literature review

**High-level Semantic Feature Detection: A New Perspective for Pedestrian Detection**

This paper discusses a methodology for pedestrian detection based on high level semantic detection. Object detection often relies on a backbone of residual networks of deep learning models which take an image of the desired data as an input, reverting it into a map of different resolution using the formula φi = fi(φi−1) = fi(fi−1(...f2(f1(I)))) φi representing the maps output. Using a pretrained standard network on ImageNet (e.g. ResNet-50 and MobileNet) multi-scale feature maps can be created with each per stage/layers of convolutional neural network (CNN) and combined into one that contains more sematic information. An experiment was conducted using the Cross Stage Partial Network (CSP) focused on evaluating the performance of its two main components: the feature extraction module and the detection head. The featured extraction module, based on ResNet-50, was responsible for processing the input images and extracting feature maps with multiple resolutions. These feature maps were able to capture information from different scales, however a downside of this method (CSP) occurs in the simple design of the detection head. The detection head made of a single 3x3 layer followed by the prediction layers offers efficiency at the cost of limited ability to capture intricate object details and context in complicated scenarios. The experiment utilized the Caltech and CityPersons datasets for pedestrian detection. Caltech comprises approximately 2.5 hours of driving video footage with extensively labelled bounding boxes, while CityPersons is a challenging large-scale pedestrian detection dataset. The feature extraction module processed the data, and 4024 test frames were extracted for evaluation. The training process employed a ResNet-50 backbone pretrained on the ImageNet dataset, a common practice in deep learning. Optimization was performed using the Adam algorithm, commonly used for training neural networks and Mini-batch training was used for the divided dataset for a more efficient result.