**人工智能导论大作业实验报告**

范成城 2020201680

1. 实验内容

图片文字内容的简单识别提取

1. 模型背景知识介绍
2. 文字识别模型所必须要有的要素：

首先是要读取输入的图像，提取图像特征，因此，需要有个卷积层用于读取图像和提取特征。由于文本序列是不定长的，因此在模型中需要引入 RNN（循环神经网络），一般是使用双向 LSTM 来处理不定长序列预测的问题。为了提升模型的适用性，最好不要要求对输入字符进行分割，直接可进行端到端的训练，这样可减少大量的分割标注工作，这时就要引入 CTC 模型（Connectionist temporal classification， 联接时间分类），来解决样本的分割对齐的问题。最后根据一定的规则，对模型输出结果进行纠正处理，输出正确结果。

1. 使用模型：CRNN模型
2. 模型特点：

1.可以进行端到端的训练

2.不需要对样本数据进行字符分割，可识别任意长度的文本列

3.模型速度快、性能好，并且模型很小（参数少）

1. 模型结构：

1.卷积层：从输入图像中提取出特征序列；

2.循环层：预测从卷积层获取的特征序列的标签分布；

3.转录层：把从循环层获取的标签分布通过去重、整合等操作转换成最终的识别结果。

（5） 卷积层介绍：

① 预处理：CRNN 对输入图像先做了缩放处理，把所有输入图像缩放到相同高度，默认是 32，宽度可任意长。

② 卷积运算

由标准的 CNN 模型中的卷积层和最大池化层组成，结构类似于 VGG，如下图：

从上图可以看出，卷积层是由一系列的卷积、最大池化、批量归一化等操作组成。

③ 提取序列特征：提取的特征序列中的向量是在特征图上从左到右按照顺序生成的，用于作为循环层的输入，每个特征向量表示了图像上一定宽度上的特征，默认的宽度是 1，也就是单个像素。由于 CRNN 已将输入图像缩放到同样高度了，因此只需按照一定的宽度提取特征即可。

（6）循环层介绍：

循环层由一个双向 LSTM 循环神经网络构成，预测特征序列中的每一个特征向量的标签分布。由于 LSTM 需要有个时间维度，在本模型中把序列的 width 当作 LSTM 的时间 time steps。其中，“Map-to-Sequence” 自定义网络层主要是做循环层误差反馈，与特征序列的转换，作为卷积层和循环层之间连接的桥梁，从而将误差从循环层反馈到卷积层。

（7）转录层介绍：

转录层是将 LSTM 网络预测的特征序列的结果进行整合，转换为最终输出的结果。在 CRNN 模型中双向 LSTM 网络层的最后连接上一个 CTC 模型，从而做到了端对端的识别。所谓 CTC 模型（Connectionist Temporal Classification，联接时间分类），主要用于解决输入数据与给定标签的对齐问题，可用于执行端到端的训练，输出不定长的序列结果。

由于输入的自然场景的文字图像，由于字符间隔、图像变形等问题，导致同个文字有不同的表现形式，但实际上都是同一个词，而引入 CTC 就是主要解决这个问题，通过 CTC 模型训练后，对结果中去掉间隔字符、去掉重复字符（如果同个字符连续出现，则表示只有 1 个字符，如果中间有间隔字符，则表示该字符出现多次）。

以上就是文本识别模型 CRNN 的介绍，该模型既可用于识别英文、数字，也可用于识别中文。一般是跟 CTPN 结合一起使用，使用 CTPN 进行文字的检测，使用 CRNN 进行文字的识别。

1. 模型搭建和训练
2. 数据集生成：

先要准备好基础数据，可根据需要准备好已标注的文本图片集。在这里，为了方便训练和测试模型，随机生成 10000 张不定长的图片数据集。通过使用 Pillow 生成图片和绘上文字，并对图片随机叠加椒盐噪声，以更加贴近现实场景。

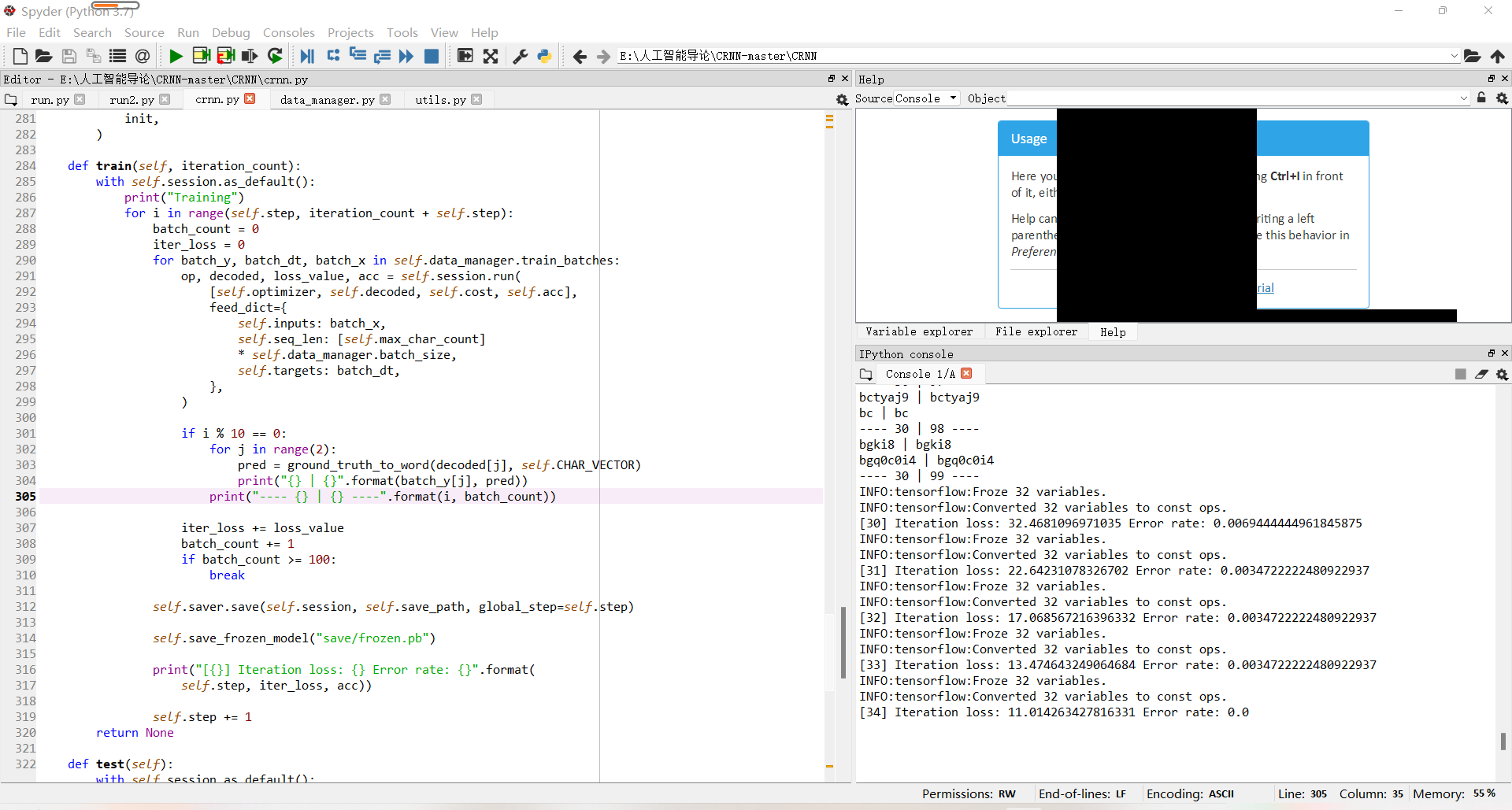


1. 模型训练：



里面参数包含了数据集文件、模型保存文件，此次实现的CRNN模型可以保存到本地，以供后续使用。倒数第三个参数是数据字典，识别的文字就从里面找，倒数第一个参数是待识别的文字类型，默认为‘en‘。

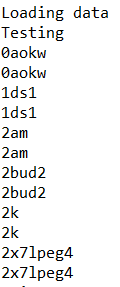
1. 结果展示：



可见，训练了35轮后，针对测试集的错误率已经降到了0。此时将模型保存到本地。

1. 模型测试效果

使用生成训练数据集的方式又生成了100个测试数据，对于较短的数据，识别效果不错，如下图所示：



（上方为正确数据，下方为识别出来的数据，上下两个成一组）

但是遇到较复杂数据，或者稍长数据，有些时候可能会出现错误，比如以下的情况：

识别漏掉前面两个字符

 识别错误前面的字符。

推测是过拟合，或者说是训练轮数不够，仅在训练集上有用。

1. 总结

本次作业是对人工智能入门技术的一次实践，受限于水平，只按照相关教程写出了初级版本的模型，并没有进行代码上的优化。希望在以后的学习生涯中，能够继续钻研人工智能相关知识，做出更好的代码。