Technical Assessment for AI Scientist (Computer Vision)

Question 1: Video-based Person Re-Identification
Candidate: Drayang Chua Kai Yang

1.0 Reason choosing Torchreid over other approach such as Multi-Camera Object Tracking via Deep Metric Learning and Person_reID_baseline_pytorch

Library

Easy to implement TorchReid because it is a python library. Besides, proper documentation had been provided for the user to understand and implement for further usage.

Consistent updating

The developer team still maintain the library. There are more than 1k forks for this library, where the huge community is able to continue further improve the library.

Lightweight deep learning model with high performance

The developer team designed a novel deep re-ID CNN, termed **omni-scale network (OSNet),** for omni-scale feature learning. The deep model had been proved to achive better performance than the standard CNN model including ResNet-50, MobileNet, shuffletNet and etc in person re-identification. It is noteworthy that OSNet has only 2.2 million parameters, which are far less than the current best-performing ResNet-based methods. Furthermore, OSNet is the deep model that achieve the best result for "market1501" and “DukeMTMC-reID” image-reid datasets. The pretrained model also show high performance for “DukeMTMC-VIdeoReID” video-reid dataset.

Moreover, since the video-based Person Reidentification algorithm is expected to employ for real-time application, a lightweight model can definitely reduce the time cost and processing power while tracking the person in a video stream. This help to make the pipeline faster and achieve inference rate that required by a real-time application.

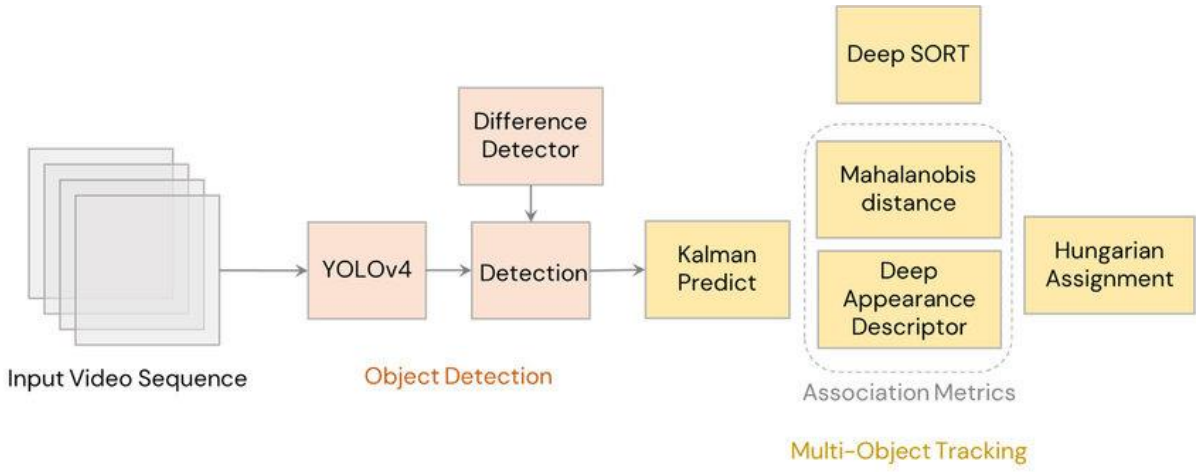
2.0 Comparison with traditional tracker such as FairMOT and DeepSort

General

DeepSort and FairMOT both are very popular methodology used in Multi-Object Tracking (MOT). Both approaches consists two main concept: object detection and re-identification (tracking).

DeepSORT

- **Anchor-based approach**
- DeepSORT is the improved version of Simple Online Realtime Tracking (SORT). SORT has 4 key components: Detection, Estimation, Data association and Identity-tracking
- Problem of SORT solved by DeepSORT:
    - identity switch
    - fails in case of occlusion
    - motion blur (vibrating camera)



- DeepSORT is implemented along with YoLo deep model, where Yolo focus on object detection, and Deep Sort focus on mult-object tracking.

Reference:

1) https://learnopencv.com/understanding-multiple-object-tracking-using-deepsort/#:~:text=DeepSORT%20can%20be%20defined%20as,offline%20just%20before%20implementing%20tracking.

FairMOT

- **Anchor-free approach, one shot MOT**
- FairMOT managed to solve the problem that encounter by the common architecture thatemploys a two-step methodology : Object Detection being done with a CNN-based detector (YOLO, Faster R-CNN) and the detected objects being passed on to another component to perform the Re-identification (Re-ID) task.
- E.g
    - bad detection can lead to difficult re-ID (“cascading effect”)
    - To re-ID, the detector’s high level ‘deep’ features are used (i.e. features are share for both detection and tracking model). Research suggests low level feature has better results for re-ID.
    - Feature of re-ID is large. Low dimension re-ID features result in higher accuracy.
- Architecture of FairMOT

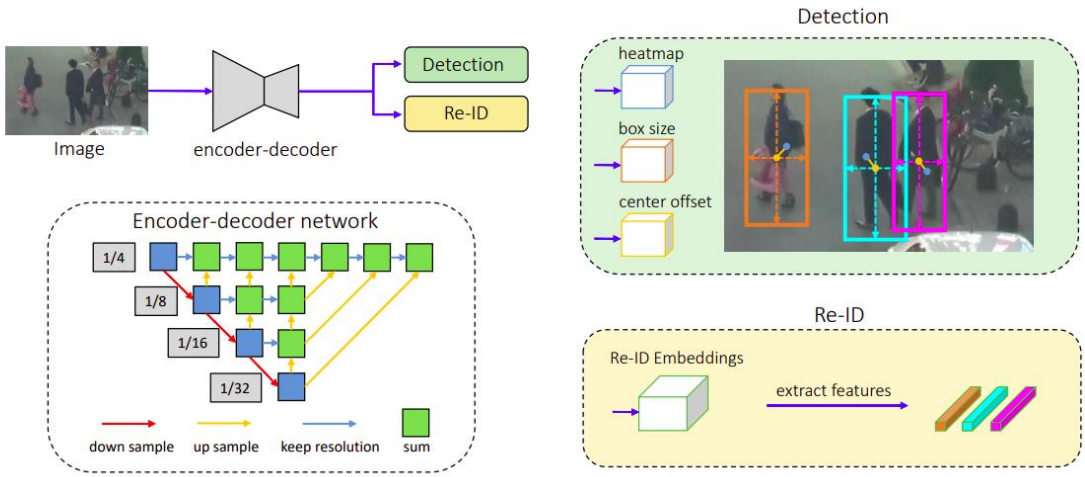


Fig. 2. Overview of our one-shot tracker *FairMOT*. The input image is first fed to an encoder-decoder network to extract high resolution feature maps (stride=4). Then we add two homogeneous branches for detecting objects and extracting re-ID features, respectively. The features at the predicted object centers are used for tracking.

    - Network
        - encoder-decoder
    - Detection branch
        - output: Heat map (estimate object center), box size, box offset
    - Re-ID branch
        - output: feature of re-ID through classification
    - Data Association
        - Somehow combine re-ID, Kalman filter and IoU

Reference:

1) https://learnopencv.com/object-tracking-and-reidentification-with-fairmot/
2) https://wandb.ai/vbagal/Multi-Object%20Tracking/reports/Yolov5_DeepSort-vs-FairMOT--Vmlldzo4Nzk0MjQ

Comparison

Task

Torchreid library (also, the other suggested approach) itself only do the work for reidentification, which mean it does not carry out the task for object detection. The input for Torchreid deep model is the bounded image that already detected and extract by the previous detection model. Hence, to achieve a real time approach

Re-identification model

FairMOT and DeepSort use standard CNNs model such as ResNET, which were originally designed for object category-level recognition tasks. However, OSNet deep model learn omni-scale feature representation, which is more important in re-identification. Torchreid developer argue that the discriminative feature used for re-ID must be of omni-scale, which is the combination of variable homogenous scale and heterogenous scale. This can make the reidentification task become more effective when we come to a more complex scenario that homogenous scales are not enough to make different. Example of homogenous feature is shirt, whereas heterogenous feature is the logo or design of the shirt.

3.0 Methodology

As mentioned earlier, the re-identification model that I used is the **Torchreid** library, specifically the **OSNet deep model**. Since OSNet itself cannot do the object detection, I implemented a similar pipeline approach as DeepSORT, by employing a YoLo model for object detection, then pass the detected result to OSNet model for person reidentification.

The datasets that I used are “market1501” image-reid dataset and “dukemtmc” video-reid dataset. Torchreid provides various commands to ease us in training the deep model. Besides, there are various pre-trained models provided by Torchreid developer team, hence, we can implement the desire model based on our need. The pretrained model weight can be found here.

Torchreid provide a comprehensive guide on the implementation of the library in person re-ID. First, we have to prepare the dataset and download into the “/reid-data” directory. The source of dataset is listed below:

1. market1501 image dataset
2. dukemtmc image dataset
3. dukemtmc video dataset

After that, we can build the deep model and define the hyperparameter based on our needs, then start training or evaluating the deep model.

The written code can refer to the “torchreid_Lauretta.ipynb” jupyter notebook file with this link.

For real time application and object detection purpose, YoloV3 deep network is used. The YOLO network had been trained on COCO dataset to determine person in the video stream, image, or real time stream. The source code can be found here.

The idea is similar to DeepSORT, where the detected person on the video stream can be marked with an unique ID and continuously tracked by the re-identification deep model (i.e. OSNet model) in our case. The image below shows the real time detection of a person.

The written code can refer to the “real_time_reid.py” python file with this link.

4.0 Result and discussion

Unfortunately, due to time constraint and technical issue, the implementation of both Yolo and Torchreid OSNet deep model for a real time person reidentification is incomplete. Hence, I mainly focus on the performance of Torchreid OSNet model.

The performance of a person re-identification model can be evaluated using mean average precision (mAP) and Cumulative matching Characteristics (CMC) curve.

Dataset	mAP	Rank-1 CMC
Market1501 (image)	82.6%	94.3%
Dukemtmc (video)	91.1%	92.6%

For both market1501 and dukemtmc datasets, OSNet possess high mAP and Rank-1 CMC. The result indicate that the model can track most of the person accurately.

Below shows a few visualizations of rank result on market1501 dataset.

5.0 Other Reference

1. What is reidentification
2. CMC evaluation metrics
3. mAP (mean Average Precision) for Object Detection
4. Real time human detection with OpenCV
5. FairMOT: Multi Object Tracking
6. Omni-Scale Feature Learning for Person Re-Identification paper
7. Torchreid documentation