）下周计划**

实现cifar数据集上的tencrop操作。

不知道现在的想法是否可行，但目标是先按当前预想完成可运行的代码。

**备注：**

上次提到的关于测试集数据预处理采用和训练集不同方法的建议，初步实验之后发现的确提高了模型准确率。



测试集取消了随机剪裁。

TOP1最高准确率：78.65

之前训练集和测试集用相同预处理方式：

最好TOP1结果：78.2