B31MV Assignment 2 - 15%
Image Classification

Due Date: 16.03.2025

Objective

Develop a comprehensive understanding of the Bag of Visual Words (BoVW) model for image classification and explore the impact of data augmentation on model performance using the CIFAR-10 dataset.

Dataset

The CIFAR-10 dataset consists of 60,000 32x32 color images in 10 classes, with 6,000 images per class. The classes include airplanes, automobiles, birds, cats, deer, dogs, frogs, horses, ships, and trucks.

Tasks

1. Data Exploration and Preprocessing (10 marks)

1. Download the CIFAR-10 dataset and visualize a sample of images from each class to familiarize yourself with the data. Include the samples in the report.

2. Preprocess the images by converting them to grayscale to simplify the feature extraction process.

2. Feature Extraction (15 marks)

1. Utilize feature detectors such as SIFT (Scale-Invariant Feature Transform) or SURF (Speeded-Up Robust Features) to identify keypoints in the images.

2. Extract feature descriptors around the detected keypoints to capture the local image information.

3. Codebook Generation (20 marks)

1. Apply the k-means clustering algorithm to the extracted descriptors to form a codebook (visual vocabulary).

2. Experiment with different sizes of codebooks (e.g., 50, 100, 200 visual words) to observe the effect on classification performance.

3. Represent each image by a histogram that quantifies the occurrence of each visual word from the codebook.

5. Classification Model Development (30 marks)

1. Divide the dataset into training, validation and test subsets.

2. Utilize a machine learning classifier such as Support Vector Machine (SVM) to train on the histograms of visual words.

3. Implement a Convolutional Neural Network (CNN). You should use a pre-trained ResNet-18 model, fine-tuning the layers for the given dataset CIFAR-10 classification. Alternatively, you may also experiment with AlexNet or any other CNN but ResNet-18 is highly recommended for best performance.

4. Assess both models' performances using metrics such as accuracy, precision, recall, and F1-score. Discuss the results and compare the performance of the SVM with the CNN model.

6. Incorporating Data Augmentation (25 marks)

1. Apply data augmentation methods such as rotations, translations, scaling, and flipping to the training images to increase the diversity of the dataset.

2. Extract features, update the codebook, and retrain the classifier using the augmented dataset.

3. Compare the performance of the model trained with augmented data to the one trained with the original data to analyze the impact of data augmentation.

Deliverables

- A detailed report documenting:

  - The methodology and rationale for each step.
  - Screenshots from the results and code.
  - Comparative analysis of model performance before and after data augmentation.

- Separate zip file with the source code.