PROJECT REPORT

Problem Statement: Deep Learning Project: AI-Powered Image Similarity Search and Recommendation System

Abstract

This project presents an unsupervised image-based recommendation and retrieval system using deep feature embeddings extracted from a pretrained convolutional neural network. A ResNet-50 model pretrained on ImageNet is employed as a feature extractor to generate compact and discriminative image representations. Without using class labels during training, the system computes similarity between images using cosine similarity on L2-normalized embeddings. The approach is evaluated on the CIFAR-10 dataset, demonstrating that meaningful semantic similarity can be achieved through transfer learning and unsupervised similarity learning. The results highlight the effectiveness of deep embeddings for content-based image retrieval tasks.

Introduction

- With the rapid growth of digital image repositories, efficient image retrieval and recommendation systems have become essential.
- Traditional text-based or metadata-driven search methods are insufficient for large-scale visual databases.
- Deep learning enables content-based image retrieval by learning rich visual representations directly from images.
- This project addresses the challenge of finding visually and semantically similar images without relying on labeled training data.
- The proposed system leverages pretrained deep neural networks to perform similarity-based image recommendation in an unsupervised manner.

Problem Statement

Deep Learning Project: AI-Powered Image Similarity Search and Recommendation System

- To design and implement an AI-powered image similarity search and recommendation system using deep learning techniques.
- To extract meaningful and discriminative visual features from images without supervised training.
- To measure similarity between images based on learned feature embeddings rather than class labels.
- To retrieve and recommend the most visually similar images for a given query image.
- To evaluate the effectiveness of the system using appropriate retrieval performance metrics.

About Dataset and Handling

- **Dataset Used**: CIFAR-10
    - Contains 60,000 color images of size 32×32 pixels.
    - Consists of 10 semantic classes: airplane, automobile, bird, cat, deer, dog, frog, horse, ship, and truck.
    - Each class has 6,000 images.
- **Data Split**:
    - The test split (10,000 images) is used for the entire experiment.
- **Preprocessing Steps**:
    - Images are resized to 224×224 to match the input size of ResNet-50.
    - Pixel values are normalized using ImageNet mean and standard deviation.
    - Images are converted into tensors compatible with PyTorch.
- **Label Usage**:
    - Class labels are not used during feature extraction or retrieval.
    - Labels are used only during evaluation to compute Precision@K.

Model Architecture

- **Backbone Network**: ResNet-50
    - A 50-layer deep convolutional neural network.
    - Utilizes residual connections to avoid vanishing gradient problems.
- **Pretraining**:
    - Model is pretrained on the ImageNet dataset.
- **Architecture Modification**:
    - The final fully connected classification layer is removed.
    - The output becomes a 2048-dimensional feature vector.

- **Feature Normalization**:
    - L2 normalization is applied to embeddings.
    - Ensures cosine similarity corresponds to angular distance in embedding space.

Method of Implementation

1. **Data Preparation**:
    - CIFAR-10 dataset is loaded using torchvision.
    - Standard ImageNet preprocessing is applied.
2. **Feature Extraction**:
    - Each image is passed through the ResNet-50 backbone.
    - A 2048-dimensional embedding is generated per image.
    - Embeddings are L2-normalized.
3. **Embedding Storage**:
    - Extracted embeddings and corresponding images are stored using Joblib.
4. **Similarity Computation**:
    - Cosine similarity is computed between a query image embedding and all database embeddings.
5. **Recommendation Generation**:
    - Images are ranked based on similarity scores.
    - Top-K most similar images are retrieved for a given query image.

Evaluation

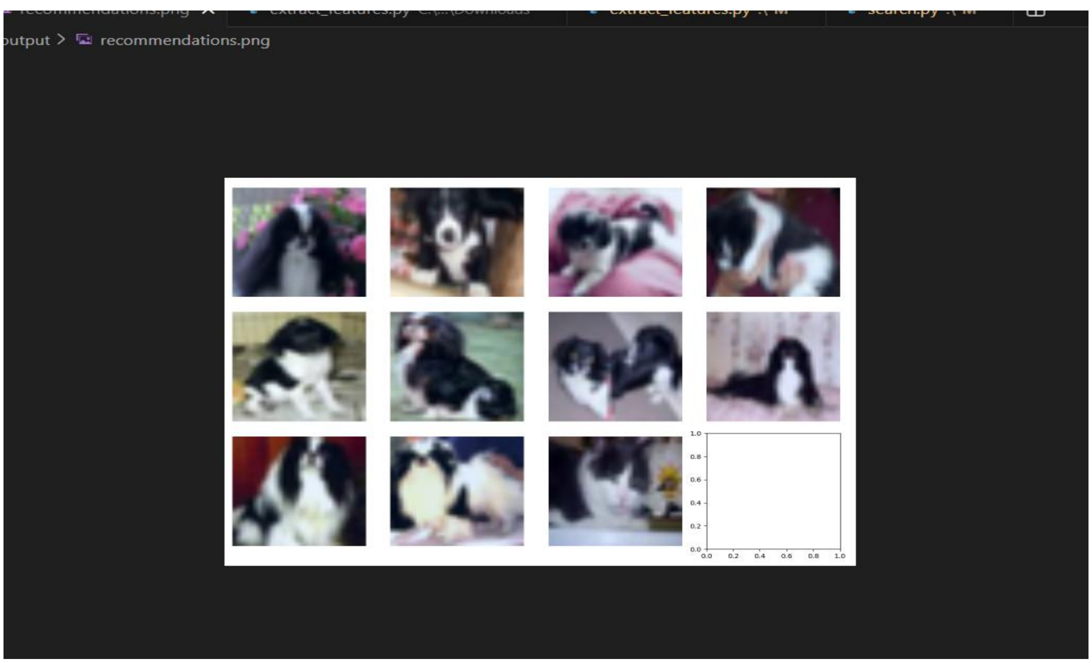
- **Evaluation Type**: Post-hoc evaluation using ground truth labels.
- **Metric Used**: Precision@K
    - Measures the fraction of retrieved images that belong to the same class as the query image.
- **Precision@K Formula**:
    - Precision@K = (Number of relevant images in top K) / K
- **Observations**:
    - The system achieved a **Precision@10 of 0.80**.
    - This indicates that, on average, 8 out of the top 10 retrieved images belong to the same semantic class as the query image.

    - High Precision@K values indicate strong semantic clustering in the embedding space.
    - The evaluation validates that pretrained embeddings capture class-level similarity without supervised fine-tuning.

Output

- The system retrieves visually and semantically similar images for a given query image.
- A grid of images is generated containing:
    - The query image.
    - The top-K recommended images.
- Results are saved as image files for visual inspection.
- The observed **Precision@10 = 0.80** demonstrates effective unsupervised image retrieval performance on the CIFAR-10 dataset.
- Precision@K score is printed to quantify retrieval performance.

```
Query Image Index: 42
Precision@10: 0.80
Saved image: output/recommendations.png
```



Future Scope

- Fine-tuning the backbone network using contrastive or self-supervised learning methods.
- Using larger and more diverse datasets for improved generalization.
- Incorporating approximate nearest neighbor search for scalability.
- Evaluating additional metrics such as Recall@K and Mean Average Precision (mAP).
- Extending the system to real-world applications such as product recommendation and medical image retrieval.

Output Video Link

[Google Drive Folder] (https://drive.google.com/drive/folders/1dxR8jXFB-ISh2IMuEw6bXqgLXAGgembC?usp=sharing)

Conclusion

- An unsupervised image-based recommendation system was successfully implemented.
- Pretrained ResNet-50 embeddings proved effective for semantic image retrieval.
- The approach eliminates the need for labeled training data.
- Cosine similarity with L2-normalized embeddings provides meaningful similarity ranking.
- The project demonstrates a practical and scalable solution for content-based image recommendation using deep learning.