# 判断色情图片

开发者：杨宁 指导教师：仲宝才

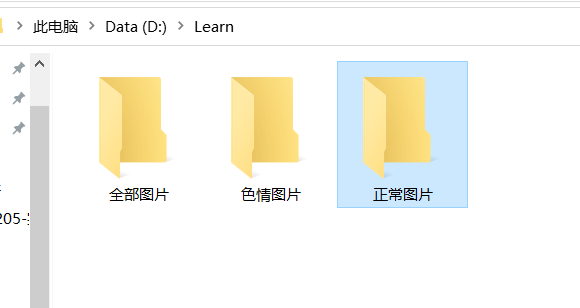
1. 项目概述

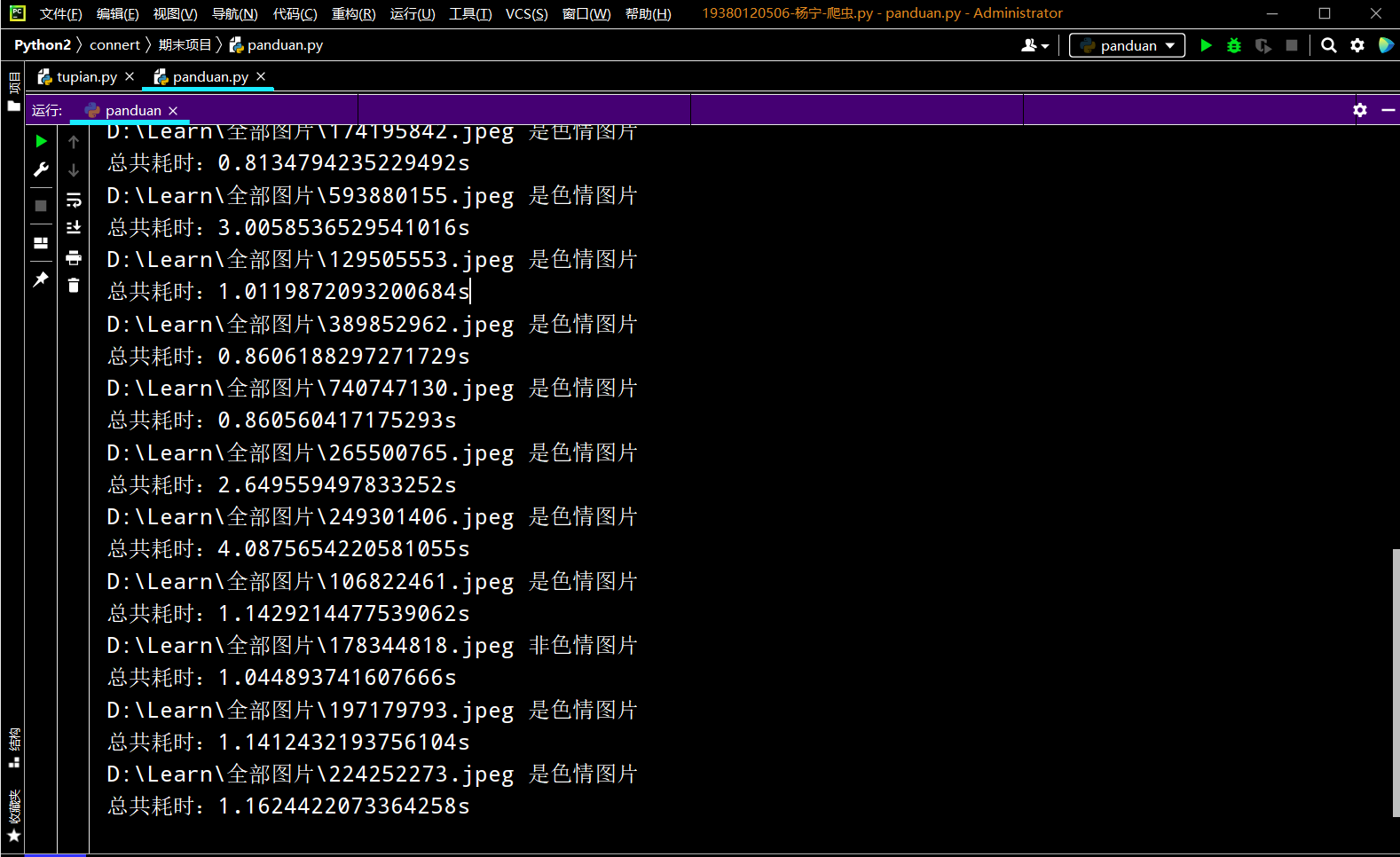
随着社会发展，网络上的一些甄别工作还是需要人工进行，例如鉴黄师。鉴黄师的工作很简单就是判断图片以及视频是否色情图片，但这项工作也给这些工作人员带来一些不便，所以有必要有一个程序来帮助他们。

根据三种判断颜色原理，利用把图片的像素原理判断图片是否是色情图片，可以十分有效的用于鉴黄师等工作，利用Python爬虫爬取网站上的图片，进行分类爬取的图片，在未来应用可以十分广泛，可以有效减轻一些工作人员。

2.项目展示

### 2.1 启动后





### 

### 2.2 项目思路

知识点：

肤色像素检测与皮肤区域划分算法

文件的打开（用的新语法）

异常处理

爬虫知识

判断是否色情：

原理：根据颜色(肤色)找出图片的皮肤区域，然后根据判断条件判断是否为色情图片。

程序执行步骤：

遍历每个像素，检测像素颜色是否为肤色  
将相邻的肤色像素归为一个皮肤区域，得到若干个皮肤区域  
剔除像素数量极少的皮肤区域

判断标准：

皮肤区域小于三个

皮肤区域的像素与图像所有像素的比值小于15%

最大皮肤区域小于总皮肤面积的45%

皮肤区域数量超过60个  
注意：

颜色标准：RGB颜色模式、HSV颜色模式、YCbCr颜色模式

创建Skin类:

用来表示像素以及像素信息(唯一的 编号（id），是/否肤色（skin），皮肤区域号（region），横坐标（x），纵坐标（y）)

Skin类的相关方法：

collections.namedtuple() 函数实际上是一个返回Python中标准元组类型子类的一个工厂方法。给相应的字段返回类，可以初始化这个类，为定义的字段传递值。

isinstane(object, classinfo) 如果参数 object 是参数 classinfo 的实例，返回真，否则假；参数 classinfo可以是一个包含若干 type 对象的元祖，如果参数 object 是其中任意一个类型的实例，返回真，否则假。

resize():图片缩小方法。基于最大宽高按比例重设图片大小，注意：这可能影响检测算法的结果。 如果没有变化返回 0，原宽度大于 maxwidth 返回 1， 原高度大于 maxheight 返回 2，原宽高大于 maxwidth, maxheight 返回 3。 maxwidth - 图片最大宽度，maxheight - 图片最大高度，传递参数时都可以设置为 False 来忽略。

parse():解析方法，region:记录相邻像素中肤色像素所在的区域号，初始化为 -1。

self.\_add\_merge() 这个方法接收两个区域号，它将会把两个区域号添加到 self.merge\_regions 中的元素中，self.merge\_regions 的每一个元素都是一个列表，这些列表中存放了 1 到多个的区域号，区域号代表的区域是连通的，需要合并。

somenamedtuple.\_replace(kwargs) 返回一个替换指定字段的值为参数的 namedtuple 实例。\_merge() 便是用来合并这些连通的皮肤区域的\_analyse\_regions()，运用之前在程序原理一节定义的非色情图像判定规则，从而得到判定结果

爬虫：

爬取百度—美女图片，根据data数据进行查找，再根据上述的内容进行判断，同时进行计时（

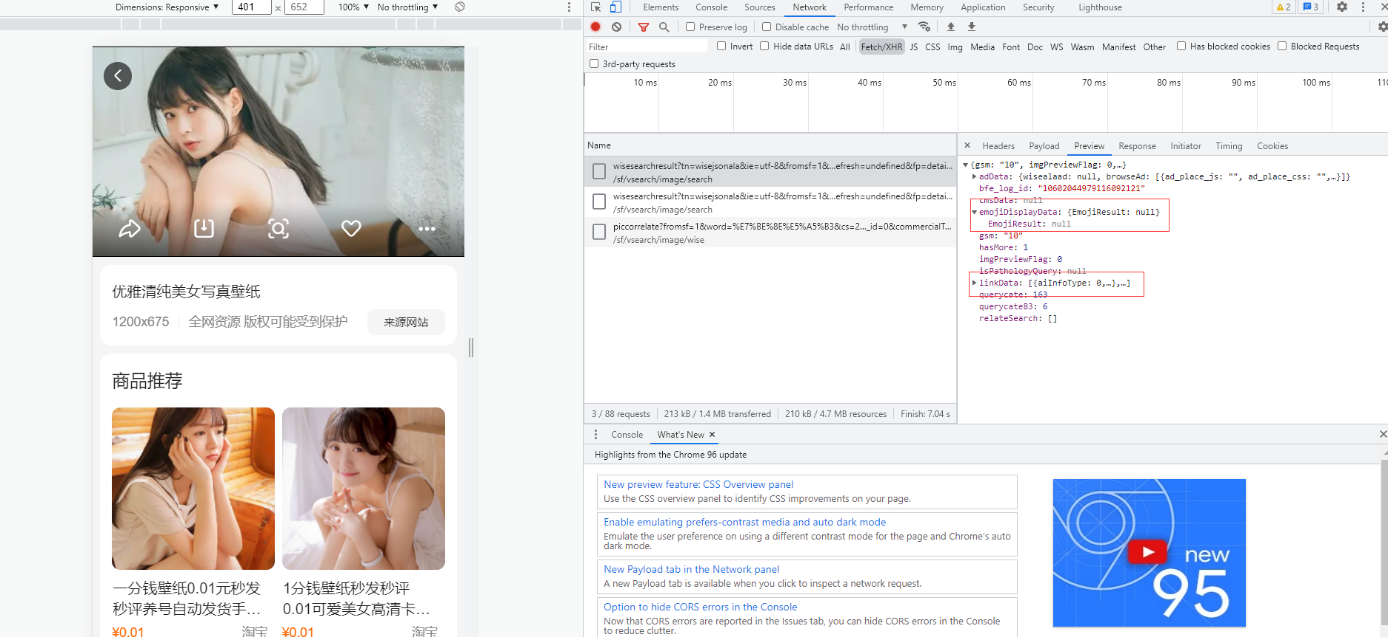
time1 = time.time()

…..

time2 = time.time()  
print('总共耗时：' + str(time2 - time1) + 's')

）

### 2.3查询的页面



3.总结

了解了色情图片检测的原理，整个实验难点是在皮肤区域的检测与整合这一方面，本实验还有许多可以改进的地方，比如肤色检测的公式，色情判定条件。根据网页的美女图片进行判断和保存，利用爬虫技术进行处理，这方面可以改进的是，设置一些常用网站进行判断，诸如360、搜狗等等。 如果技术允许可以在网上进行判断，上传到网上运用于工作中，诸如鉴黄师等等。