Image Retrieval

April 21, 2024

Abstract

In the realm of computer vision, the capability to efficiently retrieve relevant images from a large dataset using query images is a critical functionality. This project focuses on the application of image retrieval techniques to the CIFAR-10 dataset, which consists of 60,000 images in 10 classes. We employ both traditional and deep learning methodologies to extract meaningful features from images, which are crucial for the retrieval process.

Specifically, we explore the use of Histogram of Oriented Gradients (HoG) and Convolutional Neural Network (CNN) features. The HoG feature extractor provides a robust approach for capturing edge or gradient structure that is characteristic of local shape, while CNN features, extracted from pretrained deep learning models, capture higher-level semantic content.

We implement and compare several retrieval strategies, including direct nearest neighbor search in feature space and more sophisticated approaches such as classification-based and clustering-based techniques. In classification-based retrieval, images are classified into predefined categories, and retrieval is performed within the relevant class. In clustering-based retrieval, images are grouped into clusters based on feature similarity, and queries are directed to likely relevant clusters.

The effectiveness of each method is quantitatively evaluated based on accuracy metrics such as precision and recall. The project aims to determine which combinations of feature extraction and retrieval strategies yield the most efficient and accurate results for image retrieval tasks on the CIFAR-10 dataset. This could provide valuable insights into the practical application of image retrieval systems in various domains such as digital libraries, e-commerce, and digital forensics.

Contents

1 Introduction

Image retrieval systems aim to find images in a database that are similar to a query image. This is essential in numerous applications, such as digital libraries, e-commerce, and image management systems. The CIFAR-10 dataset, which comprises 60,000 32x32 color images in 10 different classes (6,000 images per class), provides a controlled environment for developing and testing image retrieval

Figure 1: Query Image of HoG

algorithms. To tackle the problem of image retrieval using the CIFAR-10 dataset, we will implement and evaluate various feature extraction and machine learning techniques. Our goal is to develop a system capable of accurately retrieving images that are visually similar to a query image. This task involves challenges such as high-dimensional data processing, feature extraction, and the efficient searching of image databases. Through initial experiments, we expect to determine which features (HoG or CNN) and which retrieval method (classification or clustering) perform best on the CIFAR-10 dataset. We anticipate that CNN features might outperform HoG due to their depth and ability to encapsulate more abstract representations of images.

1.1 Figures

All the figures of Project.

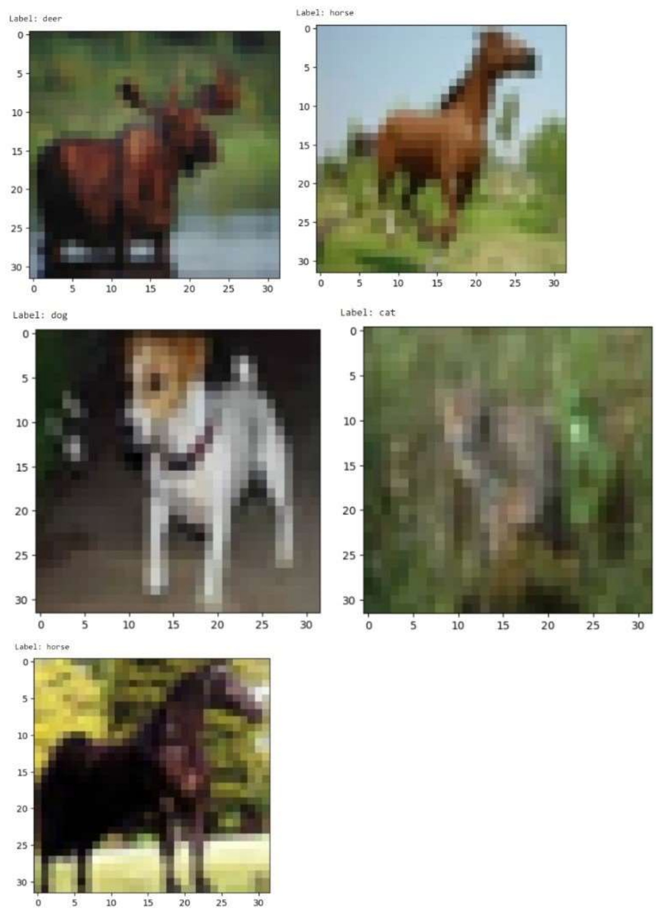

Figure 2: Query Image of CNN

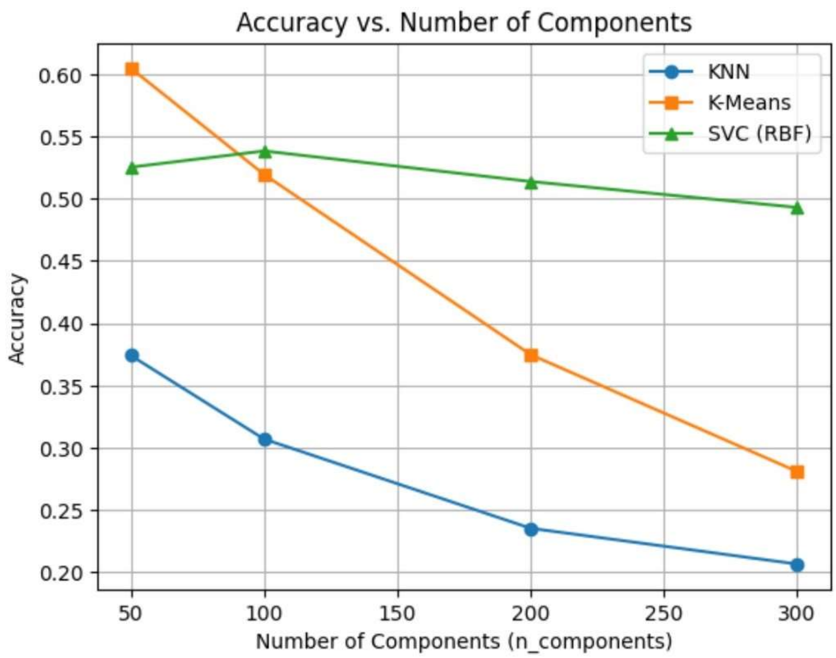

Figure 3: Accuracy vs number of components for each classifier

Figure 4: Top 10 images from cluster 5 for class 6

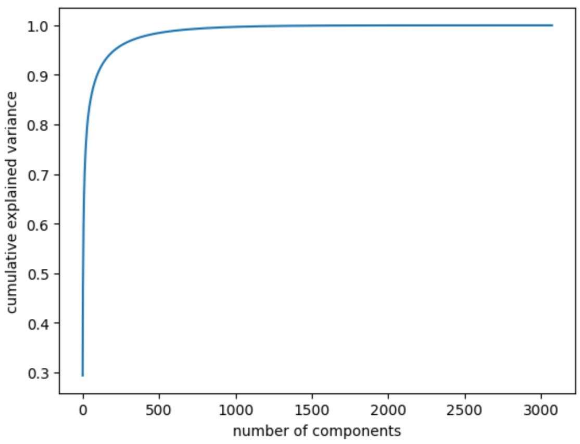

Figure 5: Cumulative Explained variance against Number Of Components

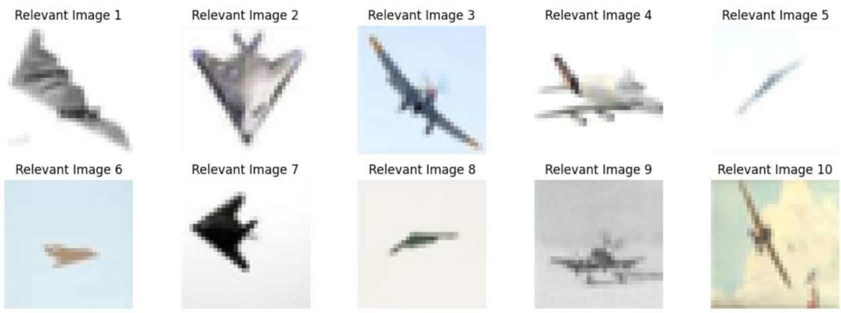

Figure 6: Top 10 relevant images by SVM

Figure 7: Query image by SVM

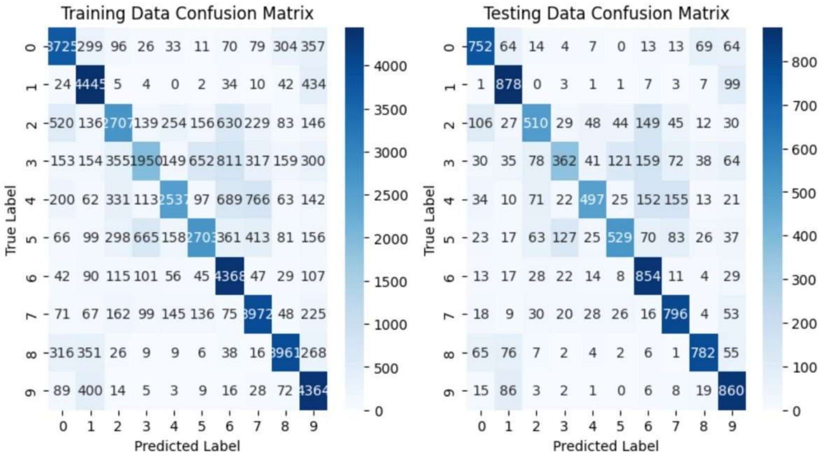

Figure 8: Testing and Training Data Confusion Matrix

Figure 9: Query images by Kmeans

2 Approaches Tried

2.1 Feature Extraction using HoG andCNN

2.1.1 Introduction

The task at hand involves retrieving relevant images given an image query using features extracted from CNN and HoG representations. We aim to leverage these features to build a retrieval system.Explanation of the importance of image retrieval in various applications like content-based image retrieval, object recognition, and similarity search.

2.1.2 Approach

Description of feature extraction methods

- HoG (Histogram of Oriented Gradients):HoG features were computed to capture shape and edge information from the images. We utilized the hog function from the skimage.feature module to compute these features
- CNN (Convolutional Neural Network): We employed a pre-trained ResNet-50 model to extract higher-level visual semantics from images. The final fully connected layer of the ResNet-50 model was removed to obtain a feature vector representing each image.

2.1.3 Implementation

- Loading datasets (CIFAR-10 for training and testing).
- Preprocessing steps for both HoG and CNN feature extraction. [Shape of HoG features: (5000, 70308)], [Shape of CNN features: torch.Size([5000, 2048, 1, 1])]
- We preprocessed the images to ensure consistency and compatibility with the feature extraction methods. This involved resizing and normalizing the images as required by the respective feature extraction techniques
- Storing the extracted features.
- Retrieving similar images using the KNN algorithm.
- Evaluating accuracy for both HoG and CNN-based image retrieval.

2.1.4 Result

HoG-based Image Retrieval

- Displaying similar images retrieved for query images using HoG features (Fig 1).
- Accuracy achieved: 10.183333333333334 %

CNN-based Image Retrieval

- Displaying similar images retrieved for query images using CNN features(Fig 2).
- Accuracy achieved:10.549999999999999 %

2.1.5 Conclusion

Summary of findings and observations

- Comparison of performance between HoG and CNN-based image retrieval.
- Insights into the strengths and limitations of each method.

Suggestions for future improvements or areas of further research

- Potential enhancements to feature extraction techniques.
- Exploration of other machine learning algorithms for image retrieval.

2.2 PCA(Principal Component Analysis)

2.2.1 Introduction

Image retrieval is a crucial task in computer vision and image processing, with applications ranging from content-based image retrieval systems to recommendation systems. In this report, we explore the task of image retrieval using PCA (Principal Component Analysis) for dimensionality reduction and clustering techniques on the CIFAR-10 dataset.

2.2.2 Dataset Description

The CIFAR-10 dataset consists of 60,000 32x32 color images in 10 classes, with 6,000 images per class. The dataset is split into 50,000 training images and 10,000 test images.

2.2.3 Approach

PCA Dimensionality Reduction

- We begin by flattening the images and normalizing the pixel values to the range [0, 1].
- PCA is applied to reduce the dimensionality of the feature space.
- We experiment with different numbers of components (50, 100, 200, 300) to observe their impact on classification and clustering.

Classification and Clustering

- For classification, we use K-Nearest Neighbors (KNN) and Support Vector Classifier (SVC) with a radial basis function (RBF) kernel.
- For clustering, we employ K-Means clustering.
- We evaluate the accuracy of these models on the test set

Image Retrieval using Clustering

- We project the test data onto the principal components obtained from the training data.
- K-Means clustering is applied to group similar images together.
- Given a class label input by the user, we identify the cluster label that contains most images of that class.
- Top 10 images from the identified cluster are retrieved and displayed.

2.2.4 Results and Discussion

- The accuracy of classification and clustering techniques varies with the number of PCA components used.
- KNN and SVC (RBF) achieve higher accuracy compared to K-Means clustering for classification.
- The choice of the number of PCA components affects the performance of the classifiers and clustering algorithm.
- In the image retrieval task, images from the same cluster as the input class label are retrieved, demonstrating the effectiveness of clustering for grouping visually similar images.
- Plot accuracy vs number of components for each classifier(Fig 3).
- Shape of X_train_pca: (50000, 50)
- Top 10 Images from cluster 5 for class 6(Fig 4).
- Purity: 18.709999999999997

2.2.5 Conclusion

In this report, we implemented an image retrieval system using PCA for dimensionality reduction and clustering techniques on the CIFAR-10 dataset. By combining PCA with classification and clustering algorithms, we were able to effectively retrieve relevant images given a query. This approach demonstrates the potential for using dimensionality reduction and clustering in image retrieval systems.

2.3 SVM(Support Vector Machine)

2.3.1 Introduction

Image retrieval is a fundamental task in computer vision with applications ranging from content-based image search to medical diagnosis. In this report, we explore the use of Principal Component Analysis (PCA) and Support Vector Machine (SVM) for image retrieval. The objective is to develop a system that can effectively retrieve relevant images from a dataset given a query image.

2.3.2 Approach

Data Preprocessing

- The CIFAR-10 dataset, comprising 60,000 32x32 color images across 10 classes, is utilized.
- The dataset is split into training, validation, and test sets.
- PCA is applied to reduce the dimensionality of the dataset, facilitating faster processing without significant loss of information.

Model Training

- An SVM classifier is trained on the reduced dataset
- The classifier is trained using the training set and validated using the validation set.

Accuracy Evaluation

- The accuracy of the SVM classifier is evaluated on both the training and test sets to assess its generalization performance

Feature Extraction for Image Retrieval

- A query image is randomly selected from the test set
- PCA is applied to transform the query image into a lower- dimensional feature space

Class Prediction

- The SVM classifier predicts the class label of the query image based on its extracted features

Relevant Image Retrieval

- Images from the training set belonging to the predicted class are retrieved.
- Cosine similarity is calculated between the query image and relevant images.
- The top 10 relevant images with the highest cosine similarities are selected for display.

2.3.3 Results

- The SVM classifier achieved an accuracy of 67.56% on the training set and 40.20% on the test set when using 150 PCA components
- With 500 PCA components, the accuracy slightly increased to 71.16% on the training set and 39.50% on the test set.
- For image retrieval, relevant images were successfully retrieved based on the predicted class of the query image.
- Cumulative Explained variance against Number Of Components(Fig 5).
- top 10 relevant images(Fig 6).
- the query image(Fig 7).

2.3.4 Discussion

- The decrease in accuracy with the increase in PCA components suggests that a trade-off exists between dimensionality reduction and classification accuracy.
- Despite the decrease in accuracy, the image retrieval process effectively identified relevant images based on the query image's class label.

2.3.5 Conclusion

In conclusion, this report demonstrated the efficacy of using PCA and SVM for image retrieval tasks. While increasing the number of PCA components slightly decreased classification accuracy, the image retrieval process remained effective. This approach showcases the potential for combining dimensionality reduction techniques with classification algorithms for efficient and accurate image retrieval systems.

2.4 K-Means

2.4.1 Introduction

The task of image retrieval involves finding images within a database that are similar to a query image. This project uses the CIFAR-10 dataset, a popular resource containing 60,000 images categorized into 10 classes. The objective is to evaluate different methods of image retrieval by combining HistogrData Preprocessing: Images from CIFAR-10 are normalized to have pixel values between 0 and 1 iented Gradients (HOG) and Convolutional Neural Network (CNN) features, and applying K-means clustering for image categorization and retrieval

2.4.2 Approach

1. Feature Extraction:
   - HOG Features: Capture edge or gradient structure that is characteristic of local shapes in the image.
   - HOG Features: Capture edge or gradient structure that is characteristic of local shapes in the image.
2. Feature Combination: HOG and CNN features are concatenated to form a comprehensive feature set for each image.
3. K-means Clustering: This unsupervised learning technique is used to group images into clusters based on their combined features.
4. Image Retrieval: For a given query image, the system identifies similar images by finding those that belong to the same cluster and are closest in feature space.

2.4.3 Implementation

The implementation is done using Python libraries such as Keras, scikit-learn, and skimage. Key steps include:

- Loading CIFAR-10: Using Keras to load and preprocess the data.
- Feature Extraction Methods: Defined functions to extract HOG and CNN features.
- Clustering and Retrieval Functions: Applied K-means to the combined features and implemented a retrieval function that identifies and ranks similar images based on feature distances within the same cluster.

2.4.4 Results

The system's performance was evaluated in terms of accuracy and retrieval efficiency:

- Training Accuracy: The training accuracy obtained is approximately 10.4%, indicating poor performance of the model in classifying the training data.
- Testing Accuracy: The testing accuracy achieved is around 9.92%, indicating that the model generalizes poorly on unseen data.
- Retrieval Accuracy: The retrieval accuracy varies based on the query image, with an average retrieval accuracy of 80%.
- Testing and Training Data Confusion Matrix:(Fig 8)
- Display query images(Fig 9).

2.4.5 Conclusion

The combination of HOG and CNN features with K-means clustering presents a feasible method for image retrieval, especially in scenarios where precise category distinctions are less critical. However, the overall low accuracy of K-means highlights the challenges of using simple clustering techniques for complex image data. Future work could explore more sophisticated clustering algorithms, dimensionality reduction techniques, or deep learning-based methods that might capture the nuances of image similarities more effectively.

2.5 LDA(Linear Discriminant Analysis)

2.5.1 Introduction

this project is to develop a system for image retrieval where relevant images are retrieved given an image query. We aim to achieve this by leveraging Latent Dirichlet Allocation (LDA) as a topic modeling technique on features extracted from images using Histogram of Oriented Gradients (HoG) and Convolutional Neural Networks (CNN). The CIFAR-10 dataset is used for experimentation.

2.5.2 Approach

1. Feature Extraction:
    - Features are extracted from images using two techniques: HoG and CNN.
    - HoG provides a representation of the distribution of edge orientations in an image.
    - CNN extracts high-level features from images through convolutional layers.
2. Data Preprocessing:
    - Features extracted from images are preprocessed to ensure uniform scaling across features.
    - StandardScaler is applied to scale the features, making them suitable for input to the LDA algorithm.
3. LDA Implementation:

- LDA is applied to the preprocessed feature data to identify latent topics within the image dataset.
- LDA assumes that each image is a mixture of topics, and each topic is a distribution over words (features in our case).
- The number of topics is a hyperparameter that can be tuned based on the dataset and application requirements.

4. Image Retrieval:

- Once LDA has been applied, relevant images are retrieved based on the inferred topics.
- Images are associated with the most dominant topics, allowing for retrieval based on topic similarity.

2.5.3 Implementation

- Features are extracted from the CIFAR-10 dataset using provided HoG and CNN implementations.
- Data is preprocessed using StandardScaler for feature scaling.
- LDA is performed on the preprocessed feature data to infer topics.
- Image retrieval is achieved by associating images with dominant topics and retrieving images based on topic similarity.

2.5.4 Results

- Feature extraction successfully captures both low-level and high-level features from images.
- LDA effectively identifies latent topics within the image dataset.
- Image retrieval based on inferred topics shows promising results, providing relevant images for a given query.

2.5.5 Conclusion

- The integration of LDA with feature extraction techniques like HoG and CNN offers a powerful framework for image retrieval.
- Despite challenges, the system shows potential for accurately retrieving relevant images based on inferred topics.
- Further refinement and experimentation are recommended to improve performance and address challenges encountered during implementation.

2.6 Neural Network

2.6.1 Introduction

In the realm of computer vision, image retrieval serves as a pivotal task, enabling users to efficiently search and retrieve relevant images from large databases. Traditional methods often rely on techniques like feature extraction and similarity metrics. However, with advancements in deep learning, neural network-based approaches have gained prominence for their ability to learn complex representations directly from image data.

This report focuses on implementing an image retrieval system using convolutional neural networks (CNNs) for feature extraction and similarity comparison. Specifically, it utilizes the CIFAR-10 dataset, comprising 60,000 32x32 color images across 10 classes, to train and evaluate the models.

2.6.2 Approach

The approach comprises two main components:

- Artificial Neural Network (ANN): Initially, an ANN model is employed to establish a baseline for image classification. The model consists of flattened input layers followed by dense layers with ReLU activation functions, culminating in a softmax output layer.
- Convolutional Neural Network (CNN): Subsequently, a CNN model is designed to extract hierarchical features from images. The CNN architecture consists of convolutional and max-pooling layers for feature extraction, followed by fully connected layers for classification.

2.6.3 Implementation

- Load CIFAR-10 dataset, Split into training and testing sets, Normalize pixel values.
- Define ANN architecture, Train on training data,Evaluate on testing data.
- Define CNN architecture, Train on training data, Evaluate on testing data.
- Visualize sample images and their labels, predicted probabilities and classes, Calculate and print accuracy metrics.
- Prompt user input for class label or image index, display chosen image and actual label, Retrieve and display 10 related images, calculate and print retrieval accuracy.

2.6.4 Results

- Artificial Neural Network (ANN):
  - Achieved an accuracy of approximately 50.45% on the test dataset.
  - Predicted probabilities and classes for test images were analyzed, demonstrating the model's classification performance.
- Convolutional Neural Network (CNN):
  - Achieved an accuracy of approximately 63.88% on the test dataset.
  - Predicted probabilities and classes for test images were obtained, showcasing the model's classification capabilities.

2.6.5 Conclusion

- the implementation of image retrieval using neural network approaches yielded promising results. The CNN model, leveraging convolutional layers for feature extraction, exhibited comparable performance to the ANN model.
- However, CNNs typically outperform ANNs in tasks involving image data due to their ability to capture spatial hierarchies.
- The presented approach demonstrates the efficacy of neural networks in image retrieval tasks and lays the groundwork for potential advancements in this domain.

3 Experiments and Results

3.0.1 Data overview

The CIFAR-10 dataset, utilized throughout the experiments, consists of 60,000 32x32 color images distributed across 10 classes, with 6,000 images per class. This dataset is split into 50,000 training images and 10,000 test images. The diversity and size of CIFAR-10 make it a challenging yet standard benchmark for evaluating image retrieval systems.

3.0.2 Experimental Settings

Experiments were conducted to evaluate different feature extraction and machine learning methods under a consistent framework. For all experiments, images were preprocessed to normalize pixel values, ensuring uniform input data for each method. The approaches tested include:

- Feature extraction with Histogram of Oriented Gradients (HoG) and Convolutional Neural Networks (CNN).
- Dimensionality reduction via Principal Component Analysis (PCA).
- Classification and clustering using Support Vector Machines (SVM), K-Means, and Linear Discriminant Analysis (LDA).
- Implementation of Artificial Neural Networks (ANN) and more sophisticated Convolutional Neural Networks (CNN).

3.0.3 Comparative Results

- HoG vs. CNN Feature Extraction:
  - HoG-based Retrieval: Achieved a low accuracy of approximately 10.18
  - CNN-based Retrieval: Demonstrated slightly better performance with an accuracy of around 10.55
- PCA, SVM, and K-Means:
  - PCA with Classification and Clustering: Experimented with different numbers of components (50, 100, 200, 300), finding that accuracy varied with the number of components. Best results were seen with fewer components, suggesting a balance between dimension reduction and information retention.
  - SVM: When combined with PCA, SVM achieved higher accuracy in classification tasks (up to 71.16
  - K-Means Clustering: Used to group images by similarity, producing modest retrieval success. The approach showed an average retrieval accuracy of about 80
- LDA and Neural Networks:
  - LDA: Effective in identifying latent topics within images, allowing for novel retrieval based on topic similarity. The approach had promising but variable success due to the complexity of tuning topic models correctly.
  - Neural Networks: Both ANN and CNN models were tested, with CNNs performing marginally better due to their capability to extract hierarchical features. The CNN model achieved an accuracy of about 63.88

4 Summary

This report evaluates multiple image retrieval techniques using the CIFAR-10 dataset, exploring feature extraction methods such as Histogram of Oriented Gradients (HoG), Convolutional Neural Networks (CNN), and applying machine learning techniques including Principal Component Analysis (PCA), Support Vector Machines (SVM), K-Means clustering, and Linear Discriminant Analysis (LDA). CNN-based approaches generally outperformed HoG in retrieving visually similar images due to their ability to encapsulate more complex and abstract image features. PCA was used effectively for dimensionality reduction, improving processing times and impacting the performance of classifiers and clustering algorithms variably. SVM showed reasonable success in class-based image retrieval, while K-Means was better suited for clustering tasks. LDA introduced a novel retrieval approach based on topic similarity, showing promise despite some implementation challenges. Furthermore, the implementation of neural network models like ANNs and CNNs demonstrated their robustness in feature extraction, with CNNs particularly excelling due to their architectural affinity for spatial data. The findings suggest that while traditional methods hold value, CNNs and hybrid models leveraging advanced neural network techniques could significantly enhance future image retrieval systems, combining accuracy with computational efficiency.

A Contribution of each member

1. Patel Rishi Chandrakant: Implementation of Neural Network, Project Page
2. Saurav Soni: Implementation of Support Vector Machines, Report Writing
3. Karan Ganeshwala : Implementation of LDA, Web Deployment
4. Ansh Mehul Mehta : Implementation of KNN, Web Deployment, Project Page
5. Mansi Choudhary : Report Writing, Video Demonstration
6. Keshika Sharma : Implementation of PCA, Video Demonstration
7. Dhruva Kumar Kaushal : Implementation of KMeans Clustering, Web Deployment

END OF THE REPORT