THU-70250043-0, Pattern Recognition (Spring 2021) **Project: 1**

大作业：商品图像检索

Lecturer: Changshui Zhang zcs@mail.tsinghua.edu.cn

Hong Zhao vzhao@tsinghua.edu.cn

Student:

1 作业主题

图像检索问题（Image Retrieval）是数字图像处理领域重要而经典的问题。随着互联网尤其是移动互联网技术的飞速发展，电子商务已经融入到我们的日常生活中，互联网商品图像检索变成了一个有着广泛实际应用的课题。近年来该问题得到很多关注，其中不乏一些顶刊、顶会上的工作 [1, 2, 3, 4]，有一些机器学习类竞赛也关注此问题。在实际应用中，一些本课程有所提及的机器学习、模式识别的方法和思路，将此问题的解决方案不断完善。

文献 [5, 6] 介绍了基于内容的图像检索的方法。早期，我们通过判断待查询图像和数据集其他图像特征相似性以实现检索。随着深度学习的发展，深度模型自动提取特征越来越流行。当前的图像检索方法按照数据有无标注可以划分为：监督、无监督、半监督、弱监督以及伪监督和自监督方法；按照模型主体结构又包括，自编码网络、孪生网络、对抗生成网络、注意力网络、循环神经网络等；按照特征的形式又可以分为：01 二值描述、实数特征描述以及聚合描述。 文献 [7] 介绍了一些机器学习方法，比如，基于 PCA、SIFT 以及 SURF 等传统的特征提取方法，结合特征编码技术实现的图像检索。

为了进一步提升检索效果，人们不再局限于图像之中，开始着眼于利用多模态信息（如视频、图像、文本等）来实现跨模态或多模态的检索。文献 [1] 对输入图像加上了一些描述文本，该文本指定了对输入图像要做出的修改，要求模型寻找出与修改后图片相似的图像。文献 [2, 3] 则关注于文本和图像之间的跨模态检索方法。 另外图像检索还涉及到数据长尾分布和鲁棒性的问题。文献 [8, 9, 10] 介绍了一些解决数据长尾分布的方法。 文献 [11, 12] 提出了一些增强鲁棒性的方法。 请大家在这些文献的基础上，继续调研，深入思考，更好地完成本次训练。

在本次课程项目中，我们聚焦于一个具体的用机器学习方法完成的商品图像检索任务。作业要求完整地完成数据处理、建模、实验、分析等过程。

2 作业内容

要求以小论文的格式提交实验报告，你的报告需要包括但不限于本章指定的基础要求，在基础要求（基础分 80 分）以外的实验、分析、讨论等（以下标注为：提高要求）均可以加分（满分 100 分，加满为止）。

注意由于任务定义和评价指标选择是相对开放的，评分时将更看重报告的完整度，在评价算法结果时将更看重改进的模型相对于基础模型的提高，而非最终结果的绝对值。同时，为了使得相对提高看起来明显而故意把基础模型性能做差的行为是严格禁止的，这一点会影响最终评分。

一、题目，摘要，关键词

二、简介（5’）

- 商品图像检索任务及主流的解决方法。
- 机器学习方法在商品图像检索中的应用（在本文档提供的文献之外，至少调研 2 篇相关文献）。
- 本文的主要贡献。

三、任务定义（5’）

- 本次项目提供的数据集中，包含图像信息与文本信息。本次作业需要完成三个任务：
  （1）仅利用图像信息完成检索。
  （2）仅利用文本信息完成检索。
  （3）同时利用图像、文本信息完成检索。
- 考虑如何处理以上三个任务，请用符号和公式定义输入和输出。

四、数据整理（5’）

- 描述数据内容。
- 如需要，请给出数据清洗、缺失值处理等具体方式。
- 可视化地观察数据分布。

五、方法设计（30’）

- 每个任务使用至少三种不同的方法进行对比（可以参考以下技术手段），不同任务可以采用相同方法。
  1. PCA
  2. SIFT

3. Deep learning-based
4. KNN
5. Boosting-based

- 如果完成该任务需要特征提取，请说明：

  （1）数据归一化处理方法；

  （2）可尝试特征间的组合；

  （3）特征分析模块或方法。

- 使用至少 1 种 ensemble 方法组合不同模型。
- （提高要求）尝试从机器学习的角度深入思考尝试对现有方法做出自己的改进。参考角度如下，可继续拓展。

  1. 长尾分布问题
  2. 鲁棒性
  3. 跨模态学习

七、实验设计及结果（15’）

- 数据集如何划分成训练、验证、测试集。
- 评价指标。请合理选择至少 3 个评价指标作为结果并给出相应的计算公式，评价指标可以是：准确率/召回率/F1 score/AUC/mAP等。表 1 显示了我们在本次作业数据上（数据详情见章节 3）简单实验的结果，其中文字信息采用的是 tf-idf 特征，图片信息采用的是预训练 resnet50 提取的特征，利用余弦距离计算相似度，并选定阈值得到结果。注意我们得出的结果是直接在训练数据上计算得到的评价指标，只能提供一个参考。

	F1	AUC	mAP
只利用文字信息	0.617	0.992	0.787
只利用图片信息	0.638	0.975	0.681
利用两种信息（并集）	0.725	-	-

表 1: 简单实验结果

- 每个任务每种模型单独的最好结果对比（列出图表并进行讨论）。
- 每个任务 Ensemble 后的最好结果对比（列出图表并进行讨论）。
- 每个任务每种模型在不同超参数下的表现（列出图表并进行讨论）。

- （提高要求）从机器学习角度改进后的实验设计及结果。

八、实验结果分析（15’）

实验结果分析的方式可以包括但不限于：

- 讨论数据和方法中每一部分的贡献（若删去/更换部分数据输入/实验设定/模型结构，结果会发生什么变化）。
- 特征的重要性分析。
- 错误分析（模型对哪些数据预测准确率高，对哪些数据预测准确率低）。
- 案例分析（在具体的案例上，不同模型表现的区别在哪里）。
- 模型和结果可视化分析。
- （提高要求）从机器学习角度改进后的实验结果分析。

九、代码接口（5’）

具体要求见第3节。

十、小组成员贡献

十一、结论

3 作业提交

3.1 作业说明

- 本次作业的数据来自 kaggle 比赛: Shopee —Price Match Guarantee。按照 kaggle 网站规则，需要大家注册参赛才可下载和使用数据。鼓励大家体验、参与比赛的全过程。但是大家不必有压力，**比赛成绩和本次作业成绩无关**。(为方便大家下载，数据集已上传至清华云盘)。
- 请将数据集自行划分为训练集、验证集和测试集, 并报告测试集上的结果。测试集**不允许**参与模型的训练。如果算力不够，允许取出完整数据集中的一部分数据作为 mini 数据集，然后在 mini 数据集中完成此次任务。
- 可以个人或 2 人小组的形式完成作业，每小组提交一份作业即可，须用独立段落说明成员贡献。(注意: 对小组完成的作业要求更高。)
- 完成作业过程中可以参考已有的代码，但要在报告中用独立段落给出详尽说明，具体到自己提交的代码中哪一个文件哪些行，并提供参考来源的链接。若参考的代码过多，将影响评分；若在未标明参考他人的部分发现了雷同现象，本次作业将判定零分。

- 编程语言不限。

3.2 提交代码说明

- 自己完成的全部训练、测试代码均需要进行提交。库函数的代码无需提交，除非自己修改了某个已有的库。
- 训练数据如有特殊处理（如划分出 mini 数据集），需在报告中说明，无需上传。中间文件和生成的结果文件不需要上传。
- 需要从所完成的模型中选择一个 ensemble 前单个性能最优的，作为可直接运行的测试模型提交。只需提交最优模型训练好的参数，其他模型提交代码即可。
- 提交的代码必须是自包含的（self-contained）。即，从下载原始数据开始，不依赖于任何中间结果，必须要能够完整地复现训练、测试过程。

3.3 提交文件说明

- 最终作业以 zip 压缩包形式提交。除压缩包命名外，目录名和文件名采用英文，防止因为操作系统不同造成乱码。提交的压缩包展开后目录结构应当如下：

  - 张三-商品图像检索.zip
    - codes/
      - README.md
      - ...
    - report.pdf

  其中，report.pdf 为报告，... 部分为代码，可以包含多个文件或目录，**代码中要求用相对路径以方便复现**。

- 需要写一个Markdown格式的 README.md，内容至少要包括：
  - 按模块描述压缩包中每部分文件的作用。
  - 运行代码所需的软件环境和软件版本，并提供从裸的操作系统开始配置所需环境的命令。
  - 下载原始数据和整理数据的命令。
  - 训练所提交的最优模型的命令。按此命令训练出的模型，不应当与用提交的模型参数加载出的最优模型性能差异过大。
  - 用上一步训练出来的模型在测试集数据上测试的命令。
  - 用提交的最优模型在测试集数据上测试的命令，以及预期的结果。
- README.md 必须包含关于环境、数据、训练、测试的所有命令。README.md 中的命令依次复制粘贴到命令行里执行，应当能够复现整个流程。提交前建议反复检查并确认这一点，复现失败将极大影响评分。

- 最终提交的压缩包体积不得超过 50MB，如模型或者测试数据太大导致超出这一限制，需要将较大的文件移出压缩包直至满足这一限制为止，并将这些文件放在互联网上（例如清华云盘）且在 README.md 中提供下载命令。

参考文献

[1] Vo N, Jiang L, Sun C, et al. Composing text and image for image retrieval-an empirical odyssey[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 6439-6448.

[2] Wang Z, Liu X, Li H, et al. Camp: Cross-modal adaptive message passing for text-image retrieval[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 5764-5773.

[3] Zhang Q, Lei Z, Zhang Z, et al. Context-aware attention network for image-text retrieval[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 3536-3545.

[4] Lang Y, He Y, Yang F, et al. Which Is Plagiarism: Fashion Image Retrieval Based on Regional Representation for Design Protection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 2595-2604.

[5] Dubey S R. A Decade Survey of Content Based Image Retrieval using Deep Learning[J]. arXiv preprint arXiv:2012.00641, 2020.

[6] Chen W, Liu Y, Wang W, et al. Deep Image Retrieval: A Survey[J]. arXiv preprint arXiv:2101.11282, 2021.

[7] Zheng L, Yang Y, Tian Q. SIFT meets CNN: A decade survey of instance retrieval[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 40(5): 1224-1244.

[8] Zhou B, Cui Q, Wei X S, et al. Bbn: Bilateral-branch network with cumulative learning for long-tailed visual recognition[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 9719-9728.

[9] Schroeder B, Tripathi S. Structured Query-Based Image Retrieval Using Scene Graphs[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2020: 178-179.

[10] Zhou X, Pan P, Zheng Y, et al. Large Scale Long-tailed Product Recognition System at Alibaba[C]//Proceedings of the 29th ACM International Conference on Information & Knowledge Management. 2020: 3353-3356.

[11] Humenberger M, Cabon Y, Guerin N, et al. Robust Image Retrieval-based Visual Localization using Kapture[J]. arXiv preprint arXiv:2007.13867, 2020.

[12] Somasundaran B V, Soundararajan R, Biswas S. Robust image retrieval by cascading a deep quality assessment network[J]. Signal Processing: Image Communication, 2020, 80: 115652.