

# 实验报告

## 算法整体流程

- 加载数据：**从指定文件夹读取图像数据，并按照给定的训练集大小分割成训练集和测试集。
- 特征提取：**使用SIFT算法从每张图像中提取特征。

特征提取步骤利用SIFT（尺度不变特征变换）算法从每张图像中提取特征。SIFT算法可以检测和描述图像中的局部特征，它对旋转、尺度缩放、亮度变化保持不变性，对视角变化具有一定的不变性。在提取过程中，SIFT算法首先检测关键点（keypoints），然后为每个关键点生成一个描述符（descriptor），描述符是一个128维的向量，描述了关键点周围的图像梯度信息。这些描述符将用于后续的图像分类。

- 构建视觉词汇表：**通过K-means算法对所有训练图像的SIFT描述符进行聚类，形成一个视觉词汇表。

构建视觉词汇表的过程类似于在文本分析中创建词汇表。在图像处理中，通过K-means聚类算法将所有训练图像的SIFT描述符进行聚类，形成视觉词汇（visual words）。每个聚类中心代表一个视觉词汇，所有图像的描述符都被映射到这些聚类中心，以此来简化和概括图像的特征信息。选择合适的聚类数（K值）是这一步骤的关键，因为它直接影响词汇表的粒度和最终分类的性能。

- SVM分类：**使用SVM进行分类训练和预测。

SVM（支持向量机）是一种强大的监督学习算法，用于分类和回归任务。在图像分类中，SVM通过在特征空间中找到一个最优的超平面，以此来区分不同的类别。SVM对于非线性和高维数据表现尤为出色，这得益于其使用核技巧（如RBF核），该技巧可以将数据映射到更高维的空间中，使得原本线性不可分的数据变得可分。在本项目中，通过将图像的特征直方图输入SVM，算法学习如何根据视觉词汇的分布来区分不同的图像类别。

- 性能评估：**生成并显示混淆矩阵和分类报告，用于评估模型性能。

## 各个函数的功能说明

- `load_images_from_folder(folder, train_size)`：**

- 功能：**从指定文件夹加载图像，并按照给定的比例分割成训练集和测试集。
- 输入：**

- `folder`: 文件夹路径。
  - `train_size`: 训练集中每类图像的数量。
  - 输出: 训练集和测试集的图像及其标签。
2. `extract_sift_features(X)`:
- 功能: 使用SIFT算法从图像中提取特征描述符和关键点。
  - 输入:
    - `X`: 图像数据列表。
  - 输出: 图像的描述符和关键点列表。
3. `show_features_in_img(X, keypoints)`:
- 功能: 在图像上绘制并显示关键点。
  - 输入:
    - `X`: 单个图像数据。
    - `keypoints`: 图像的关键点。
  - 输出: 显示含关键点的图像。
4. `build_vocabulary(descriptors_list, k)`:
- 功能: 使用K-means算法从SIFT描述符中创建视觉词汇表。
  - 输入:
    - `descriptors_list`: 所有图像的SIFT描述符列表。
    - `k`: 聚类中心的数量。
  - 输出: 训练好的K-means模型。
5. `features_to_histogram(features, vocab_model)`:
- 功能: 根据视觉词汇表将图像的SIFT描述符转换为特征直方图。
  - 输入:
    - `features`: 单个图像的SIFT描述符。
    - `vocab_model`: 训练好的K-means聚类模型。
  - 输出: 图像的特征直方图。
6. `plot_histogram(histogram)`:

- 功能：绘制特征直方图。
- 输入：
  - `histogram`：图像的特征直方图。

7. `plot_confusion_matrix(cm, class_labels, title="Confusion Matrix")`:

- 功能：绘制并显示混淆矩阵。
- 输入：
  - `cm`：混淆矩阵。
  - `class_labels`：类别标签。
- 输出：显示混淆矩阵。

## 实验结果

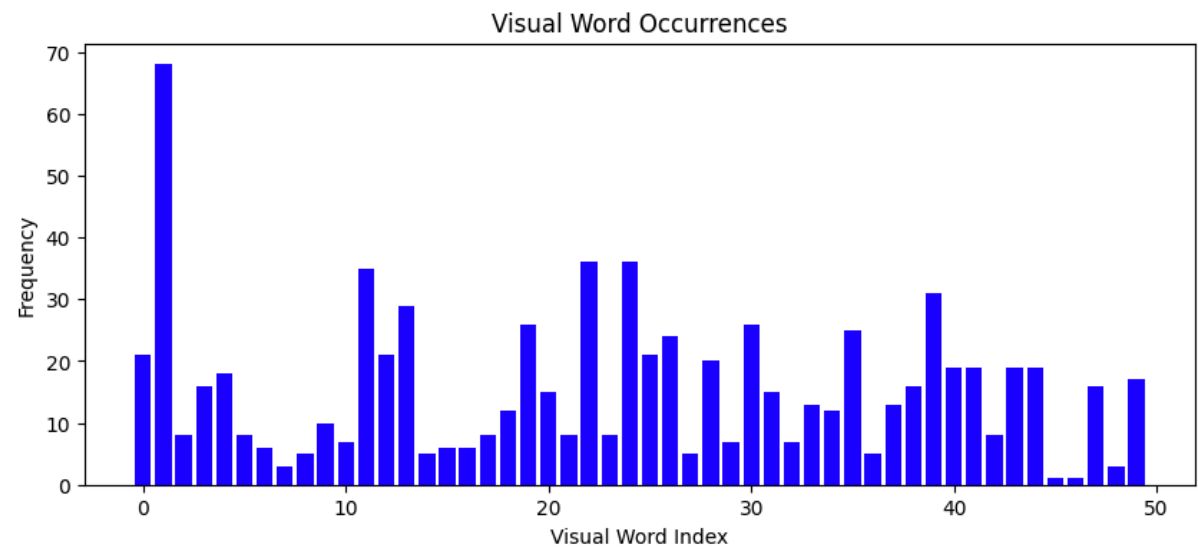
keypoints展示，能初步判断一下K值的数量级

Picture Keypoints



词库展示

能够看到某张图具体的词库分布，在这里归一化然后比较余弦距离可以一定程度上比较图像相似度



## 场景分类结果

本实验测试了K值分别为50，100，200时分类器的表现，结果如下

- 从结果来看，高斯核的svm相比同维度的线性核svm效果更好。
- 比较两个核对于k值变化的反应，线性核在k值变化的时候性能更加不稳定，因为其无法非线性的划分空间，对空间分布本身的情况比较敏感。
- 高斯核k值变化时表现稳定，并且有随着k值增加效果变好的趋势。

参数	precision	recall	f1-score
test(50, 'linear')	0.52	0.49	0.50
test(100, 'linear')	0.50	0.48	0.49
test(200, 'linear')	0.53	0.51	0.51
test(500, 'linear')	0.57	0.54	0.55
test(50, 'rbf')	0.56	0.54	0.54
test(100, 'rbf')	0.57	0.55	0.55
test(200, 'rbf')	0.58	0.57	0.56
test(500, 'rbf')	0.60	0.59	0.58

由于篇幅限制，下面仅展示效果最好的参数下的混淆矩阵和结果 (k=500,'rbf')

	precision	recall	f1-score	support
00	0.29	0.44	0.35	66
01	0.82	0.91	0.86	91
02	0.44	0.27	0.33	161
03	0.27	0.37	0.31	60
04	0.48	0.35	0.40	139
05	0.59	0.75	0.66	210
06	0.75	0.86	0.80	178
07	0.52	0.70	0.60	110
08	0.59	0.58	0.59	158
09	0.66	0.64	0.65	224
10	0.67	0.44	0.53	260
11	0.57	0.58	0.58	142
12	0.72	0.57	0.64	206
13	0.40	0.75	0.52	65
14	0.63	0.63	0.63	165
accuracy			0.59	2235
macro avg	0.56	0.59	0.56	2235
weighted avg	0.60	0.59	0.58	2235

