Pedestrian detection using Convolutional Neural Networks

Problem Statement

Pedestrian detection is problem of detecting the location of individuals who are walking on a particular indoor and outdoor environment. It is an essential and significant task in any intelligent video surveillance system, as it provides the fundamental information for semantic understanding of the video footages. These applications need real-time detection performance for timely decision making, by using limited computing power and resources available in an embedded device. In this research, a novel Convolutional Neural Network based object detection method will be exploited for pedestrian detection in the Indian context. Nowadays, Deep Learning based solutions are applied to the problem of pedestrian detection. One of the challenges in applying Convolutional Neural Network based pedestrian detection is, applying such models for the Indian context where regular and traditional outfits of the people highly vary from the outfits of the people in the regular datasets. One common challenge for any CNN based pedestrian detection is to meet the real time processing requirements where the Deep Learning model should run on embedded devices with limited processing power and energy.

Background

Despite the challenges, pedestrian detection still remains an active research area in computer vision in recent years. Numerous approaches have been proposed over these years. It has an obvious extension to automotive applications due to the potential for improving safety systems. Many car manufacturers (e.g. Volvo, Ford, GM, and Nissan) offer this as an ADAS option in 2017.

Traditional pedestrian detectors are trained to search for pedestrians in the video frame by scanning the whole frame. The detector would “fire” if the image features inside the local search window meet certain criteria. The drawback of this approach is that the performance can be easily affected by background clutter and occlusions. Sometimes, pedestrians are modeled as collections of parts. Part hypotheses are firstly generated by learning local features, which include edgelet and orientation features. Though this approach is attractive, part detection itself is a difficult task.

Recently patch based pedestrian detection approach combines both the detection and segmentation with the name Implicit Shape Model (ISM). The advantage of this approach is only a small number of training images are required. When the conditions permit (fixed camera, stationary lighting conditions, etc.), background subtraction can help to detect pedestrians. Background subtraction classifies the pixels of video streams as either background, where no motion is detected, or foreground, where motion is detected. This algorithm has been extended to the detection of humans in 3D video streams.

Methodology

Architecture of the pedestrian detection is shown in figure 1**.**

Step 1: Data collection and dataset preparation

This will involve collection of images from various sources and formatting them, annotating them with ground truth object bounding boxes.

Step 2: Developing a CNN based pedestrian detection model

Popular pretrained CNN feature extraction models such VGG16, ResNet and object detection frameworks such as SSD (shown in Figure 2), YOLO will be exploited for this task.

Step 3: Training and experimentation on datasets

The pedestrian detection model will be trained both on the large scale datasets such as Pascal VOC 2007, KITTI Vision and dataset that will be populated based on Indian context as part of this project.

Step 4: Deployment and analysis on real life scenario

The trained and tested pedestrian detection model will be deployed in a real-life scenario for further analysis where both positive and failure cases will be leveraged for further improvement in the methodology.

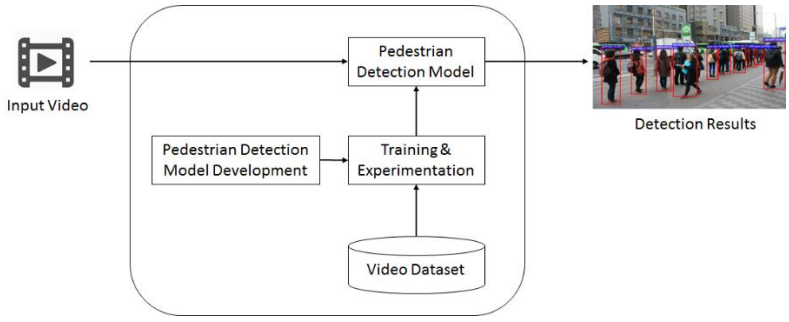

Figure 1. Architecture of Pedestrian Detection

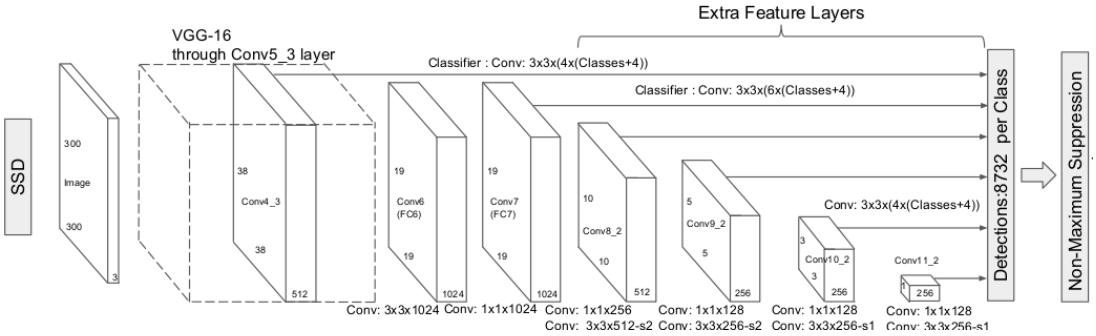

Figure 2. Architecture of CNN Based Object detection [Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y. and Berg, A.C., 2016, October. Ssd: Single shot multibox detector. In European conference on computer vision (pp. 21-37). Springer, Cham.]

Experimental Design

Dataset

Pedestrian detection datasets such as Pascal VOC 2007, KITTI Vision will used for experimentation and evaluation.

Evaluation Measures

Measures such as accuracy and Mean Average Precision (MAP) will be computed by comparing the detected pedestrian bounding boxes and ground truth boxes from the datasets.

Software and Hardware Requirements

Python based Computer Vision and Deep Learning libraries will be exploited for the development and experimentation of the project. Tools such as Anaconda Python, and libraries such as OpenCV, Tensorflow, and Keras will be utilized for this process. Training will be conducted on NVIDIA GPUs for training the end-to-end version of CNN based object detection model.