Vidyavardhini's College of Engineering & Technology

Department of Computer Engineering

To creating and training an Object Detector
Date of Performance: 04/09/2023
Date of Submission: 10/09/2023

Aim: To creating and training an Object Detector

Objective: Bag of Words(BOW) in Computer Vision and detecting cars in a scene.

Theory:

Bag of Words is a concept borrowed from natural language processing (NLP) that is used to represent and analyze visual information, particularly in the context of image classification and object recognition. The bag of words model is a simplified and effective way to represent images as histograms of visual words or features.

- Feature Extraction:

The first step is to extract local features from an image. Commonly used feature extraction techniques include Scale-Invariant Feature Transform (SIFT), Speeded-Up Robust Features (SURF), and Histogram of Oriented Gradients (HOG). These techniques identify distinctive points, edges, or texture patterns in the image.

- Vocabulary Building:

Once features are extracted from a set of images, a vocabulary of visual words is created. This vocabulary is essentially a dictionary of representative visual features. Clustering algorithms like k-means are often used to group similar features together, resulting in a predefined set of visual words.

- Feature Encoding:

In this step, each local feature in an image is assigned to the nearest visual word in the vocabulary. This process is called feature encoding or vector quantization. A common method for encoding is the "nearest neighbor" approach, where each feature is assigned to the visual word that is closest in feature space.

- Histogram Representation:

After feature encoding, the image is represented as a histogram of visual words. The histogram counts how many times each visual word appears in the image. This histogram represents the image's visual content in a bag of words fashion, where the order and spatial arrangement of the features are ignored.

- Classification or Retrieval:

The histograms generated for images can be used for various tasks. For instance, in image classification, the histograms are fed into machine learning classifiers (e.g., Support Vector Machines or Random Forests) to classify images into predefined categories. In image retrieval, the histograms are used to find similar images based on their visual content.

In computer vision, the bag-of-words model (BoW model), sometimes called bag-of-visual-words model can be applied to image classification or retrieval, by treating image features as words. In document classification, a bag of words is a sparse vector of occurrence counts of words; that is, a sparse histogram over the vocabulary. In computer vision, a bag of visual words is a vector of occurrence counts of a vocabulary of local image features.

Code :-

```
import cv2

import sklearn.feature_extraction.image

import numpy as np
```

```
# Load the image

image = cv2.imread('/content/BOW.jpg')

# Convert the image to grayscale

grayscale_image = cv2.cvtColor(image, cv2.COLOR_BGR2GRAY)
```

```
# Extract the SIFT features

sift = cv2.SIFT_create()

#sift = cv2.SIFT()

keypoints, descriptors = sift.detectAndCompute(grayscale_image, None)

from sklearn.cluster import KMeans

# Cluster the descriptors using k-means

n_clusters = 100

kmeans = sklearn.cluster.KMeans(n_clusters=n_clusters)

labels = kmeans.fit_predict(descriptors)

# Create a bag of visual words

bag_of_words = np.zeros(n_clusters)

for label in labels:

    bag_of_words[label] += 1

# Print the bag of words

print(bag_of_words)
```

Final Output:-

```
[ 301. 1498. 456. 277. 595. 326. 318. 643. 288. 391. 327. 301. 535. 348.
626. 501. 518. 456. 380. 308. 392. 353. 544. 474. 359. 971. 356. 586. 397.
315. 275. 312. 920. 268. 435. 290. 362. 380. 396. 519. 739. 477. 363. 463.
520. 1025. 370. 446. 356. 909. 161. 232. 348. 457. 364. 914. 349. 376.
447. 213. 348. 337. 359. 470. 392. 304. 299. 589. 339. 374. 305. 315. 254.
412. 374. 397. 452. 312. 310. 319. 478. 502. 421. 523. 372. 399. 222. 357.
385. 307. 412. 283. 341. 371. 442. 387. 336. 1924. 320. 334.]
```

Conclusion: In conclusion, our study on the bag of words (BOW) model simplifies images into histograms of visual words, making it computationally efficient and robust to minor variations in scale, rotation, and viewpoint. However, it does lose some spatial and structural information, which more advanced techniques like Convolutional Neural Networks (CNNs) aim to capture.