

图像检索系统设计报告

黎科荣 202304060821 电子 2301 班

一、引言

随着计算机视觉与模式识别技术的快速发展，图像理解与检索已成为智能信息处理领域中的重要研究方向。传统图像处理方法通常依赖人工设计的特征，如颜色直方图、纹理描述子和局部关键点特征，在早期视觉任务中取得了一定成效。然而，这类方法在复杂场景下对高层语义信息的表达能力有限。近年来，以卷积神经网络（CNN）为代表的深度学习方法在图像分类与特征表示方面表现出显著优势，能够自动学习多层次的特征表示，但其对数据规模和计算资源的依赖较高。

在此背景下，如何在传统特征与深度特征之间进行有效对比与融合，充分发挥二者的互补优势，成为智能信息处理中的一个重要问题。CIFAR-100 数据集包含类别丰富、分辨率较低的自然图像，既适合传统特征分析，也适合深度模型训练，为相关研究提供了良好的实验平台。

本课程设计以 CIFAR-100 数据集为研究对象，围绕图像特征提取与相似性检索展开。课题目标包括：一是实现基于传统视觉特征（如颜色与局部特征）的图像表示与检索方法；二是利用预训练的深度卷积神经网络提取高层语义特征并进行比较分析；三是探索不同特征在图像检索任务中的适用性，并为后续的特征融合方法奠定实验基础。通过本课题的研究，加深对智能信息处理相关理论与方法的理解，并提升综合运用图像处理与深度学习技术解决实际问题的能力。

二、相关技术简介

在图像理解与检索任务中，特征表示与相似性度量是核心技术环节。传统图像特征通常由人工设计，侧重于对图像低层视觉信息的描述，如颜色分布、纹理结构和局部关键点等。其中，颜色直方图通过统计不同颜色空间中的像素分布，能够有效反映图像的整体颜色特性；基于局部关键点的特征（如 SIFT）则通过检测并描述局部显著区域，在一定程度上具有尺度与旋转不变性，适合刻画图像的局部结构信息。这类特征计算成本较低、可解释性强，但对复杂语义和全局结构的表达能力有限。

卷积神经网络（CNN）特征是近年来广泛应用于图像分析的重要技术。CNN 通过多层卷积和非线性变换，能够从原始图像中逐级学习由低层到高层的特征表示，最终形成具有较强语义判别能力的特征向量。相较于传统特征，CNN 特征在复杂场景和多类别任务中表现出更强的鲁棒性和泛化能力，尤其适合刻画图像的整体结构和语义信息。在实际应用中，常利用预训练模型提取中高层特征作为图像的全局表示。

在图像检索任务中，相似性度量用于衡量不同图像特征向量之间的相似程度。常用的距离度量方法包括欧氏距离和余弦相似度。欧氏距离关注特征向量在空间中的绝对差异，适用于特征分布较为均匀的情况；余弦相似度则通过计算向量夹角来衡量方向一致性，更关注特征的相对分布特征，常用于高维特征空间。合理选择特征表示方式与距离度量方法，是提升图像检索性能

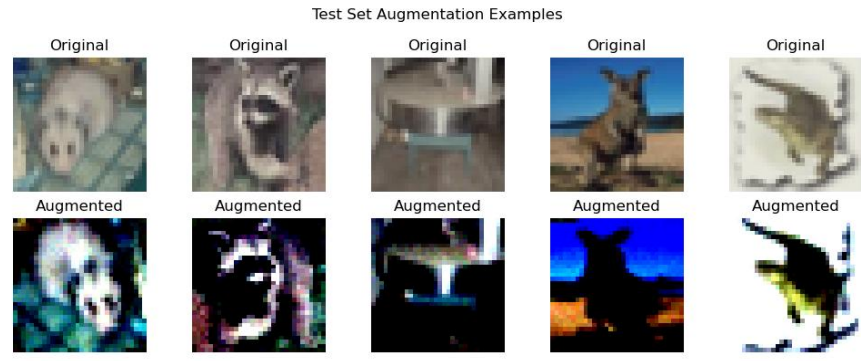
能的重要因素。

三、系统设计与实现

检索系统整体采用模块化设计，主要由离线处理模块和在线检索模块组成。离线阶段主要完成数据预处理、特征提取与特征库构建；在线阶段则针对查询图像提取对应特征，并通过相似性度量在特征库中进行匹配，最终返回最相似的图像结果。该设计能够有效减少重复计算，提高检索效率，同时便于不同特征方法的对比与扩展。

① 数据预处理模块

在数据预处理阶段，针对 CIFAR-100 图像分辨率较小的特点，对原始图像进行统一读取与格式转换，并在训练阶段引入随机旋转、水平翻转和随机缩放裁剪等几何增强操作，以提高模型的泛化能力。预处理后的数据被统一保存，保证后续传统特征提取与深度特征训练使用同一数据基础，从而增强实验结果的可比性。



② 传统特征提取与检索模块

传统特征模块主要关注图像的低层视觉信息。系统中实现了基于颜色分布和局部结构的特征提取方法，例如颜色直方图和 SIFT 特征。对于每张图像，首先提取对应的传统特征向量，并在离线阶段构建特征库。在检索阶段，针对查询图像计算同类特征，并采用欧氏距离或余弦相似度对特征向量进行匹配，根据相似度排序返回最相近的图像结果。该模块在颜色和纹理信息较为显著的类别上表现较为稳定。

③ 深度神经网络训练与特征提取模块

在深度特征提取方面，系统选用在 ImageNet 数据集上预训练的 ResNet-50 网络作为基础模型，并在 CIFAR-100 数据集上进行微调训练。通过调整网络的部分结构和训练策略，使模型能够更好地适应目标数据集。训练完成后，选取网络中全连接层之前的高层特征作为图像的深度表示，并在离线阶段构建深度特征库。在检索阶段，查询图像通过同一网络提取特征，并利用相似度度量方法完成匹配与排序。该模块在依赖整体结构和语义信息的类别上具有明显优势。

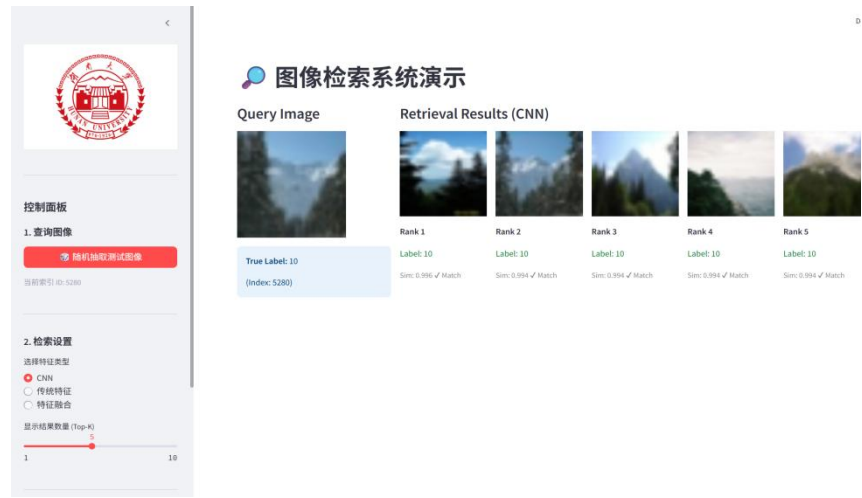
④ 特征融合与检索模块

为进一步提升检索性能，系统在此基础上实现了传统特征与 CNN 特征

的融合检索方法。具体而言，将同一图像对应的传统特征和深度特征进行特征级拼接或在相似度层面进行加权融合，从而构建更加丰富的图像表示。在检索过程中，融合后的特征向量用于相似性计算，有效结合了传统特征对颜色与局部细节的描述能力，以及 CNN 特征对高层语义的表达能力。实验结果表明，特征融合方法在整体检索效果和稳定性方面具有一定优势。

⑤ 检索结果展示与分析

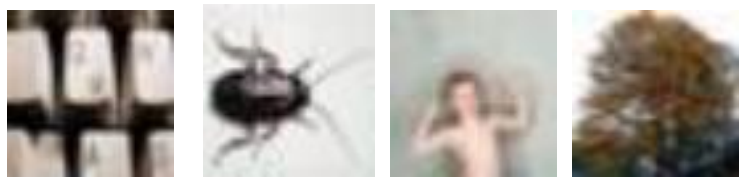
我们使用 streamlit 搭建了一个简单 web 界面，将检索结果以可视化方式进行展示，包括一张查询图像及其最相似的若干张结果图像。通过对不同特征方法下的检索结果进行对比分析，可以直观地观察传统特征、深度特征及其融合方法在不同类别上的表现差异，为后续实验分析和性能评估提供依据。



四、实验设置

① 数据集说明

本实验采用 CIFAR-100 数据集作为研究对象。该数据集包含 100 个类别的彩色自然图像，每个类别包含 600 张样本，其中训练集与测试集分别包含 50,000 张和 10,000 张图像。所有图像的分辨率均为 32×32 像素，涵盖动物、交通工具、日常物品等多种场景，具有类别丰富、语义多样的特点。该数据集广泛应用于图像分类与检索任务，适合作为传统特征与深度特征对比分析的实验平台。在实验中，训练集用于模型训练及特征库构建，测试集用于查询图像的检索与性能评估。为保证实验的一致性，传统特征与深度特征均基于相同的数据划分进行提取与对比。



② 参数设置

在数据预处理阶段，对训练集图像引入随机旋转、水平翻转以及随机缩放裁剪等几何增强操作，以提升模型的泛化能力，测试集仅进行必要的格式

转换，不引入随机增强。

在深度特征提取实验中，选用在 ImageNet 数据集上预训练的 ResNet-50 网络作为基础模型，并在 CIFAR-100 数据集上进行微调训练。训练过程中将 epochs 设置为 500, batch_size 设置为 128, 损失函数采用 CrossEntropyLoss, 采用随机梯度下降 (SGD) 优化算法, 动量参数设置为 0.9, 权重衰减系数为 5×10^{-4} , 初始学习率根据实验经验设定为 0.1, 并在训练过程中逐步衰减。

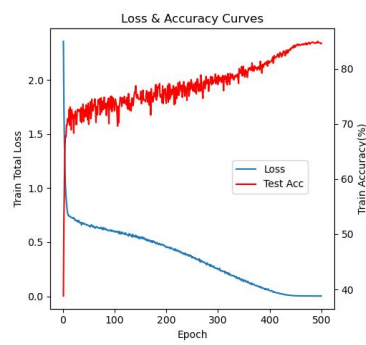
在图像检索阶段，针对传统特征、深度特征以及融合特征分别构建特征库。相似性度量方法选用欧氏距离和余弦相似度，并根据不同特征类型进行对比实验。

③ 评价指标

为定量评估检索系统的性能，本实验主要采用 Top-K 检索准确率作为评价指标。对于每一张查询图像，若其返回的前 K 张检索结果中包含与查询图像同类别的样本，则认为该查询命中。通过对测试集中所有查询样本的统计，计算不同 K 值下的平均检索准确率，以衡量系统的整体检索性能。

五、实验结果与分析

在模型训练阶段，对基于 ImageNet 预训练的 ResNet-50 在 CIFAR-100 数据集上的微调过程进行了统计与分析。从训练曲线来看，随着训练轮数的增加，模型的训练损失 (loss) 整体呈现出逐步下降的趋势，并在后期趋于稳定，表明网络参数逐渐收敛。同时，训练集和测试集上的分类准确率 (accuracy) 持续提升，在初期增长较快，随后进入平缓阶段，说明模型在有效学习判别特征的同时，未出现明显的震荡或发散现象。通过冻结部分网络层，仅对高层语义相关模块进行微调，在保证训练稳定性的同时降低了过拟合风险。



epochs	Max test accuracy
100	75.02%
200	77.29%
300	79.02%
400	81.43%
500	84.95%

在图像检索实验中，分别基于传统特征、CNN 深度特征以及二者融合特征进行了对比分析。实验结果表明，基于传统特征的检索方法在颜色分布较为稳定、外观差异明显的类别上具有一定优势，但在类别语义相近或形状复杂的场景中，检索结果容易受到背景和局部纹理干扰。相比之下，基于 CNN 深度特征的检索方法在整体结构和语义层面表现更加稳定，能够较好地区分不同类别的图像，但在部分依赖颜色细节的类别中仍存在一定误检情况。

SIFT + COSINE | Query idx=1681

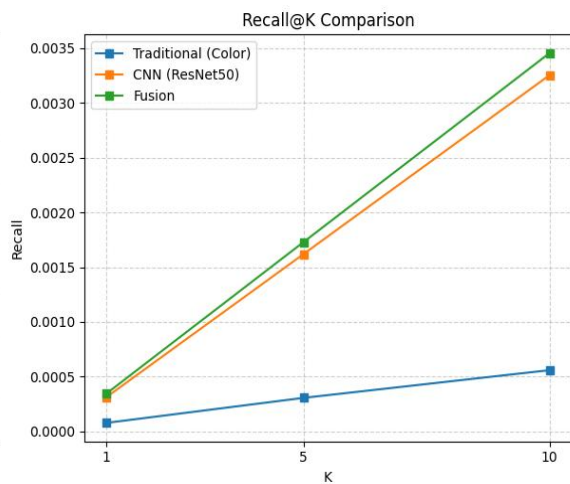
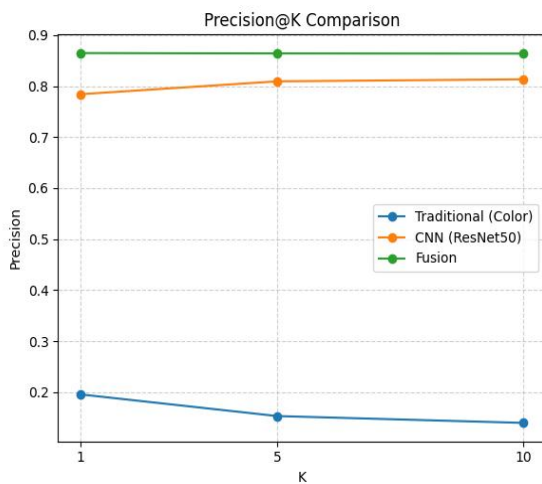


CNN + COSINE | Query idx=1681



进一步的实验结果显示，将传统特征与 CNN 特征进行融合后，检索性能在多数情况下得到提升。融合方法在保持 CNN 特征语义判别能力的同时，引入了传统特征对颜色与局部细节的补充信息，使检索结果在准确性和稳定性方面均有所改善。通过对检索结果的可视化分析可以观察到，融合特征能够在更多查询样本中返回类别一致且视觉相似度较高的图像，验证了不同特征之间的互补性。

fusion_feature + COSINE | Query idx=1681



	P@1	P@5	P@10	R@1	R@5	R@10
Traditional (Color)	0.1958	0.15328	0.13990	0.000078	0.000307	0.000560
CNN (ResNet50)	0.7840	0.80924	0.81324	0.000314	0.001618	0.003253
Fusion	0.8647	0.86392	0.86358	0.000346	0.001728	0.003454

六、个人负责的主要部分

在本课程设计中，我主要负责数据集的选取与整理，以及深度学习模型的训练与相关实验配置工作。在数据集方面，综合考虑数据规模、类别丰富性以及任务难度，选用 **CIFAR-100** 数据集作为实验对象，并对其数据结构、类别划分和样本分布进行了分析。在此基础上，完成了数据预处理与增强方案的设计，为后续特征提取和检索实验提供了统一且可靠的数据基础。

在模型训练方面，我负责搭建并训练基于 **ResNet-50** 的深度特征提取模型。具体包括挑选和加载 **ImageNet** 预训练权重、调整网络结构以适配 **CIFAR-100** 分类任务，并制定合理的微调策略，通过冻结部分网络层、仅训练高层语义相关模块来平衡训练效率与模型性能。同时，对训练过程中的关键参数（如学习率、优化器及训练轮数等）进行了多组实验对比，并记录训练过程中损失函数和准确率的变化情况，为后续实验分析提供依据。通过上述工作，成功获得了稳定可用的深度特征模型，为系统整体实验的顺利开展奠定了基础。

七、个人总结

通过本次图像检索系统的课程设计，我系统地完成了从数据集选择、模型训练到实验分析的完整实践过程，对图像检索与特征表示相关技术有了更加深入的理解。在实验过程中，我围绕 **CIFAR-100** 数据集，完成了数据预处理与增强方案的设计，并基于 **ImageNet** 预训练的 **ResNet-50** 模型进行了微调训练，掌握了迁移学习中冻结与解冻网络层的基本策略，以及训练参数对模型性能的影响。通过对训练过程中损失函数和准确率变化的分析，我进一步加深了对深度模型收敛特性与泛化能力的认识。

此外，在对比传统特征与深度特征的检索实验中，我直观体会到不同特征在表达能力和适用场景上的差异，并认识到特征融合在提升系统整体性能中的重要作用。本课程设计不仅巩固了课堂所学的智能信息处理相关理论知识，也显著提升了我在实际工程中进行数据处理、模型训练和实验分析的综合能力，为后续深入学习计算机视觉与模式识别方向的内容奠定了良好基础。