INT408 – ASSESSMENT 2

Assessment Number	1
Contribution to Overall Marks	40%
Submission Deadline	20-Dec -2020

Assessment Objective

This assessment aims at evaluating students' ability to exploit the deep learning knowledge, which is accumulated during lectures, and after-class study, to analyze, design, implement, develop, test and document the pedestrian detection algorithm using Mask R-CNN framework [1]. The assessment will be based on the Pytorch software.

General Guidelines

1. The descriptions in the Problem Specifications are required to be analyzed with mathematic equations, combined with the explanations of all elements in each equation.
2. The modified part of the source codes are required to include in the report.
3. The final pedestrian detection performance that you obtain should be reported in the lab report. Meanwhile, the screenshot of the final performance results are also required in the report.
4. For the final performance results that you obtained, the numeric quantitative results are required. In addition, is also important to include some subjective image examples in the report with correct bounding boxes marked in green, whereas false bounding boxes marked in red.

Pedestrian detection using Mask R-CNN

Overall Description:

This lab is to use the Pytorch software and Mask R-CNN framework [1] for pedestrian detection. Pedestrian detection [2] aims to predict bounding boxes of all the pedestrian instances in an image. It has attracted much attention within the computer vision community in recent years as an important component for many human-centric applications, such as self-driving vehicles, person re-identification, video surveillance and robotics.

Mask R-CNN is a framework with both object detection and segmentation using deep convolutional neural network. The following images are examples of pedestrian detection.

Figure 1. Examples of pedestrian detection.

Problem Specifications:

1. Please describe the 2 key components in the Mask R-CNN framework: the RoIPooling layer and the loss functions in the framework. (20%)

2. Please describe the object detection performance metric, mAP (Mean Average Precision), and explain why it can well reflect the object detection accuracy. (20%)

3. Please train (or fine-tune) and test the framework on one of the existing pedestrian detection datasets, and report the final AP performance that you have achieved. The dataset in this lab is PennFudanPed [3]. Please also report some pedestrian detection examples by including the images and bounding boxes. (40%).

4. Propose your own method to further improve the pedestrian detection performance or reduce the model size based on the Mask R-CNN framework, and compare different methods with the performance you obtained and explain why. (20%)

Hint:

1. In literature, there are existing methods that improve the Mask R-CNN to reduce the model size which fit the deep learning model to smart devices. One such work is published in [3]. You can refer to [3] for problem 4.
2. This lab is based on the official tutorial of Mask R-CNN from pytorch, and it is really a good try to read the document and play the code in the Colab.

https://pytorch.org/tutorials/intermediate/torchvision_tutorial.html

https://colab.research.google.com/github/pytorch/vision/blob/temp-tutorial/tutorials/torchvision_finetuning_instance_segmentation.ipynb

[1] He K, Gkioxari G, Dollár P, et al. Mask RCNN. arXiv e-prints, Article[J]. arXiv preprint arXiv:1703.06870, 2017.

[2] Tong Xiao, Shuang Li, Bochao Wang, Liang Lin, and Xiaogang Wang. Joint detection and identification feature learning for person search. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jul 2017.

[3] https://www.cis.upenn.edu/~jshi/ped_html/

[4] Sandler M, Howard A, Zhu M, et al. Mobilenetv2: Inverted residuals and linear bottlenecks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 4510-4520.