

行人重識別在真實世界之應用

Evaluation of Person Reidentification (Re-ID) techniques with Real-World data

指導教授：藍崑展

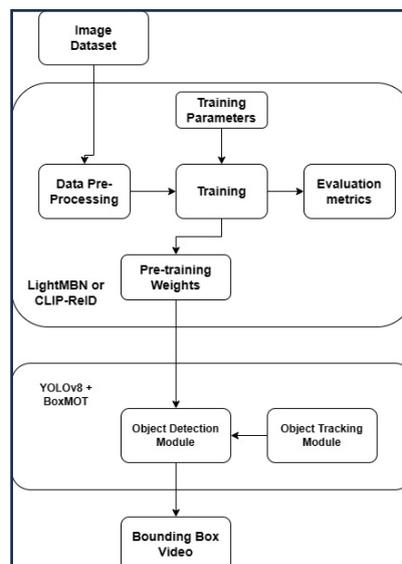
專題成員：張奕翔

開發工具：Vim, Conda, PyTorch, YOLOv8

測試環境：Linux, Windows 10

一、簡介：

Person Reidentification (Re-ID) is an increasingly crucial task in computer vision which is used to enhance security and surveillance systems by accurately tracking multiple individuals across different camera perspectives. This is done through feature extraction of the entire appearance of an individual, including clothing, and pose, to reidentify them in different locations or times. While many Re-ID techniques have been proposed with State-of-the-art (SOTA) performance, however, many of those performances are based on controlled environment benchmark tests and lacked real-world applicability. This project aims to address this gap by evaluating a few SOTA Re-ID techniques to observe their effectiveness and practical limitations by feeding them real-world input data. The models will be trained using both benchmark and custom datasets, and evaluated on precision and training time. The resultant Re-ID weights will be leveraged by an object detection, and tracking model on multiple custom real-world scenario video data to produce a bounding box video with unique identifiers for each individual. By analysing the Re-ID models' performance on real-world data, this project aims to identify which technique is more suitable for practical security applications requiring accurate person tracking.



二、 測試結果

Using standard benchmark tests such as Market1501, LightMBN outperforms CLIP-ReID as it produces good average and rank-1 results while also being fairly efficient in training time. Similarly, when training YOLOv8 on the resultant Market1501 weights on real-world video data such as Figure 2 below (Market environment), LightMBN is able to produce promising results, showing most visible individuals to the camera while needing less computation time than CLIP-ReID. However, when training on smaller sized custom datasets, CLIP-ReID outperforms LightMBN on average precision metrics due to the leveraging of CLIP's pretrained features.

When trained on real-world data with higher levels of occlusion and noise, both CLIP-ReID and LightMBN struggle with performance, and the resultant bounding box video struggles to identify the individuals. This is due to the limited amount of feature information that can be extracted from noisy, occluded data, and as a result both models are unable to make accurate image classifications.



Figure 2: Sample bounding box video on street camera footage trained on CLIP-ReID