ELG 5163 – Machine Vision
Techno-blog

Name: Xu Zhang (0300062651), University of Ottawa
General topic: 13. Shape Recognition and Classification
Selected sub-topic: Pedestrian Detection

Abstract

In this micro-lecture, I will present a pedestrian detection method based on deep neural network model, to automatically learn a pedestrian's inherent features in order to identify that pedestrian in the complex background.

Background

Pedestrian detection is an essential and significant task in intelligent video surveillance system and It has an obvious extension to automotive applications due to the potential for improving safety systems. However, pedestrian detection is a challenging task due to its complexity background as well as various body sizes and postures.

Traditional pedestrian detection methods are built on handcrafted features and shallow trainable architectures which focus on capturing low-level feature extraction of pedestrians, but these methods has to find a good trade-off between the accuracy and efficiency of computation.

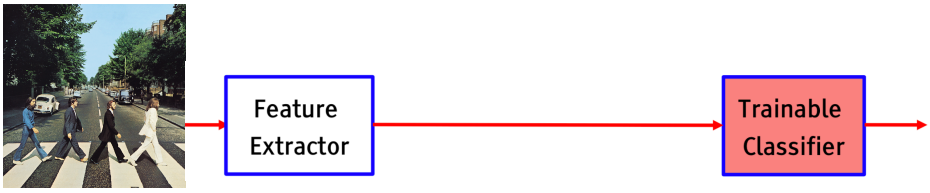

Figure 1 Traditional Pedestrian Detection Methods

Recently, with the development of deep learning technology, instead of extracting the low-level feature as traditional methods, deep learning method can discover multiple levels of representation of the images through multiple layers and can be considered as feature extractor.

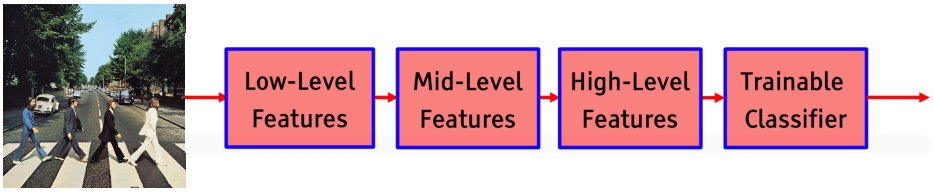

Figure 2 Deep Learning Pedestrian Detection Methods

Methodology

There are three related algorithms which are all based on deep learning: RCNN, Fast R-CNN and Faster R-CNN. The difference is that Faster RCNN can implements the end-to-end training in the convolution neural network which means it implements all the processes in deep nets: feature extraction, region proposals and region classification as well as bounding box regression, other two still need to implement some of the processes by hand-crafted.

Because pedestrian detection is a specific part of the object detection, so I select Faster R-CNN algorithm as my method in this blog. The framework of the methodology is shown below:

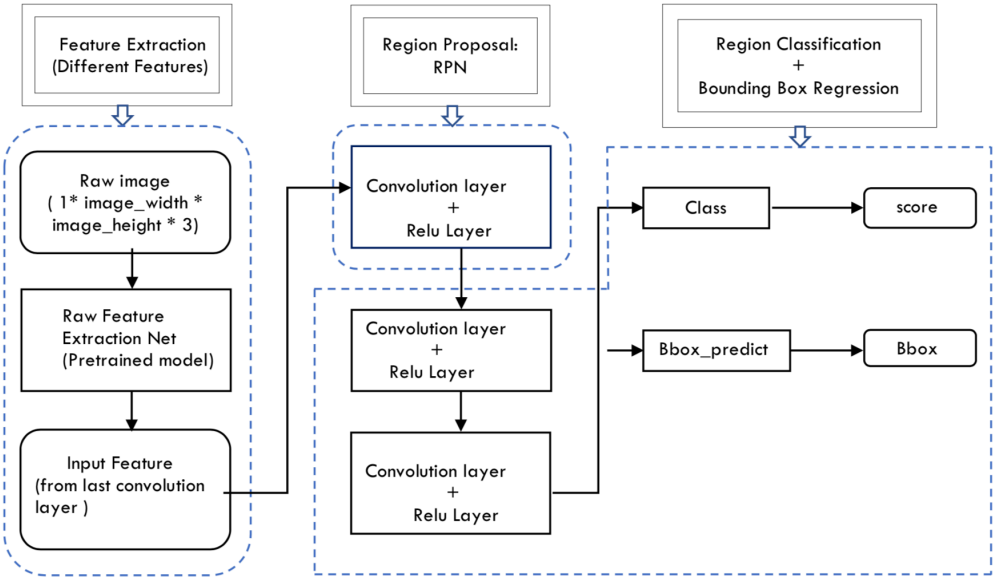

Figure 3 Framework of Methodology

In this framework, Feature Extraction Network(FEN) using a pretrained model such as VGG16 which can extract the raw feature of the input images from the last convolution layer of the model. Region Proposal Network(RPN) can get the region proposals over the feature map.

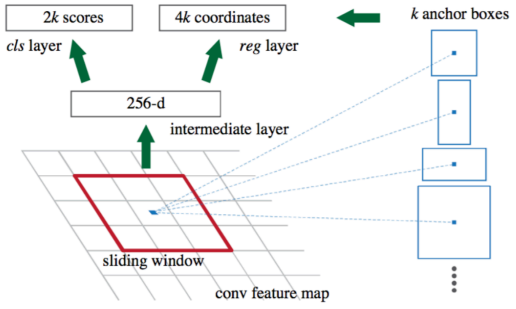

Figure 4 RPN in Faster R-CNN

Region Classification part will predict the probability of each anchor box including a pedestrian. And bounding box regression part can predict four offset coordinates for each anchor box. Through the backpropagation, the performance will be optimized via calculating the gradient of the loss function, minimizing the loss and adjusting the parameters, like the weights and bias.

Implementation

1. Load the dataset

The dataset I used in this blog is *Caltech Pedestrian Dataset*, which consists of approximately 250,000 frames with a total of 350,000 bounding boxes and 2300 unique pedestrians were annotated. After load the dataset, we need to split the data set into a training set(60%), cross validation set(20%) and test set(20%).

2. Set up optimization algorithm

In this blog, I choose *Stochastic gradient descent with momentum*(SGDM) as my optimization algorithm, which is a stochastic approximation of gradient descent optimization. We can use *trainingOptions* function in Matlab to set the parameters of SGDM as:

MaxEpochs = 5, MiniBatchSize = 1, InitialLearnRate = 1e-3

3. Train the network

The Faster R-CNN pedestrian detection network is composed of a feature extraction network followed by two sub-networks. The feature extraction network is typically a pretrained CNN such as VGG16. The first sub-network following the feature extraction network is a region proposal network (RPN) trained to generate object proposals. The second sub-network is trained to predict the actual class of each proposal. We can use *trainFasterRCNNObjectDetector* function in Matlab to automatically add the sub-networks for this model.

4. Evaluate Detector Using Test Set

Evaluate the model on test set to measure the trained model's performance. Matlab *Computer Vision Toolbox™* provides evaluation function *evaluateDetectionPrecision* to get average precision matrix of this model:

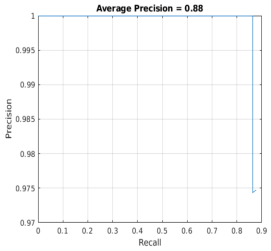

Figure 5 Average Precision Matrix of this model

5. Apply trained model on new image

Figure 6 Apply trained model on new image

References

- Chen, L. (2017). Pedestrian Detection Based on Deep Learning. UC Riverside. ProQuest ID: Chen_ucr_0032N_13061. Merritt ID: ark:/13030/m5sf7s4z. Retrieved from https://escholarship.org/uc/item/9q04r9hg

- Zhong-Qiu Z., Peng Z., Shou-tao X., Xindong W. (2015). Object Detection with Deep Learning: A Review. arXiv:1807.05511.

- Girshick, R. "Fast R-CNN." Proceedings of the 2015 IEEE International Conference on Computer Vision. Santiago, Chile, Dec. 2015, pp. 1440-1448.

- Zitnick, C. L., and P. Dollar. "Edge Boxes: Locating Object Proposals from Edges." European Conference on Computer Vision. Zurich, Switzerland, Sept. 2014, pp. 391-405.

- Uijlings, J. R. R., K. E. A. van de Sande, T. Gevers, and A. W. M. Smeulders. "Selective Search for Object Recognition."_ International Journal of Computer Vision_. Vol. 104, Number 2, Sept. 2013, pp. 154-171.

- Girshick, R. "Fast R-CNN." Proceedings of the 2015 IEEE International Conference on Computer Vision. Santiago, Chile, Dec. 2015, pp. 1440-1448.

- Ren, S., K. He, R. Gershick, and J. Sun. "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks." IEEE Transactions of Pattern Analysis and Machine Intelligence. Vol. 39, Issue 6, June 2017, pp. 1137-1149.

- Girshick, Ross, et al. "Rich feature hierarchies for accurate object detection and semantic segmentation." Proceedings of the IEEE conference on computer vision and pattern recognition. 2014.

- He, Kaiming, et al. "Spatial pyramid pooling in deep convolutional networks for visual recognition." IEEE transactions on pattern analysis and machine intelligence 37.9 (2015): 1904-1916.

- Zhang, Liliang, et al. "Is faster r-cnn doing well for pedestrian detection?." European Conference on Computer Vision. Springer International Publishing, 2016.