体系结构设计

本项目采用面向对象设计，在本项目中将会主要介绍项目任务的模块划分并确定类与类之间的逻辑关系。

总体设计

（系统面向对象，使用B/S结构）

总体类图

用户界面层

控制层



业务逻辑层

数据持久层

类图说明

类图综述

用户界面层

模型性能查看界面类功能是当系统管理员点击按钮来查看模型性能时显示与模型性能相关的系统界面，其与查看模型分类测试准确度类、查看模型分类测试耗时类、查看模型分类训练速度类，选择测试数据集类以及选择测试模型类之间都是组合关系，即页面上应分别有实现这几个类的功能的按钮。

查看模型分类测试准确度类实现按下查看模型分类测试准确度按钮的后续操作，其他类功能相似。

各个查看不同性能的类还实现显示模型性能接口。

控制层

查看模型分类测试准确度查看控制类，查看模型分类测试耗时查看控制类和查看模型训练速度查看控制类与模型性能查看控制类的关系是继承关系。

模型性能查看控制可设定模型性能控制参数，三个子类利用父类中设置好的参数进行不同性能参数查看的后续控制。

业务逻辑层

模型性能类与模型性能获取类是组合关系，与模型性能参数保存类是聚合关系，其还与测试模型类和依照误差要求训练模型类有关联关系。

而模型性能获取类与获得模型分类测试耗时类、获得模型分类测试准确度类和获得模型训练速度类有继承关系。

获得模型分类测试耗时类和获得模型分类测试准确度类与测试模型存在依赖关系，获得模型训练速度类与训练模型类存在依赖关系

预处理类与数据集预处理类以及单个图片预处理类存在组合关系，后两者与图像规范化类存在依赖关系。

数据持久层

模型导入类与数据集导入类实现文件导入接口，模型性能表类与保存模型性能参数类以及读取模型性能参数类存在组合关系。

关系描述

···

|  |  |  |  |
| --- | --- | --- | --- |
|  | 机器学习 | 用户交互 | 数据库操作 |
| 预处理 | √ |  |  |
| 查看模型性能 | √ | √ | √ |

包图

类设计

|  |  |  |
| --- | --- | --- |
| 类 | 属性 | 操作 |
| 数据预处理 | 经过预处理的图片数据集  规范统一图片像素规格 | 提取图片信息，舌图提取，舌体分割，周围像素合并，单个像素分割成为多个像素，并行预处理 |
| 查看模型性能 | 模型训练误差要求  测试集准确率  测试集分类时间  模型训练达到要求误差迭代代数 | 查看模型训练达到误差要求速度，查看模型对测试数据集的准确率，查看模型对测试集的测试时间，修改模型训练误差要求的值 |

类交互设计

数据预处理

首先确定图片数量，若图片数量大于1，可通过并行处理加快预处理速度，提高模块运行效率。

接下来预处理模块可以提取图片像素相关信息。

对输入的舌图进行自动识别, 分离舌图的有效部分, 一般采用基于边缘检测方法完成从背景中提取舌像。

判断此时图片的像素大小大于规范图片的像素大小，对图片进行像素合并操作，即将周围几个像素进行合并操作，合并成一个像素，如果几个像素的颜色有差别那么就取这几个像素的平均值。若此时图片的像素大小小于规范图片的像素大小，那么应该对图片进行像素裁切操作，即按照需求，将单个像素裁切为多个像素。最后将处理完成的图片保存。



查看模型性能

系统管理员通过点击系统界面上的不同按钮来确定查看机器学习模块的不同性能参数。

当选择查看模型的训练速度时，模型训练将以系统管理员设定或系统预设的误差要求作为测试参数来判断其训练速度。随着训练的不断进行，模型不断优化迭代，误差不断减小，当达到或小于误差要求时记录此时的迭代代数，将其作为模型的训练速度的参考标准，但是如果在我们所限制的迭代代数下，无法达到要求的误差，那么系统将会返回系统管理员一个消息，提醒其在限制的迭代代数下，无法达到误差要求，但是其可以通过修改误差要求来尝试得到模型训练的速度的参考标准。

查看模型对测试集进行分类的性能时，系统管理员应先选择测试数据集，并将数据集输入预处理模块，对测试集中带标记的图片进行数据预处理得到规范统一图片格式的数据集合。

将处理好了的图片输入模型，模型对输入的片进行分类。

分类结果与图片的标记进行比对，如果相同则分类正确，如果出现差异的则分类错误。记录整个集合的准确度作为测试集的分类准确率，同时对整个分类过程计时，整个分类过程所用的时间作为测试集的分类耗时。

最后结果保存在模型使用的类中。当系统管理员需要查看时和通过点击界面上的按钮查看训练集准确度和训练级分类耗时。



数据库库设计

模型性能属性组

|  |  |  |  |
| --- | --- | --- | --- |
| 实体名 | 主键 | 属性 | 外键 |
| 模型性能参数 | 模块ID，测试数据集ID | 模块ID，测试数据集ID，模型训练的误差要求，模型训练达到误差要求的迭代次数，测试集的测试分类准确率，测试集测试分类耗时 | 无 |
| 机器学习模型 | 模型ID | 模型ID，模型建立时间，模型路径 | 模型ID |
| 数据集 | 数据集ID | 数据集ID，数据集路径 | 数据集ID |



模型性能参数记录由模型ID和测试数据集ID两个属性构成的属性组唯一标识。其他属性说明模块性能。机器学习模型实体记录由模型ID唯一标识，数据集实体记录由数据集ID唯一标识，同时，模型ID与数据集ID作为两个记录的外键与模型性能参数记录存在关系。