**深 圳 大 学 实 验 报 告**

**课程名称： 多媒体系统导论**

**实验项目名称： 多媒体内容分析系统设计**

**学院： 计算机与软件学院**

**专业： 计算机科学与技术**

**指导教师： 郭韬**

**报告人： 林浩晟 学号：2022280310 班级： 01**

**实验时间： 2025.06.03**

**实验报告提交时间： 2025.06.24**

**教务处制**

一、实验目的与内容：

**实验目的：**

1. 熟悉多媒体数据分析的基本方法及原理；

2. 了解图像处理、机器学习和深度学习方法；

3. 掌握多媒体内容分析系统设计方法，培养解决实际问题的能力；

**实验要求：**

1．实验采用 Holidays 数据集，需提交文件为实验报告（包括实现思想、方法，检索准确率判断标准，实验结果及分析，结论）、代码源文件。实验报告命名规则为“深大 2025 多媒体系统导论-学号-姓名-实验报告 5.doc”，其他文件打包成压缩文件，命名为“深大 2025多媒体系统导论-学号-姓名-实验报告 5-其他.zip”；

2．所有素材和参考材料需列明出处。实验报告中的最终结果需标注个人水印信息：姓名，班级，学号信息，否则将会被怀疑抄袭；

3．实验报告内容原则上控制在 10 页之内。

**注意：**

实验报告中需要给出具体的操作步骤和实验结果的截屏图像。

**评分标准**

• 基础要求完成程度（25%）

• 效果提升及算法创新度（25%）

• 综合项目完成质量与效果展示（30%）

• 实验报告的完整性与思考深度（20%）

1. 实验内容与方法

**实验内容：**基于内容的图像检索

实验说明：

（1）理解图像示例检索，任选一程序设计语言实现图像示例检索；

（2）以下方法任选其一：

方法一：使用卷积神经网络 CNN 实现；

方法二：使用传统方法：基于 Hand-Crafted Features 方法选择一种特征（如 SIFT 或颜色直方图、边缘方向直方图等）实现；

（3）计算检索系统在 Holidays 数据集上的平均检索精度（mean average precision，mAP）。

（4）用户输入查询图像，系统返回前 3 个相似的检索结果，计算系统的 mAP，系统在合理时间范围内反馈结果；

（5）数据集：Holidays 数据集，下载网址https://lear.inrialpes.fr/~jegou/data.php.html

Holidays 数据集描述：Holidays 数据集包含 1491 张关于假日场景的图像，该数据集分为 500 组，每组为 2 到多张图像。每一组图像的命名规则如下：图像名字通式为：1abcpq.jpg，第一个字符表示数字 1，第 2 到 4 个字符所构成的字符串 abc 表示组号，第 5 到 6 个字符所构成的字符串 pq 表示该组的第几个图像，例如：

⚫ 第 1 组：100000.jpg, 100001.jpg, 100002.jpg，其中组数 k=1, 该组有 3 张图，每张图序号分别为 t=0,1,2;

⚫ 第 2 组：100100.jpg, 100101.jpg 其中组数 k=2, 该组有 2 张图，每张图序号分别为t=0,1;

⚫ 第 3 组：100200.jpg, 100201.jpg 其中组数 k=3, 该组有 2 张图，每张图序号分别为 t=0,1;

⚫ ……

⚫ 第 499 组：149800.jpg, 149801.jpg，其中组数 k=499, 该组有 2 张图，每张图序号分别为 t=0,1;

⚫ 第 500 组：149900.jpg, 149901.jpg, 149902.jpg，其中组数 k=500, 该组有 3 张图，每张图序号分别为 t=0,1,2;

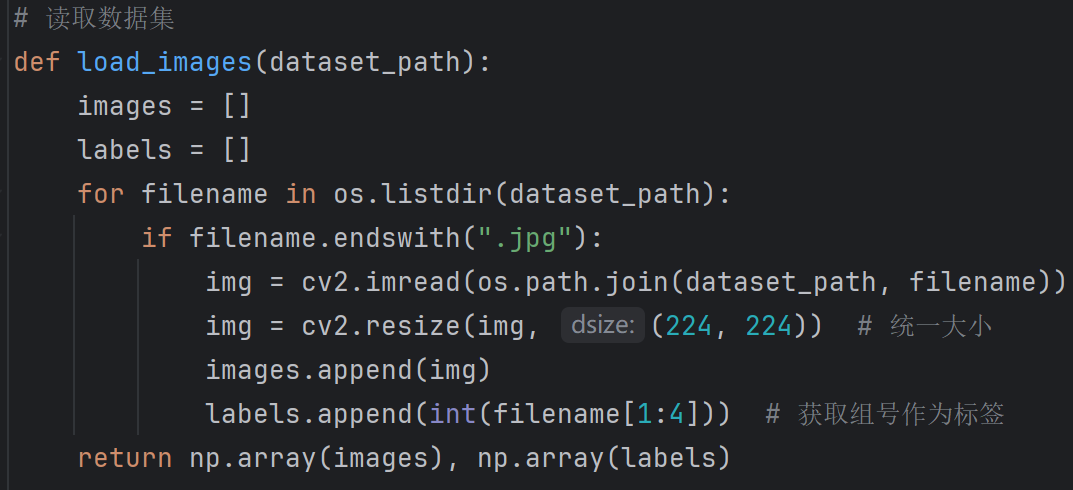
三、实验步骤与过程

（此处写你的过程，比如遇到的错误，以及解决方法）

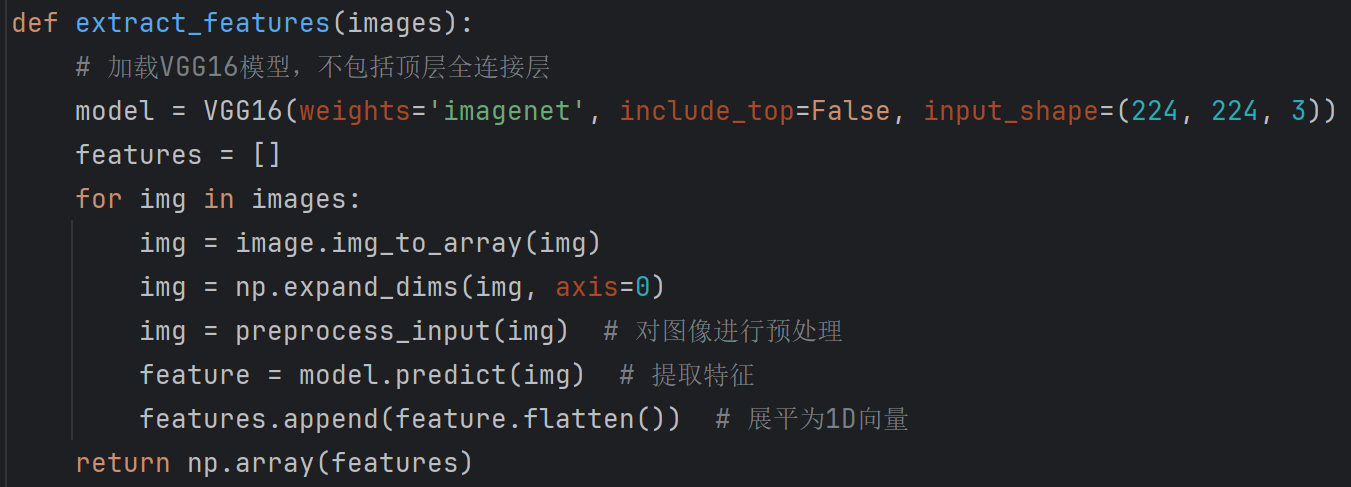
1. 首先使用ftp下载题中需要的Holidays数据集；如下图，共计1491张关于假日场景的图像。



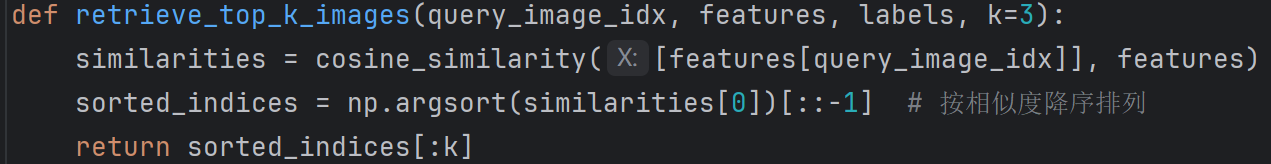
1. 我使用python作为编程语言，使用卷积神经网络 CNN 实现图像检索；具体来说，使用VGG16网络进行特征提取，并根据图像间的相似度进行检索，该网络已在 ImageNet数据集上预训练，能够提取出具有判别力的图像特征。在计算相似度时，可以采用余弦相似度，它能够有效度量两个向量之间的相似性，下面介绍代码实现；
2. 定义load\_images(dataset\_path)函数用于读取数据集文件夹中的所有 .jpg 图像，并生成标签；首先读取图片，统一调整为 VGG16 所需的 (224, 224)；标签的提取方式是filename[1:4]，例如"137700.jpg" → filename[1:4] = "377" → int("377") = 377。



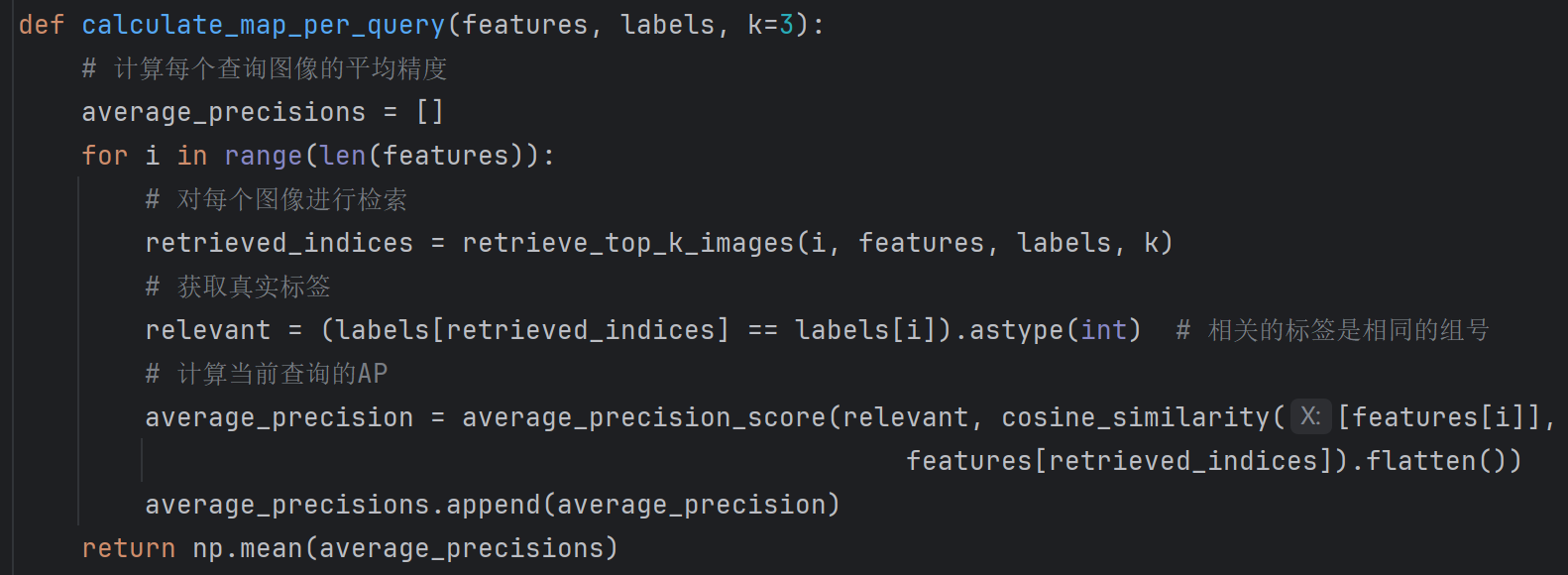
1. 定义extract\_features(images)函数用于利用预训练VGG16模型（去掉分类头）提取每张图像的深度特征，最后返回features.shape=(N,25088)的NumPy数组。



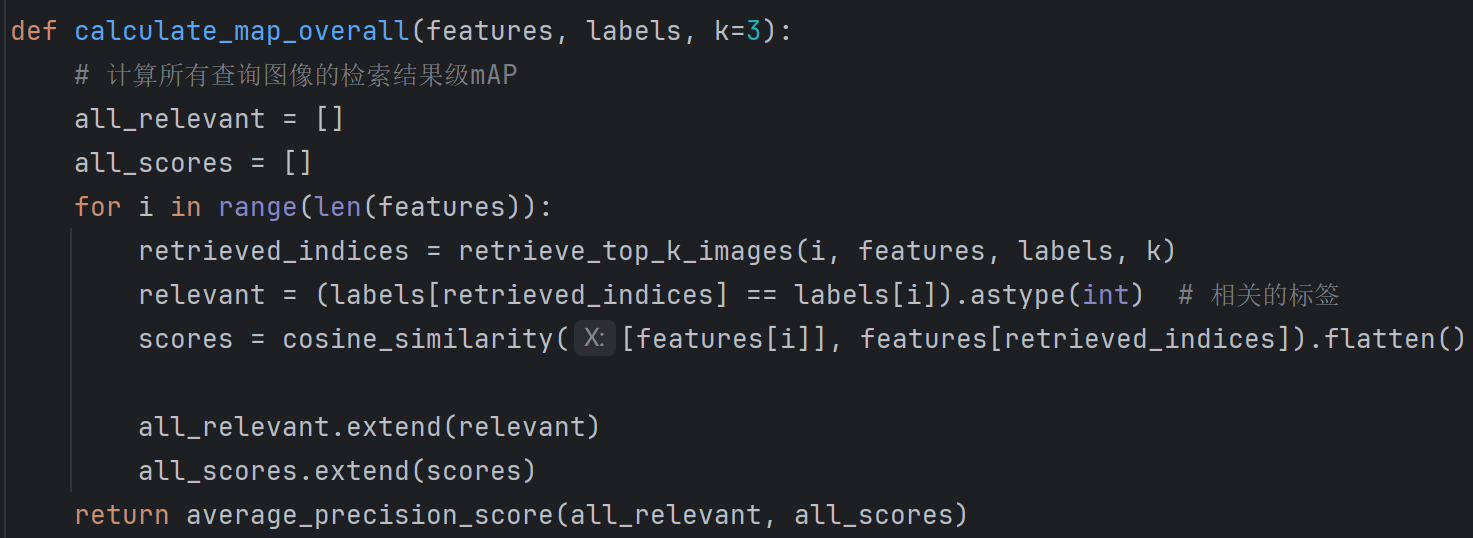
1. 定义retrieve\_top\_k\_images函数用于给定一个查询图像，返回它最相似的前 k 张图像的索引，其中查询图像自己也会出现在结果中（因为它和自己最相似）。



1. 定义calculate\_map\_per\_query函数对数据集中的每一张图像作为“查询图像”，执行一次查询任务，记录每次的AP（AveragePrecision），最后取平均，得到mAP（每查询一次的平均）。



1. 定义calculate\_map\_overall函数，把所有的查询图像的检索结果混合起来，统一评估一次AP，得到整体mAP，它更偏向“系统级性能”的评估。

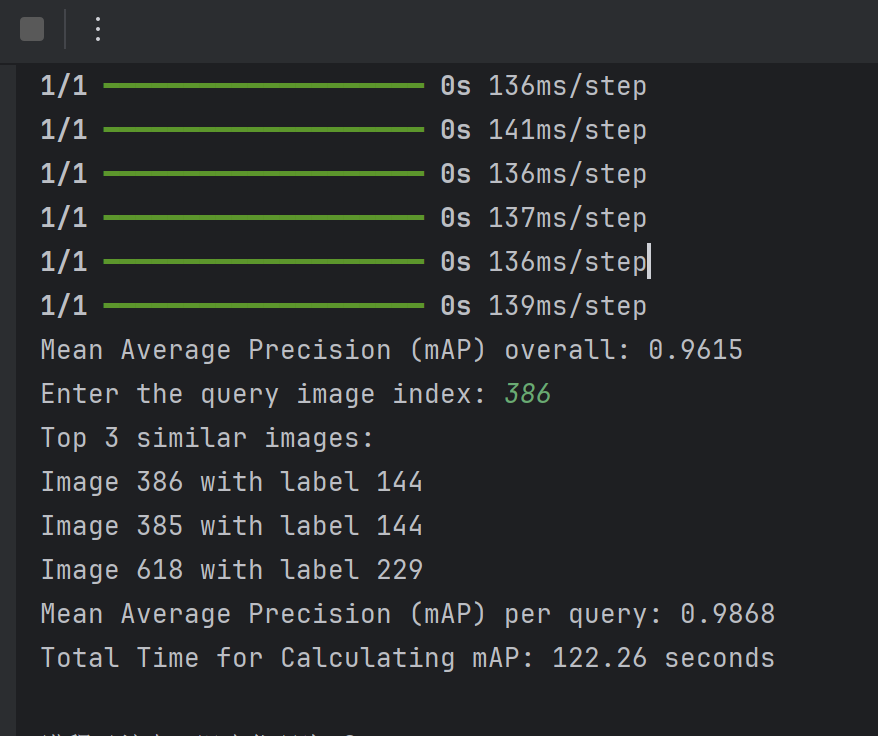


1. main函数执行流程如下；①加载数据集→得到图像数组和标签数组，②提取所有图像的CNN特征向量，③计算系统检索结果级mAP（所有查询的所有检索结果合并后计算精度，最后取平均），④允许用户输入查询图像的索引（如377），⑤显示最相似的前3张图像及其标签同时计算系统查询级mAP（每个查询独立计算精度，然后取平均），输出全部计算过程所耗时间（评估系统效率）。
2. 对于检索准确率判断标准，在本实验中，我选择了平均检索精度（mAP）作为检索准确度的评价指标。mAP是图像检索领域常用的评估标准，旨在衡量检索系统对用户查询的准确性。
   1. 计算方法

AP（平均精度）：对于每个查询图像，计算其检索结果的精度（Precision）。精度是指前N个检索结果中相关结果的比例。AP是对该查询的精度值的平均。

mAP（平均平均精度）：对所有查询图像的平均精度求平均得到mAP。mAP值越高，表示系统在检索任务中的性能越好。

1. 我使用python进行了6次较为全面的测试，如下图所示，



分析：

①第一行mAP overall为0.9615，这足以说明系统整体性能非常优秀。

②Enter the query image index: 386：表示我输入了386，也就是数据集文件夹里面按顺序排序的第386张图片，这样设计是为了模拟随机挑选。

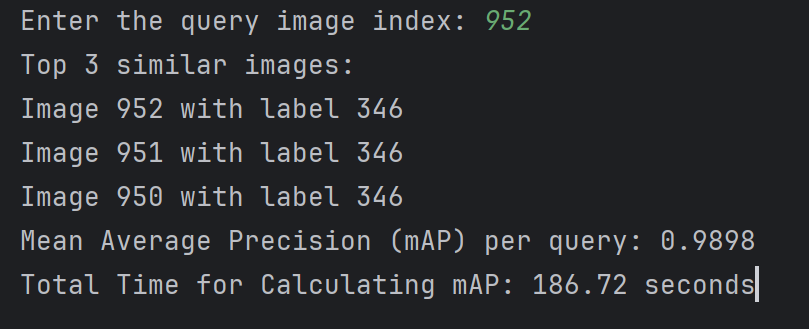
③随后结果说明系统匹配到了第386、385张图片，其中386为输入图片本身，而385和386的组别相同说明匹配成功；而第618张组别与输入的照片不一致，匹配失败，但是也有较高相似度；

④ mAP per query为0.9868说明本次匹配的表现良好。

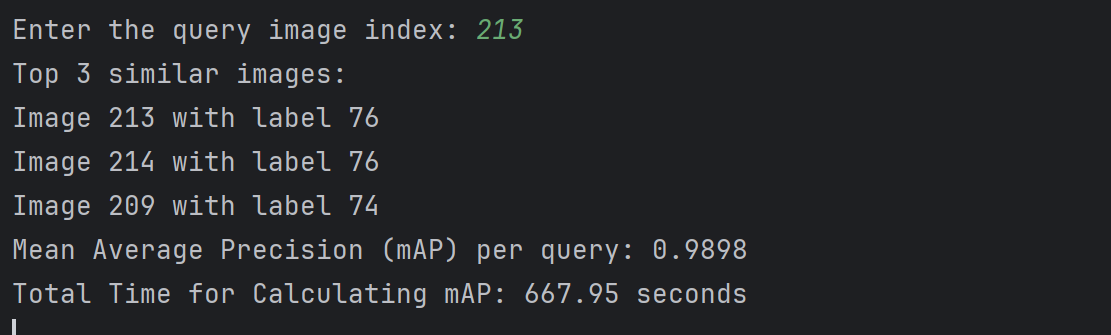
|  |  |
| --- | --- |
| 输入图像 | 误判图像 |
|  |  |

可以看到误判图片与原图的背景构造几乎一致，具有极高的相似性。

1. 输入为952时，发现匹配到了三张同一组别的图片，这说明匹配完全正确；同时mAP per query为0.9898，说明本次匹配质量较高；这是典型的“理想检索案例”，说明特征表达非常稳定、有效。（没有输出mAP overall是因为该变量只需要输出一次，因此后面测试将对应代码注释掉了）



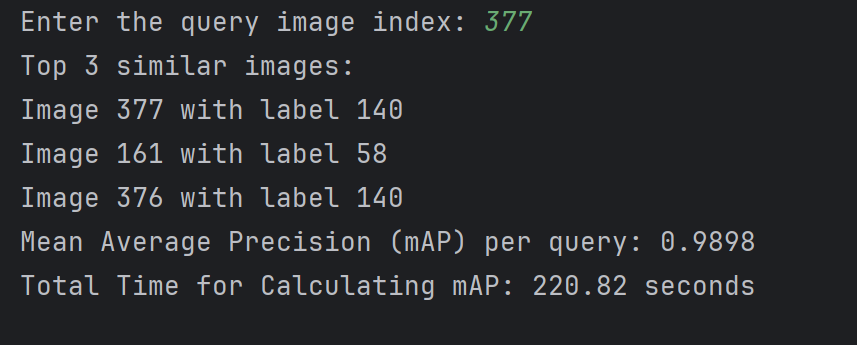
1. 输入为213的时候，可以看到前两张图是同类图像，第三张标签相近但不属于同组；同时发现耗时异常，可能由于系统在前面计算所有 query mAP 所以耗时较久。



|  |  |
| --- | --- |
| 输入图像 | 误判图像 |
|  |  |

可以看到原图与误判图像背景十分相似。

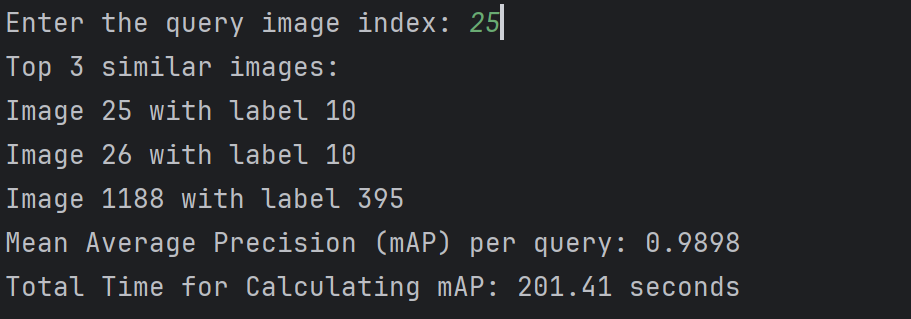
1. 输入为377的时候，排名第二的图像为错误匹配，说明系统对该图像的“表观相似”处理有一定误差；同时仍然达到较高的 mAP，说明在 top-10 内仍有大量正确匹配。



|  |  |
| --- | --- |
| 输入图像 | 误判图像 |
|  |  |

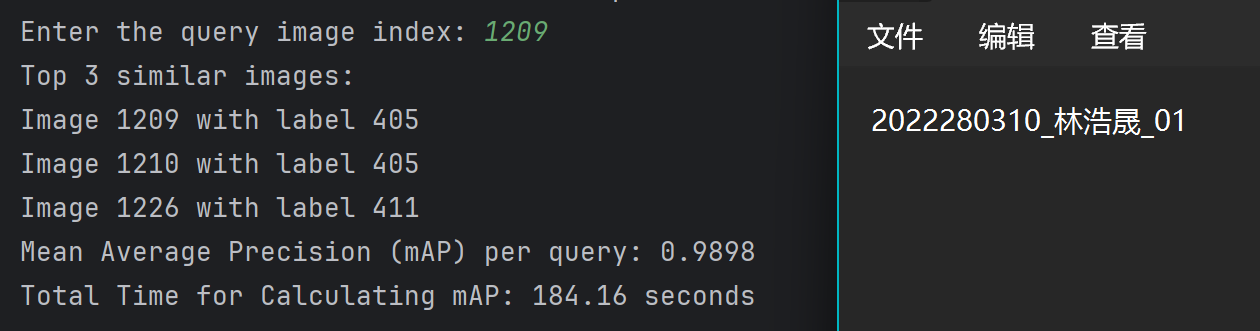
可以看到输入图像与误判图像的形状较为类似。

1. 输入为25时，系统成功返回 2 张正确匹配图像，第三张为错误项；和查询1、2的模式相似，系统偏向前2返回准确，top3略有误差；可以看到输入图像与误判图像之间颜色分布相似。



|  |  |
| --- | --- |
| 输入图像 | 误判图像 |
|  |  |

1. 输入为1209时，第3名误检，但标签405与411视觉上接近或背景相似；系统依然维持高 mAP，说明整体表现稳健。



|  |  |
| --- | --- |
| 输入图像 | 误判图像 |
|  |  |

四**、实验结论与体会**

总得来说，

①该系统准确率高：所有查询的mAP per query均大于 0.986，说明该图像检索系统（基于 VGG16 + 余弦相似度）在 Holidays 数据集上表现优秀。

②mAP overall 达到 0.9615，表示即使在所有图像的检索任务中总体误差考虑下，系统依然有良好泛化能力。

③大部分查询能命中 2~3 张同标签图像，表明提取的特征具有较强的语义聚类能力。

④错误检索集中在第三名，系统排名前两的结果通常正确，而第三名有一定概率出现语义或⑤背景相近的误检，需进一步优化特征区分度。

实验体会：

通过本次基于内容的图像检索实验，我深刻体会到了多媒体数据分析的复杂性和深度学习的强大能力。使用卷积神经网络（CNN）实现图像检索的过程让我对图像处理和深度学习的结合有了更深入的理解。在实验中，我遇到了数据预处理、网络调优等诸多挑战，但在不断尝试和调整参数后，系统最终能够返回较为准确的检索结果。计算平均检索精度（mAP）的过程让我更加注重结果的准确性和效率。这次实验不仅让我掌握了多媒体内容分析系统设计的方法，还培养了我解决实际问题的能力，让我对多媒体数据分析领域充满了兴趣和信心。

|  |
| --- |
| 指导教师批阅意见：  成绩评定：  指导教师签字：  年 月 日 |
| 备注： |