**实验主题:医学ANA图像分类**

**1.实验目的:**

图像分类，顾名思义，是一个输入图像，输出对该图像内容分类的描述的问题。医学图像常常面临数据量少，标注困难的情况。本实验需要在较少量且不平衡的数据中设计一个有效的神经网络达到图像训练分类的目的。

**2.实验环境与数据准备:**

2.1实验环境

2.1.1 硬件环境

PC机一台

2.1.2 软件环境

Anaconda3+Tensorflow-gpu1.4+Keras2.1（采用清华镜像网安装）

2.2数据准备

2.2.1 数据下载

数据地址：<https://pan.baidu.com/s/1cqdOIofxYT9zjbxAj_ZKTQ>

提取码：y79m

2.2.2 数据介绍

VGHTC数据集是专业人员用彩色数码相机采集的HEp-2细胞基质图像，其血清稀释度为1:80。对于细胞基质图像，专业人员利用荧光显微镜观察样本切片时，由于样本很小，因此以640倍的放大率进行观察并采集，最终采集的图像大小为3126×2352。该数据集中一共包含有260张样本图像，对应四种不同的染色模式，即：粗斑点型（Coarse speckled, CS）样本、细斑点型（Fine speckled, FS）样本、核仁型（Nucleolar, NU）样本和周边型（Peripheral, PE）样本。并且，数据集中，167张样本图像是粗斑点型的，2020型的，38张样本图像是核仁型的，35张样本图像是周边型的。



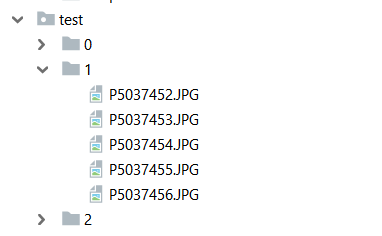
**3.实验操作步骤(占60%):(1小时)**

3.1数据预处理

首先要将数据5,5等分为训练集和测试集。可分别在两个文本文件记录训练集以及测试集的图片路径、名称、类别标签（可以将类别标签与数字对应）等信息。在preprocess.py实现。记录格式和all\_list.csv相同

测试图片集放置格式

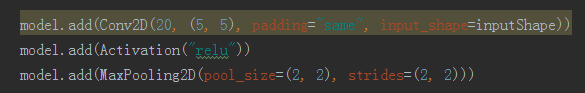


3.2搭建网络

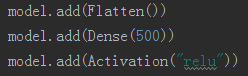
1.初始化模型



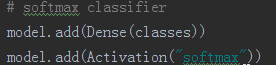
2.接下来设置卷积层、激活函数、池化层等



3.构建全连接网络层



4.softmax 分类器



Vgg\_m.py是提供的参考网络，自行学习后可修改。

3.3训练网络

1.定义超参数

我们需要为训练设置一些参数，比如训练的epoches，batch\_szie等。这些参数不是随便设的，比如batch\_size的数值取决于你电脑内存的大小，内存越大，batch\_size就可以设为大一点。又比如norm\_size（图片归一化尺寸）是根据你得到的数据集，经过分析后得出的。

2.导入数据以及标签，按8:2分为训练集和验证集

接下来我们需要读入图片，归一化图片至恰当的尺寸以及对应的标签信息。

3.训练网络

在正式训练之前可以使用ImageDataGenerator来对我们的小数据集进行数据增强（对数据集图像进行随机旋转、移动、翻转、剪切等），以加强模型的泛化能力。

然后初始化已搭建好的神经网络，选择优化器优化参数，损失函数，性能评估标准后，就可以使用训练数据开始训练了。训练50个epochs大约只需要几分钟。训练完成后保存最佳模型。（运行train.py时根据main函数注释赋值参数）

3.4评估结果

最后是对测试数据进行预测，并评估结果。这里得到模型最终的损失值和准确率。加载上一步骤的模型以及权重，以相同的方式导入测试数据并归一化图片尺寸用网络预测类别得到模型评估结果。

**4.修改实验题目(占40%):（1小时）**

可以修改不同的网络结构、调整不同的网络参数或者尝试使用别的任务已训练好的网络用迁移学习等不同方法提高实验结果。

**5.修改实验解答步骤:**

1.可以添加删减网络层数、卷积核数目、卷积核大小等等。

2.调整不同的初始化网络参数等

3.使用预定义的网络以及参数，根据自己的需要冻结一些层，再使用本实验数据重新训练模型调整权重。

**6.参考文献**

1.Keras中文文档<https://keras.io/zh/>

2. Jiaqian L , Kuo-Kun T , Yi H Z , et al. Staining Pattern Classification of Antinuclear Autoantibodies Based on Block Segmentation in Indirect Immunofluorescence Images[J]. PLoS ONE, 2014, 9(12):e113132-.