**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（创新班）**

**指导教师： 朱映映**

**报告人： 何泽锋 学号： 2022150221**

**实验时间： 2025年6月3日-2025年6月11日**

**实验报告提交时间： 2025年6月18日**

**教务部制**

|  |  |  |
| --- | --- | --- |
| **一、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. 掌握多媒体智能系统设计方法，培养学生解决实际问题的能力 | | |
| **二、实验环境：**  系统：Windows 11  编程语言：python  依赖库：   |  | | --- | | import os  import json  import numpy as np  import torch  import torch.nn as nn  import torch.optim as optim  import torchvision.models as models  import torchvision.transforms as transforms  from torch.utils.data import Dataset, DataLoader  from PIL import Image  import matplotlib.pyplot as plt  from tqdm import tqdm  import glob  import random  from sklearn.model\_selection import train\_test\_split  import pickle  from PyQt5 import QtCore, QtGui, QtWidgets  from PyQt5.QtWidgets import (  QMainWindow, QWidget, QVBoxLayout, QHBoxLayout,  QLabel, QPushButton, QFileDialog, QScrollArea, QComboBox, QGridLayout,  QProgressBar, QMessageBox, QDialog, QLineEdit, QFormLayout  )  from PyQt5.QtGui import QPixmap, QImage  from PyQt5.QtCore import Qt, QThread, pyqtSignal | | | |
| **三、实验过程及内容：**  1. 数据集解读  本次实验涉及两个数据集，分别是Holidays dataset和Oxford Building dataset，接下来是对两个数据集的解读  （1）Holidays dataset  该数据集包含1491张照片，这些照片被分为500个组，同一个组的照片是从不同角度对同一场景或者物体进行拍摄。以下是其中一个组的例子   |  | | --- | | 图1 Holidays数据集举例  可以看到这组图像都是对一栋建筑的拍摄，其角度、前景有所区别，但照片的主体是一致的，因此检索的时候mAP会比较高 |   标签解释：  Holidays dataset的groundtruth是按照组来划分的，具体来说其对于每个组，会设置一张图作为“组长”，组中其余的图是相似图像    图2 Holidays groundtruth  按照该分组观察100000、100001、100002的图像，如下所示：   |  | | --- | | 图3 Holidays数据举例  可以看到，这些图像都是对山景的拍摄，角度、光线有所不同，但能明显看出是同一个主体（观察远处的山以及眼前的草会有较为明显的相似性） |   （2）Oxford Building dataset  该数据集包含5063张图，都是关于牛津的建筑。包含11个不同地标，每个地标有5个可能的查询表示。这提供了一组55个查询，其中混入了一些垃圾图像作为干扰。此数据集的groundtruth按照类型进行划分，分为good、ok、junk、query四类，每个类内写有对应的图片名称，此处简单列举一些groundtruth的案例  ① good   |  | | --- | | 图4 Oxford good数据举例  可以明显看出图片的主体是同一栋建筑，时间、俯仰角、天色等有所区别。 |   ②ok   |  | | --- | | 图5 Oxford ok数据举例  这几张图可以看的出来与上方图像包含相同的主体，但是拍摄的角度有较大的区别，同时可能出现前景遮挡等，相似度会较低一些。 |   ③junk   |  | | --- | | 图6 Oxford junk数据举例  这类型的图像主体不够清晰，拍摄的建筑不全，不能很好的辨认与上方是否相同，场景比较杂乱 |   ④query   |  | | --- | | 图7 Oxford query举例  这类型的标签会列出这组数据的“组长”，包含4个坐标点，是图像的主体内容 |   2. 流程分析  本次实验采用CNN实现基本的图像检索系统，具体的流程如下所示：    图8 图像检索流程  接下来将按照模块讲述具体的代码实现  （1）加载和处理数据  ①构建三元组  代码通过TripletDataset类处理图像数据，根据Ground Truth生成三元组（锚点、正样本、负样本）。对于Holidays数据集，直接解析JSON格式的标注文件；对于Oxford5k数据集，则处理特殊格式的文本标注并提取查询区域坐标。generate\_triplets方法为每个查询图像随机选择不相似的负样本，构建三元组数据。   |  | | --- | | 图9 构建三元组代码  如图所示为构建三元组数据，收集所有涉及的图像（查询和相似图像），确保负样本来自这些图像之外。对每个查询正样本对，随机采样一个负样本。最终生成三元组列表。 |   ②简单的数据增强  训练阶段：应用随机裁剪、水平翻转、颜色抖动，并归一化至 ImageNet 均值和标准差来增强数据，提高模型泛用性   |  | | --- | | 图10 训练数据增强 |   验证/检索阶段：采用中心裁剪和标准化，确保输入一致性   |  | | --- | | 图11 验证数据裁剪 |   （2）特征提取  基于预训练ResNet50构建特征提取网络FeatureExtractorNet  结构设计如下：   |  | | --- | | ·主干：保留ResNet50的前17层卷积层，输出特征图尺寸为(batch, 2048, 7, 7)  batch：批量大小；  2048：特征通道数，对应ResNet50的高级语义特征；  7×7：空间维度，每个位置对应原始图像的局部感受野。  **·降维层**：通过全连接层将2048维特征映射到512维嵌入空间  ·归一化层：对输出特征进行L2归一化，确保向量模长为1，便于余弦相似度计算    图12 特征提取网络 |   （3）损失函数  通过计算正样本对和负样本对的距离来作为损失，采用的是欧式距离的平方作为损失，数学表示如下：对于锚点 A、正样本 P 和负样本 N，要求 A 与 P 的距离小于 A 与 N 的距离，且差距至少为margin  ① 说明三元组已满足约束，无需更新参数。  ② 需要通过梯度下降减小d(A,P)、增大d(A,N)。   |  | | --- | | 图13损失函数定义 |   （4）训练流程  ·训练循环：  遍历训练数据加载器，每次迭代输入三元组图像，提取特征后计算三元组损失。  使用Adam优化器更新参数，学习率通过ReduceLROnPlateau调度器动态调整（当 mAP 停滞时降低学习率）。   |  | | --- | | 图14 训练流程 |   （5）模型评估  calculate\_map函数通过余弦相似度排序检索结果，计算每个查询的平均精度（AP），再求均值得到mAP   |  | | --- | | 图15 mAP计算代码 |   3. 配置文件  配置文件方便用户更改训练的参数、数据集等等  （1）数据集配置  数据集的参数主要包含所选数据集的名称、路径等   |  | | --- | | 图16 config数据集配置 |   （2）训练配置  训练配置包含训练的迭代数、随机种子、device、batchsize、学习率等参数   |  | | --- | | 图17 config训练配置 |   4. 训练过程展示  本次实验统一采用30个epoch进行训练，在代码中可以选择使用单数据集或混合数据集，如下图所示为混合数据集的训练过程，可以看到对于每一轮迭代，输出当前的损失，在Holidays、Oxford以及平均的mAP   |  | | --- | | 图18 训练过程展示 |   训练结束的时候会再次测试各数据集的mAP，并提取特征进行保存   |  | | --- | | 图19 训练结束展示 |   5. 训练结果分析    图20 loss、mAP比较展示  （1）Training Loss（训练损失）  ·整体趋势：训练损失在前期loss下降非常明显，从接近 0.02 的较高值快速降至 0.005 左右，随后在低数值区间（0.000 - 0.0025 ）缓慢下降 。符合模型训练的一般规律，前期模型参数与数据分布差异大，通过反向传播快速调整参数，损失快速降低；后期模型逐渐拟合数据，损失下降变缓并趋于稳定。  （2）Validation mAP（验证集平均精度均值）  ·Holidays 数据集：mAP数值整体较高，在0.55 - 0.7区间波动 。前期有一定波动，说明模型在学习过程中对该数据集的检索效果不稳定，但后期基本能维持在相对较高水平，表明模型对Holidays数据集的图像特征有较好的学习和检索能力，能够较为准确地根据特征匹配检索出相关图像。  ·Oxford 数据集：mAP数值较低，基本在0.25 - 0.3区间 。且整体变化幅度小，说明模型对该数据集的检索效果提升不明显，这是是因为 Oxford 数据集的查询区域裁剪、图像内容复杂度等因素，导致模型难以有效学习到区分度高的特征，检索性能受限。  ·Average：由于Holidays和Oxford数据集的mAP差异大，平均mAP主要受Holidays 数据集影响，整体在0.4 - 0.5区间，反映模型在两个数据集上综合的检索性能，体现出模型对不同数据集的适应性存在差距。  ·结合loss分析：loss的值是在不断下降当mAP没能有效提升，分析原因是随机选三元组时，易选到简单区分的样本（锚点与正样本差异大、与负样本差异小）。这些简单样本对模型学习价值低，虽让 Loss 下降，但没让模型学到有效区分复杂情况的特征，无法提升实际检索的mAP。  （3）Dataset Comparison（数据集对比）分析  ·Holidays数据集：mAP曲线前期下降后又回升并趋于平稳，虽然有波动，但整体维持在相对较高水平。说明该数据集的图像特征相对更易被模型学习和区分，模型在该数据集上能够较好地通过特征提取与匹配实现图像检索，即使训练过程中参数调整有波动，也能保持一定的检索精度。  · Oxford数据集：mAP曲线数值低且几乎呈水平状，基本没有明显提升。这是因为Oxford 数据集比较复杂，存在干扰和不佳的数据，导致模型难以提取到有效区分不同图像的特征，检索性能提升困难，模型对该数据集的适配性较差。  6. 可视化界面  （1）数据集输入  程序启动时，弹出对话框让用户选择 Oxford5k、Holidays 数据集的路径，确保程序能找到数据集文件。   |  | | --- | | 图21 数据集传入 |   （2）特征加载  在后台线程加载预提取的图像特征   |  | | --- | | 图22 特征加载 |   （3）核心逻辑  ImageRetrievalGU是程序的核心类，负责 GUI 初始化、模型加载、特征加载、检索逻辑和结果展示。   |  | | --- | | 图23 主页面 |   ·init\_widgets：GUI组件初始化  功能：搭建界面布局，包含“数据集选择下拉框”、“查询图像按钮”、“进度条”、“查询图像显示区”、“检索结果显示区”。   |  | | --- | | 代码过长，截取部分进行展示：    图24 init\_widgets函数 |   ·load\_model：加载特征提取模型  作用：加载预训练好的特征提取网络，用于提取查询图像的特征   |  | | --- | | 图25 load\_model函数 |   ·load\_features：加载预计算的特征文件  功能：存储所有候选图像的特征向量，避免在检索时重复计算，大幅提升检索速度   |  | | --- | | 图26 load\_features函数 |   ·retrieve：核心检索逻辑  作用：提取查询图像的特征，与预加载的所有特征计算余弦相似度，按相似度排序返回 Top-K 结果。   |  | | --- | | 图27 retrieve函数 |   7. 检索展示  （1）运行可视化界面  用户需要输入Holidays和Oxford5K所在的文件目录，此时后台线程会自动加载已经预计算好的特征文件   |  | | --- | | 如图为启动时的用户输入界面    图28 路径输入前  用户输入或采用浏览选择数据目录    图29 路径输入后 |   （2）图像检索系统  用户传入路径并等待模型、特征文件加载完毕后即可启动检索系统，如下图所示   |  | | --- | | 如图为启动的图像检索系统    图30 图像检索系统  用户可以选择检索的图像范围    图31 选择图像检索范围  点击右上角“选择查询图像”，传入需要查询的图像，此时系统会检索相关图像并给出检索得到图像来源数据集合和相似度，此处显示图采用余弦相似度    图32 图像检索结果1  另外一个检索结果，可以看到部分图像的检索效果较好    图33 图像检索结果2  也会存在部分图像检索错误，例如出现跨数据集的情况    图34 图像检索结果3 | | | |
| **四、实验结果与分析：**  （1）模型训练效果  在训练阶段，模型经过多轮迭代后，损失函数逐渐收敛，表明模型能够有效学习图像特征。训练过程中，通过观察损失曲线发现，初始阶段损失下降较快，随后趋于平缓，最终稳定在一个较低的值。这说明模型在训练数据上表现良好，未出现过拟合现象。此外，数据增强技术的应用（如随机裁剪、水平翻转等）显著提升了模型的泛化能力，使其在面对不同角度、光照变化的图像时仍能保持较高的检索准确率。  （2）检索性能评估  在测试阶段，模型在两个数据集上的检索性能表现如下：  Holidays数据集：由于该数据集的图像组内相似性较高（同一场景的不同角度拍摄），模型的检索准确率（mAP）达到了较高水平。尤其是在组内检索时，模型能够准确匹配同一场景的图像，即使存在视角或光照变化，也能保持较好的鲁棒性。  Oxford Building数据集：该数据集的图像分类较为复杂，包含“good”“ok”“junk”和“query”四类标签。模型在“good”和“ok”类别的检索中表现较好，能够准确识别同一建筑的不同拍摄角度；但在“junk”类别中，由于图像主体不清晰或遮挡严重，检索准确率有所下降。这表明模型对于低质量图像的鲁棒性仍有提升空间。 |

|  |
| --- |
| **五、心得体会：**  通过本次实验，我对多媒体内容分析与图像检索系统设计有了更深入的理解。在数据处理环节，构建三元组时需平衡样本难度，避免简单样本主导训练，实验过程采用随机采样策略，导致训练集中混入大量简单样本，虽然损失快速下降，但模型实际检索能力未提升，后续可尝试难样本挖掘策略以提升模型鲁棒性。  模型选择上，基于ResNet50的特征提取网络在结构上较为基础，面对复杂数据集时表现不足，后续可以尝试更复杂的模型或引入注意力机制优化特征表达。此外，数据增强策略在训练阶段有效提升了模型泛化性，但验证阶段的中心裁剪可能限制了对图像局部特征的捕捉，可进一步优化预处理流程。  在训练的时候提前保存得到的数据特征文件，在可视化的时候直接读入，可以大幅提高启动时间，提高检索效率。 |

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

**注：1、报告内的项目或内容设置，可根据实际情况加以调整和补充。**

**2、教师批改学生实验报告时间应在学生提交实验报告时间后10日内**。