摘 要：

本文复刻了一项基于内容的图像分类技术。该算法旨在通过提取图像的局部特征并构建词袋模型，实现对图像的自动分类。在先前的研究中，该算法已经被证明在图像分类领域具有一定的准确性和实用性，因此本文旨在对其进行复制和验证。

首先，本文介绍了研究背景和课题意义，阐述了数字图像数据管理的挑战和基于内容的图像分类技术的重要性。随后，详细描述了词袋模型的构建过程，包括特征提取和聚类算法生成视觉词汇的步骤。在特征提取阶段，采用了SIFT算法对图像进行局部特征点提取，为后续的词袋构建奠定了基础。在聚类算法阶段，利用K-Means算法对提取的特征点进行聚类，生成视觉词汇作为词袋的基础单元。

接着，本文介绍了支持向量机分类器的原理和应用，说明了其在图像分类中的优势和适用性。支持向量机作为一种强大的分类器，能够在高维空间中构建最优的分类超平面，对图像分类具有较高的准确性和泛化能力。

本文复现了先前研究中的实验设置和方法，对图像数据集进行了处理，并利用相同的算法流程进行实验。实验结果显示，复制的算法在图像分类任务上取得了与先前研究相符合的结果，验证了其可复现性和有效性。

最后，本文对实验结果进行了分析和总结，讨论了词袋模型中视觉词汇数量对分类效果的影响，以及支持向量机分类器的性能表现。同时，也对该算法的局限性和未来改进方向进行了探讨，为进一步研究和应用提供了参考和启示。通过复刻先前研究的图像分类算法，本文为相关领域的研究提供了重要的复现性和验证性支持，为后续研究工作奠定了基础。

关键词：词袋模型、特征提取、聚类算法、支持向量机、SIFT算法、K-Means算法。

Summary:

This article reproduces a content-based image classification technique. This algorithm aims to achieve automatic image classification by extracting local features of the image and constructing a word bag model. In previous studies, this algorithm has been proven to have certain accuracy and practicality in the field of image classification, so this article aims to replicate and verify it.

Firstly, this article introduces the research background and significance of the topic, and elaborates on the challenges of digital image data management and the importance of content-based image classification technology. Subsequently, the construction process of the word bag model was described in detail, including the steps of feature extraction and clustering algorithm to generate visual vocabulary. In the feature extraction stage, the SIFT algorithm was used to extract local feature points from the image, laying the foundation for subsequent word bag construction. In the clustering algorithm stage, the K-Means algorithm is used to cluster the extracted feature points and generate visual vocabulary as the basic unit of the word bag.

Next, this article introduces the principle and application of support vector machine classifiers, explaining their advantages and applicability in image classification. Support Vector Machine, as a powerful classifier, can construct the optimal classification hyperplane in high-dimensional space, and has high accuracy and generalization ability for image classification.

In the experimental section, this article reproduced the experimental settings and methods from previous research, processed the image dataset, and conducted experiments using the same algorithm flow. The experimental results showed that the replicated algorithm achieved consistent results with previous studies in image classification tasks, verifying its reproducibility and effectiveness.

Finally, this article analyzes and summarizes the experimental results, discusses the impact of visual vocabulary quantity on classification performance in word bag models, and the performance of support vector machine classifiers. At the same time, the limitations and future improvement directions of the algorithm were also discussed, providing reference and inspiration for further research and application. By replicating the previously studied image classification algorithms, this article provides important reproducibility and validation support for research in related fields, laying the foundation for subsequent research work.

**Key** **words**: Bag of word model, Feature extraction, Clustering, SVM, SIFT, K-Means.

目录

[1.绪论 - 1 -](#_Toc166661327)

[**1.1 课题背景** - 1 -](#_Toc166661328)

[**1.2 词袋模型简介** - 1 -](#_Toc166661329)

[**1.3 步骤分析** - 2 -](#_Toc166661330)

[2.数据集处理 - 2 -](#_Toc166661331)

[**2.1 数据集准备** - 2 -](#_Toc166661332)

[**2.2 数据集划分** - 3 -](#_Toc166661333)

[3.词袋模型构建 - 3 -](#_Toc166661334)

[**3.1 词袋模型** - 3 -](#_Toc166661335)

[**3.2 特征提取** - 4 -](#_Toc166661336)

[**3.3 生成词袋** - 4 -](#_Toc166661337)

[4.图像描述 - 5 -](#_Toc166661338)

[**4.1 特征直方图** - 5 -](#_Toc166661339)

[**4.2 使用SPM算法的特征直方图** - 5 -](#_Toc166661340)

[5.训练 - 6 -](#_Toc166661341)

[**5.1 分类器** - 6 -](#_Toc166661342)

[**5.2 SVM** - 6 -](#_Toc166661343)

[6.结果分析 - 7 -](#_Toc166661344)

[**6.1 预测评估** - 7 -](#_Toc166661345)

[**6.2 总结** - 7 -](#_Toc166661346)

# 绪论

**1.1 课题背景**

随着计算机技术和互联网的迅猛发展，以及数字图像获取技术的不断进步，我们生活中涌现出了大量的数字图像数据。然而，传统的人工方式对这些图像进行分类、组织和管理已经变得相当耗时耗力。因此，迫切需要一种能够自动分析处理图像内容的方法，以快速、规范、自动地进行图像数据的组织、归类和管理。

早期图像分类主要依赖于文本特征，需要人工为图像标注文本，采用基于文本的分类模式。然而，这种方法存在明显的缺陷，需要耗费大量时间和人力，并且分类效果不尽如人意。随着计算机技术和数字化图像技术的发展，图像库规模的不断扩大，传统的人工标注方式已经不再适用。因此，研究重点逐渐转移到基于图像内容分析的自动分类方法上。

基于内容的图像分类技术无需人工标注语义信息，而是直接对图像内容进行处理和分析，利用图像的底层视觉特征进行分类。这种技术集成了机器学习、模式识别、计算机视觉和图像处理等多个领域的研究成果，具有重要的研究意义和实际应用价值。

未来，随着人工智能技术的不断进步，基于内容的图像分类技术将会变得更加智能和高效。这将为图像数据的管理和应用带来新的可能性，推动数字图像在各个领域的广泛应用和发展。

**1.2 词袋模型简介**

词袋模型(Bag of words model)最初用于文本分类中，在信息检索中，词袋模型假定对于一个文本，忽略其词序和语法，句法，将其仅仅看做是一个词集合，或者说是词的一个组合，文本中每个词的出现都是独立的，不依赖于其他词 是否出现，或者说当这篇文章的作者在任意一个位置选择一个词汇都不受前面句子的影响而独立选择的。

而在图像分类中，图像被视为是一些与位置无关的局部区域的集合，因此这些图像中的局部区域就等同于文本中的单词了。在不同的图像中，局部区域的分布是不同的。因此，可以利用提取的局部区域的分布对图像进行识别。图像分类和文本分类的不同点在于，在文本分类的词袋模型算法中，字典是已存在的，不需要通过学习获得；而在图像分类中，词袋模型算法需要通过监督或非监督的学习来获得视觉词典。

**1.3 步骤分析**

基于词袋模型的图像分类算法一般分为四步：首先，对图像进行局部特征向量的提取。为了取得很好的分类效果，提取的特征向量需要具备不同程度的不变性，如旋转，缩放，平移等不变性；然后，利用上一步得到的特征向量集，抽取其中有代表性的向量，使用聚类算法得到视觉单词，进而形成视觉词典；接着，对每一张图像进行向量化表达，即将每一张图像的特征点使用视觉词典描述，生成频率直方图。最后，设计并训练分类器，利用图像中单词的分布进行图像分类。

根据上述过程分析，可以将算法流程细化为以下9步，下面几章将根据以下9步做具体实现：

1. 提取数据集中的样本，并划分训练集和测试集
2. 对于训练集的所有图片，提取图片的 SIFT 特征点，并对 SIFT 特征点向量归一化
3. 对所有 SIFT 特征点使用聚类算法分为 n 类
4. 将 n 类特征点的中心点作为视觉词汇，生成词袋（字典）
5. 使用论文中的 SPM 算法生成图片的特征向量
6. 将图片分为 3 种尺度，分别为 1\*1、 2\*2、4\*4 大小，统计不同尺度下的特 征直方图
7. 将不同尺度下的特征直方图合并，组合成为一个 21（ 1 + 4 + 16 ） \* n 维 度大小的特征向量
8. 将图片的特征向量作为数据集，使用支持向量机算法完成分类任务
9. 对测试集图片做和训练集相同的数据处理操作
10. 使用支持向量机模型对测试集图片的类别进行预测
11. 评估预测结果，并生成分类报告和输出混淆矩阵

# 数据集处理

**2.1 数据集准备**

在开始实现算法流程的第一步之前，首先需要准备数据集。数据集应包含全体用于训练和测试的图像样本。本文使用scene\_categories数据集，该数据集包含15个类别（文件夹名就是类别名00~14），每个类中前 150号的样本作为训练样本，15个类一共2250张训练样本；剩下的样本构成测试集，15个类一共2235张测试样本。

**2.2 数据集划分**

设计DatasetProcessor类，输入数据集路径及训练集数量，例如输入dataset和150，divTrainAndTest将遍历dataset下的全部15个文件夹，获取其文件夹名作为类名及当前目录下的图片数据，将图片和类名作为整体存放到属性imgTrainSet和imgTestSet中。

# 词袋模型构建

**3.1 词袋模型**

为了表示一段文本，我们可以将其仅仅看作是若干个词汇的集合，例如有如下两个文档：

1：Bob likes to play basketball, Jim likes too.

2：Bob also likes to play football games.

根据上面的两个句子，我们可以提取出一个共同的词典D，这个词典一共包含10个不同的单词，利用词典的索引号，上面两个文档每一个都可以用一个10维向量表示：

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 序号 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| D | Bob | like | to | play | basketball | also | football | games | Jim | too |
| 1 | 1 | 2 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 1 |
| 2 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0 |

同样的，为了表示一幅图像，我们可以将每一张图像看作文档，即若干个“视觉词汇”的集合，与文档描述相同，视觉词汇相互之间没有顺序。将图片中的特征提取出来后，可以通过聚类算法将大量特征归为较小的词袋，之后就可以使用生成的视觉词袋将每张图像描述出来。

图表

描述已自动生成

图3.1-1 图像的向量描述

**3.2 特征提取**

上节提到，生成词袋需要先得到句子中的单词。但在图像中，图像的词汇不像文本文档中的那样是现成的，我们需要首先从图像中提取出相互独立的视觉词汇，这些视觉词汇需要具备以下的特性：局部性、放缩不变性、旋转不变性、光照不变性。因此，本文选择使用SIFT算法进行特征提取。

由于SIFT算法具有光照不变性，因此可以先将图片数据进行灰度化处理，这样能够提高SIFT检测的速度，将图片进行SIFT检测，得到若干个128维的特征点，这些特征点就可以看作构成该图像的单词，例如在下面这张图片中，SIFT检测到了257个特征点，即该图像由257个128维度的单词构成。对训练集中的所有图像进行SIFT特征提取，最终得到1133764个128维的特征点。



图3.2-1 图像的SIFT特征

**3.3 生成词袋**

通过使用SIFT算法，我们对2250张训练样本一共提取到1133764个特征点，但是这样的视觉词汇数量太庞大了，我们需要将其中差别不大的归为同一类，最终使用可以接受的数据量描述图像。这里就需要使用到聚类算法，本文采用K-Means算法。

K-Means算法又称K均值算法，属于聚类（clustering）算法的一种，是应用最广泛的聚类算法之一。其原理是要每个簇内每个点到中心点的距离的和最小，即：

具体实现步骤如下：

1. 随机选择k个样本作为初始簇类的均值向量；
2. 将每个样本数据集划分离它距离最近的簇；
3. 根据每个样本所属的簇，更新簇类的均值向量；
4. 重复（1）（2）步，当达到设置的迭代次数或簇类的均值向量不再改变时，模型构建完成，输出聚类算法结果。

使用K-Means算法可以将1133764个特征点聚类为50个视觉词汇，这样就可以使用这50个视觉词汇作为图像数据的词袋。

# 图像描述

**4.1 特征直方图**

在文本表示中，提取词袋后就需要对文档使用词袋进行描述。在图像处理中，根据上一步得到的词袋，下面需要将SIFT得到的特征点使用50维的词袋模型进行特征向量描述。算法思想是先生成长度为50的零向量，通过计算每个特征点到每个视觉词汇的距离，并找到距离最近的视觉词汇的索引，在零向量的索引下加一，这样就能将一张图片表示为一个50维的向量。

**4.2 使用SPM算法的特征直方图**

但是上述计算全图的分布特征会丢失图像的局部/细节信息，且视觉词汇的顺序会影响最终分类结果，因此可以使用SPM算法解决上述问题。SPM即Spatial Pyramid Matching，是一种利用空间金字塔进行图像匹配、识别、分类的算法。例如：在用“视觉词汇”频率直方图描述图像时，选择3种尺度。假设下图有100个特征，level0就是100维，共100个值。level1是把图像4等分切，单独计算每一块直方图，这就有4个直方图，共400个值，level2是16等分切原图，得到16个直方图，共1600个值。最后，用100+400+1600=2100个值来表示这个图像。

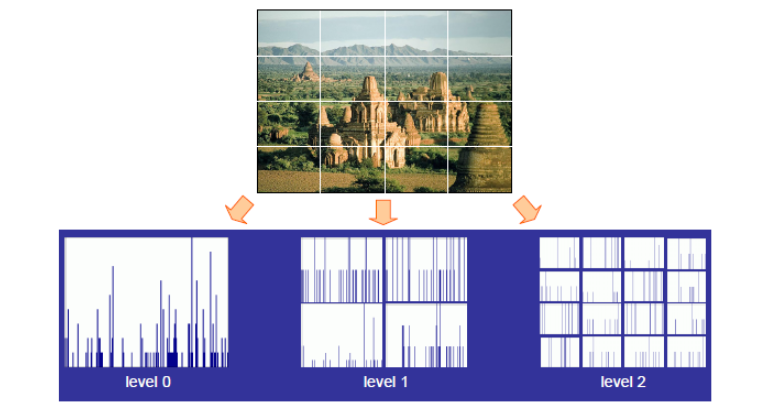


图4.2-1 SPM示意

# 训练

**5.1 分类器**

在机器学习中，分类器主要分为判别式和生成式两种类型。判别式分类器的目标是直接对类别进行建模，通过直接学习决策函数或者条件概率分布，思想是在有限样本条件下建立判别函数，不考虑样本的产生模型，直接研究预测模型，典型的例子包括逻辑回归、支持向量机和决策树等。相比之下，生成式分类器的目标是对每个类别的数据分布进行建模，由数据学习联合概率分布求出条件概率作为预测模型，即生成模型根据后验概率进行分类两者的主要区别在于，判别式分类器关注类别之间的边界，适用于边界明显的情况，而生成式分类器关注整个数据分布，适用于复杂数据情况。

通过聚类得到的词典，即可将图片的特征转化为对应的直方图。这样每幅图片就可以用一个词典向量来表示了，然后将这些向量作为分类器的输入进行训练即可得到分类器。本文的图像识别可以看作二分类问题，即判断当前图像是不是某个标签，输出的结果只有0和1，因此选择SVM（支持向量机）进行分类。

**5.2 SVM**

支持向量机（英语：Support Vector Machine，常简称为SVM）是一种监督式学习的方法，可广泛地应用于统计分类以及回归分析。

支持向量机将向量映射到一个更高维的空间里，在这个空间里建立有一个最大间隔超平面。在分开数据的超平面的两边建有两个互相平行的超平面。建立方向合适的分隔超平面使两个与之平行的超平面间的距离最大化。其假定为，平行超平面间的距离或差距越大，分类器的总误差越小。

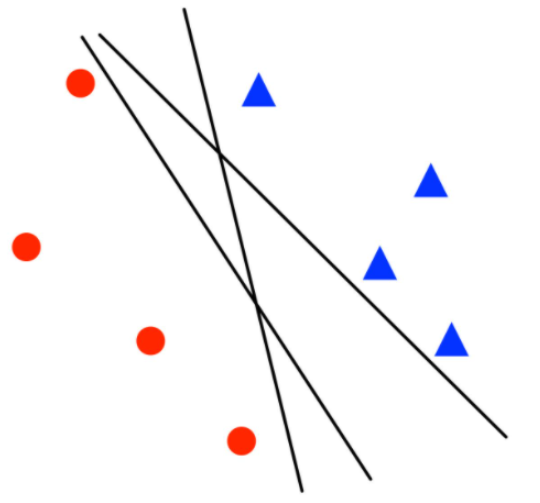


图5.2-1 SVM示意

# 结果分析

**6.1 预测评估**

测试图片使用scene\_categories 数据集每个类150张图片后面的所有图片，15各类共2235个样本，视觉词汇设计为100个，测试平台Intel i7-11800H,系统Windows11。

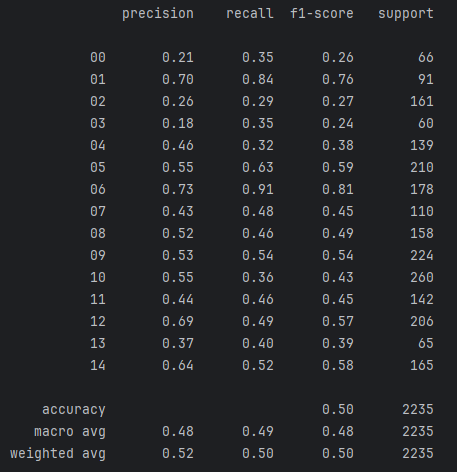
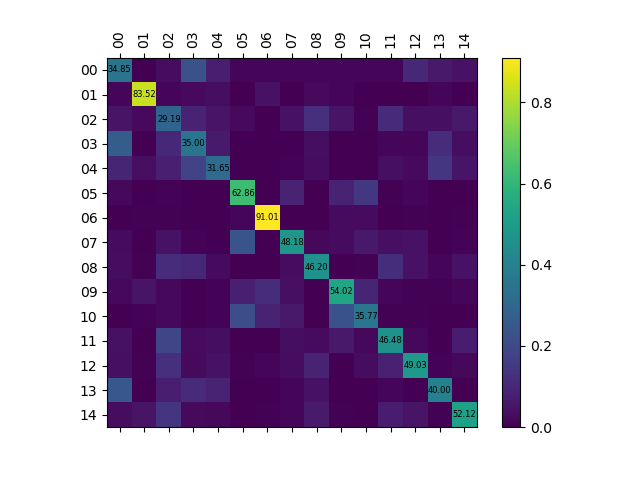
 

图6.1-1 评估结果 图6.1-2 混淆矩阵

**6.2 总结**

在研究过程中发现词袋中视觉词汇的数量对预测结果影响显著。通过调整KMeans聚类的类数，可以得到不同的分类效果，类数越大，分类效果越好，这反映了词袋模型中词汇数量对图像分类精度的影响。此外，特征点提取的算法也对分类精度有一定影响，该模型需要一种局部的、鲁棒性好的特征提取算法做支撑。

本文通过构建词袋模型和应用支持向量机分类器，详细了解了基于词袋模型的图像分类算法的实现步骤，并对词袋模型在实际数据集上进行了评估和分析。