Expt_9

Pooja Naskar- 30

Aim: To Creating and Training an Object Detector

Objective: Bag of Words BOW in computer version Detecting cars in a scene

Theory:

Creating and Training an object detector:

The process of building and training an object detector from start can be difficult and resource-consuming. Usually, PyTorch or TensorFlow, two deep learning frameworks, are used. It usually takes a lot of computational resources, particularly strong GPUs, and a solid grasp of deep learning principles to train object detectors from scratch. Using pre-trained models on big datasets and refining them on smaller, domain-specific datasets is beneficial for many object recognition applications.

Following are the steps for creating and training an object detector:

1. Data Collection and Annotation: Gather and annotate a dataset of photos containing the things you wish to identify.

2. Data Preprocessing: Adjust the image sizes and standards. To make the dataset more diverse, add more information to it.

3. Select a Deep Learning Framework: Choose between a specialized object identification library like Detectron2 (based on PyTorch) or TensorFlow Object identification API, or a deep learning framework like TensorFlow or PyTorch.

4. Model Architecture: Select an object detection architecture, like SSD (Single Shot MultiBox Detector), YOLO (You Only Look Once), or Faster R-CNN.

5. Loss Function: Choose a loss function that makes sense for detecting objects.

6. Training: Use your annotated dataset to train the model.

7. Assessment: Utilize an independent validation dataset to test the trained model.

Bag - Of - Words:

In computer vision and natural language processing, the Bag of Words (BoW) method is frequently utilized for text and picture analysis. It's an easy-to-use and efficient method of expressing data for a variety of applications, including image classification, text classification, and document retrieval. BoW is used to create fixed-length feature vectors from variable-length sequences (like words or pictures)..

BOW in Computer Vision:

By interpreting image attributes as words, the bag-of-words model (BoW model), also known as the bag-of-visual-words model, can be used in computer vision for image classification or retrieval. A bag of words, or sparse histogram over the vocabulary, is a sparse vector of word occurrence counts in document categorization. A bag of visual words in computer vision is a vector of local image feature occurrence counts.

Here's how BoW is typically applied in computer vision:

- Feature Extraction
- Clustering
- Feature Quantization
- Histogram Generation
- Normalization
- Image Classification
- Image Retrieval

Detecting Cars Example:

In computer vision, the Bag of Words (BoW) model is used to recognize cars. To get started, gather a tagged dataset that includes both car and non-car images. The next step is to extract local features from these pictures, including additional distinguishing patterns or the Histogram of Oriented Gradients (HOG). Car identification is based on these qualities.

Using a method similar to K-Means clustering, the BoW methodology clusters these local features into visual words or codewords. Each cluster center represents a visual word in the generated clusters, which create a visual lexicon or codebook.Once you have the

visuallanguage, you can effectively quantize the local attributes from each image into discrete visual words by assigning them to the closest visual word in the codebook. To convert the picture data into a format that can be used for machine learning, this step is essential.

You create a histogram for every image, counting the instances of every visual word. photos of vehicles can be distinguished from non-car photos by their distinctive histograms, which match the distribution of visual terms connected to cars

These histograms can be used as feature vectors to train an automobile identification model by classifying each training sample as either "car" or "non-car." For this, machine learning classifiers like Random Forests and Support Vector Machines (SVMs) are frequently used.

In actuality, you extract local features, make histograms, and input them into the trained classifier in order to identify automobiles in fresh photos. After that, the classifier produces predictions that enable the pictures of cars to be identified. Non-maximum suppression is one post-processing technique that can assist improve the detected automobile zones and remove overlapping or duplicate detections.

Conclusion:

A practical computer vision tool for object recognition and car detection is the Bag of Words Model. The BoW model can be applied to the detection of cars in order to determine their location and presence inside an image or video frame. Since it uses machine learning, language construction, and visual features to identify cars in images or scenarios, it is a versatile solution for a range of applications, including traffic surveillance and autonomous vehicles.

We were able to successfully investigate the Bag of Words Model method for automobile detection in this experiment.