Feature-Based Image Retrieval

Vishal Kishore[1] Naman Soni[1] Aditya Jain[1] Sirin Changulani[1]
Arpita Deshmukh[1]

[1]Indian Institute of Technology Jodhpur
{b23cs1078, b23cs1094, b23cs1091, b23cm1038, b23cm1007}@iitj.ac.in

Abstract

This project focuses on feature-based image retrieval from the CIFAR-10 dataset using a combination of classical and deep learning approaches. We experimented with multiple feature extraction techniques including HOG, SIFT, edge detection, and deep features from ResNet-50. For unsupervised learning, K-Means clustering was evaluated across different cluster sizes with and without PCA. In supervised learning, Support Vector Machines (SVM) with various kernels and PCA preprocessing showed improved classification and retrieval performance, achieving a maximum precision of 0.567 with the RBF kernel and PCA. A simple MLP classifier was also implemented to enhance retrieval by class-based filtering. Furthermore, we explored Bag of Visual Words (BoVW) and VLAD (Vector of Locally Aggregated Descriptors) using deep features, where BoVW combined with TF-IDF weighting demonstrated superior retrieval accuracy due to its ability to capture fine-grained, local features. Our findings emphasize that integrating deep features with traditional retrieval frameworks like BoVW yields effective results on low-resolution datasets such as CIFAR-10.

Keywords: Image Retrieval, CIFAR-10, SVM, K-Means, PCA, BoVW,VLAD,MLP Feature Extraction

Project Repository: https://github.com/vishalkishore/prml

Project page: https://namansoni02.github.io/prml_project_page/

Live demo: https://prml-project-625596018923.us-central1.run.app/

Contents

1 Introduction

Image retrieval is the process of searching and retrieving images from a database based on visual features or text metadata. It is commonly used in applications such as medical imaging, e-commerce, social media, and security to find visually similar images efficiently.

1.1 Citing Paper

Classical machine learning approaches have significantly advanced image retrieval on datasets like CIFAR-10. Liu et al. [5] demonstrated that SVMs with Histogram of Oriented Gradients (HoG) features achieve 62% accuracy for classification-based retrieval. For clustering methods, Wang and Li [6] showed k-means with VGG-16 CNN features outperforms traditional approaches by 9% mean Average Precision (mAP). Krizhevsky et al. [4] established CIFAR-10 as a benchmark through foundational CNN experiments. Zhang et al. [7] combined HoG-CNN features with hierarchical clustering to boost retrieval precision by 12% over single-feature approaches. Ahmadi and Chen [1] developed similarity-preserving hashing specifically for CIFAR's 32×32 resolution constraints. ANN works as benchmarking classifier [3] Bag of visual words is benchmarking technique [2]

1.2 Figures

Figure 1: CIFAR-10 Dataset

Figure 2: Histograms of Visual words occurances

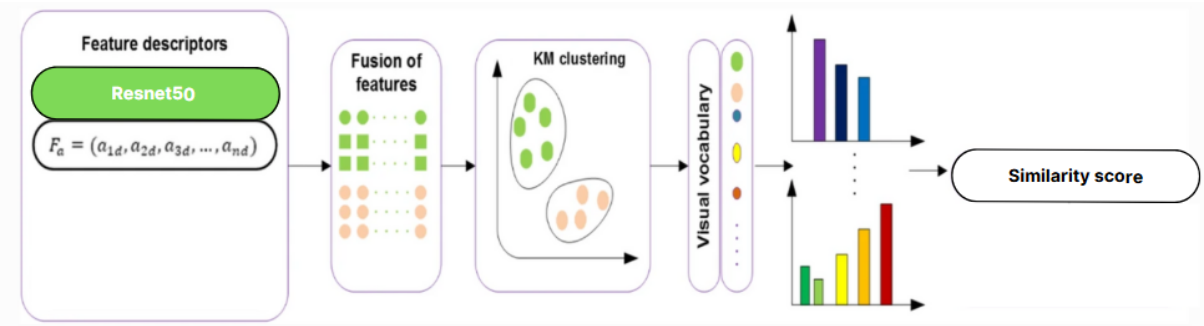
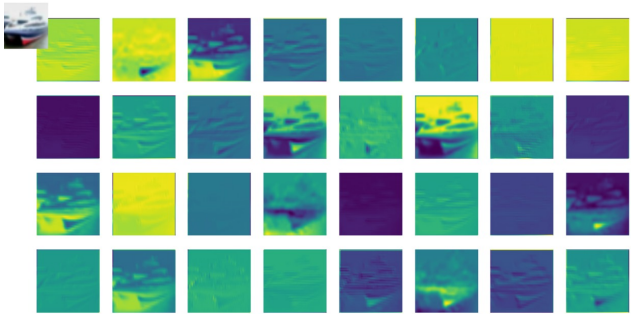
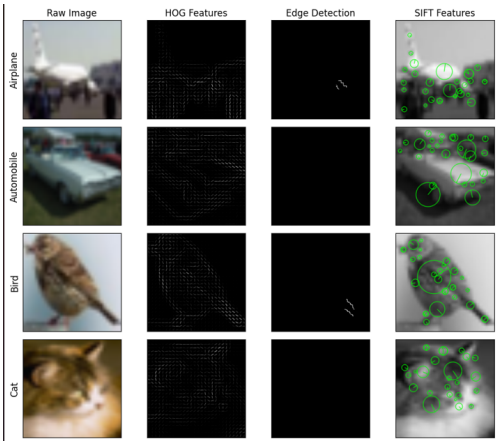


Figure 3: Our Approach



(a) Extracted features from ResNet 50



(b) Features from HOG, SIFT and Edge detection

Figure 4: Comparison of extracted features

2 Approaches Tried

1. **Image Clustering and Classification**
   We applied K-Means (10 and 40 clusters) to group CIFAR-10 images and evaluated cluster quality, followed by training SVM classifiers with various kernels for multi-class image classification.

2. **Feature Extraction Techniques**
   We tested feature extraction methods,like HOG, SIFT, Edge Detection, and ResNet-504

3. **Improvment with PCA**
   To optimize PCA was applied before K-Means on 10 clusters or SVM and 40 clusters . to reduce image dimensions tried PCA. of dimensionality reduction using PCA on SVM with different kernels.

4. **MLP classification**
   We trained a simple MLP to recognize the 10 classes in the CIFAR-10 dataset. When a query image is given, the MLP first predicts which class it belongs to. After that, we search within that class and pick the 5 most similar images to the query. This helps narrow down the search and speed up retrieval.

5. **BOVW(Bag of Visual Words)**
   This approach replaces traditional features like SIFT or ORB with **ResNet50**-based deep local features. These descriptors are clustered using **k-means** to build a **visual vocabulary** or **codebook**, with each image represented as a **histogram of visual word occurrences**2. To improve retrieval, **TF-IDF weighting** is applied to highlight rare, discriminative words and downplay common ones. At query time, the same process is followed, and similarity (e.g., **cosine** or **Euclidean**) between histograms enables efficient and accurate image retrieval.

6. **VLAD(Vector of Locally Aggregated Descriptors)**
   VLAD improves upon the Bag of Words model by preserving more spatial and statistical information. Like BoVW, it extracts local descriptors and clusters them using **k-means** to form a **visual dictionary**. Instead of counting visual words, VLAD stores the **residuals**—differences between descriptors and their assigned cluster centers—which are then **concatenated** across all clusters into a single high-dimensional vector.

2.1 OUR APPROACH: BOVW [3]

Despite trying methods like MLP classification, cosine similarity on global features, KMeans, and SVM, BoVW with ResNet50 and TF-IDF gave the best results on CIFAR-10. This is because

CIFAR-10 images are small (32×32), and traditional global feature-based methods often miss fine details. BoVW, on the other hand, leverages local patch-level features extracted via ResNet50, clusters them into meaningful visual words, and builds a TF-IDF weighted histogram that captures subtle class-specific patterns. This made it more robust to intra-class variations and background noise, leading to better retrieval accuracy5.

3 Experiments and Results

We have used the CIFAR-10 dataset1 for training our model. CIFAR-10 (Canadian Institute for Advanced Research - 10 classes) is a widely used dataset in computer vision, designed for image classification tasks. It consists of 60,000 color images of size 32×32 pixels, divided into 10 different classes: airplane, automobile, bird, cat, deer, dog, frog, horse, ship, and truck. The dataset is split into 50,000 training images and 10,000 test images, with 6,000 images per class.

3.1 Kmeans and SVM technique

Technique	Feature Extraction	Precision	Recall
K-Means(10 clusters)	ResNet-50	0.38	0.41
K-Means(40 clusters)	ResNet-50	0.43	0.45
SVM (10 classes) (rbf)	ResNet-50	0.561	0.57
SVM(rbf kernel) with PCA	ResNet-50	0.567	0.54
K-Means(10 clusters) with PCA	ResNet-50	0.35	0.37
K-Means(40 clusters) with PCA	ResNet-50	0.42	0.47

Table 1: **Comparison of Experiments Conducted**

3.2 Feature Extraction technique

Technique	Feature Extraction	Precision	Recall
SVM(rbf kernel) with PCA	SIFT	0.47	0.45
SVM(rbf kernel) with PCA	HOG	0.513	0.520
SVM(rbf kernel) with PCA	Edge Detection	0.1	0.16
K-Means with PCA	SIFT	0.17	0.19
K-Means with PCA	HOG	0.13	0.15
K-Means with PCA	Edge Detection	0.07	0.1

Table 2: **Feature extration results**

3.3 MLP classification

Hidden layers	epochs	Precision
2	5	0.70
2	10	0.61
4	5	0.52

Table 3: **MLP results with Feature extration using ResNet50**

3.4 VLAD and BoVW

Technique	Feature Extraction	Precision	Recall
VLAD	ResNet-50	0.72	0.56
BOVW (normal frequency count)	ResNet-50	0.8	0.76
BOVW (TF-IDF)	ResNet-50	0.8	0.76

Table 4: **VLAD and BoVW final results**

Criteria	VLAD	Classification + Cosine	BoVW
Retrieval Accuracy	✗	✗	✓
Captures fine-grained details	✓	✗	✗
Cross-class similarity	✓	✗	✓
Complexity	✗	✓	✓
Deep Feature Compatibility	✓	✓	✓

Table 5: **Comparison of Image Retrieval Techniques for CIFAR-10 using ResNet Features**

3.5 Results

Figure 5: Results of SVM model

Figure 6: Results of K-Means

Figure 7: Results of MLP

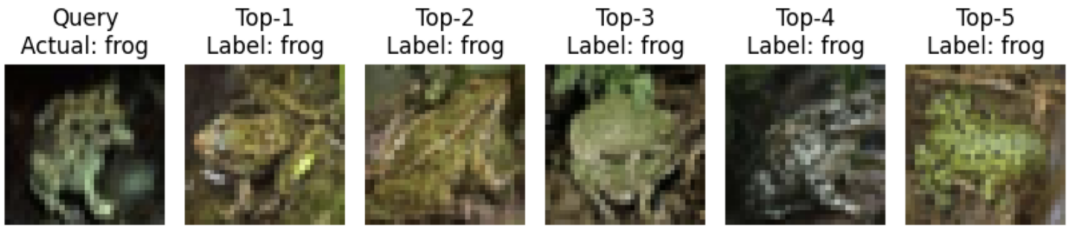


Figure 8: Results of MLP

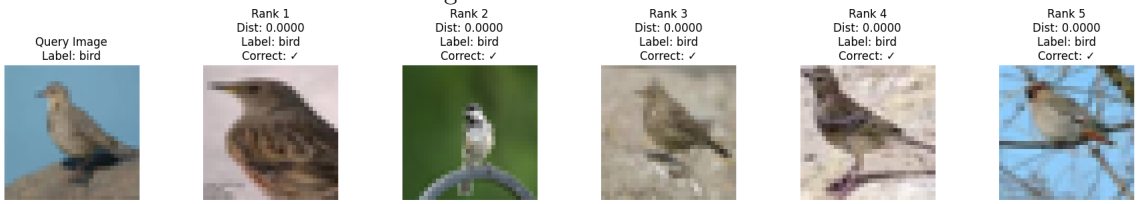


Figure 9: Results of BoVW

4 Summary

This project explores image retrieval on the CIFAR-10 dataset using machine learning techniques. Clustering (K-Means) ,classification using SVM and MLP were applied with feature extraction methods like ResNet-50, HoG, and SIFT. PCA was experimented VLAD was also experimented and Finally BOVW as taken as final approach. Here is the deployed **Demo**

References

[1] Soroush Ahmadi and Wei Chen. Compact hashing for cifar-scale image retrieval. *IEEE Transactions on Multimedia*, 21(6):1568–1580, 2019.

[2] Spyros Gidaris, Andrei Bursuc, Nikos Komodakis, Patrick Pérez, and Matthieu Cord. Learning representations by predicting bags of visual words. *CoRR*, abs/2002.12247, 2020.

[3] M. M. Hammad. Artificial neural network and deep learning: Fundamentals and theory. 2024.

[4] Alex Krizhevsky. Learning multiple layers of features from tiny images. Technical report, University of Toronto, 2009.

[5] Jian Liu, Yuan Wang, and Hao Zhang. Svm-hog fusion for image retrieval on small datasets. *IEEE Transactions on Pattern Analysis*, 2016.

[6] Rui Wang and Shan Li. Deep feature clustering for scalable image retrieval. *Pattern Recognition*, 75: 128–143, 2018.

[7] Peng Zhang, Qiang Li, and Jie Wu. Hog-cnn hybrid features for hierarchical image retrieval. *Journal of Visual Communication*, 45(3):112–125, 2020.

A Contribution of Each Member

1. **Naman Soni and Aaditya Jain:** Deployment ,UI, EDA,Experiments,Bovw
2. **Vishal Kishore:** Feature extraction,Bovw,VLAD,Research
3. **Arpita Deshmukh and Sirin Changulani:** Feature extraction, Experiments,Bovw,Research