

進度報告 (人員跨鏡追蹤)

林晉德

2021-7-22

outline



- 1 跨境追蹤
 - 問題定義
 - 特徵提取與比對
 - 檢測的準則(指標)
 - 傳統預訓練的分法
 - 盲測式(世超論文方法)
 - 訓練集與測試集
- 2 實驗目標：電資學院三樓跨鏡追蹤
- 3 世超架構與問題及目前採用架構比較
- 4 工作細項及規劃

行人跨鏡追蹤 (Person Re-ID)

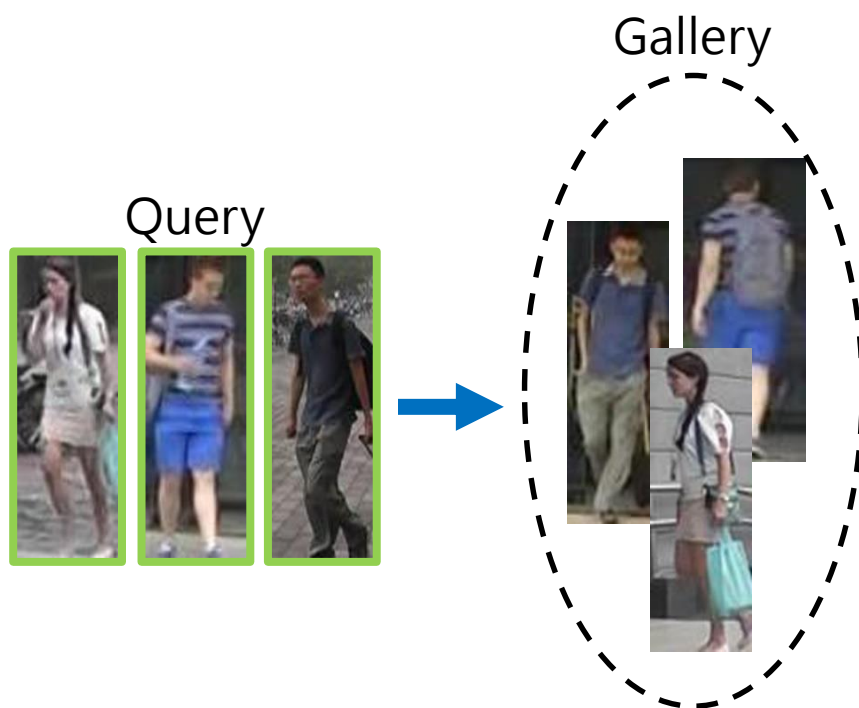
- 行人跨鏡追蹤(Person Re-Identification, Person Re-ID)目的為從包含其他攝影機畫面之行人圖像的比對圖庫 (Gallery) 中尋找與查詢圖像 (Query) 匹配的行人。



封閉式 & 開放式 Person Re-ID

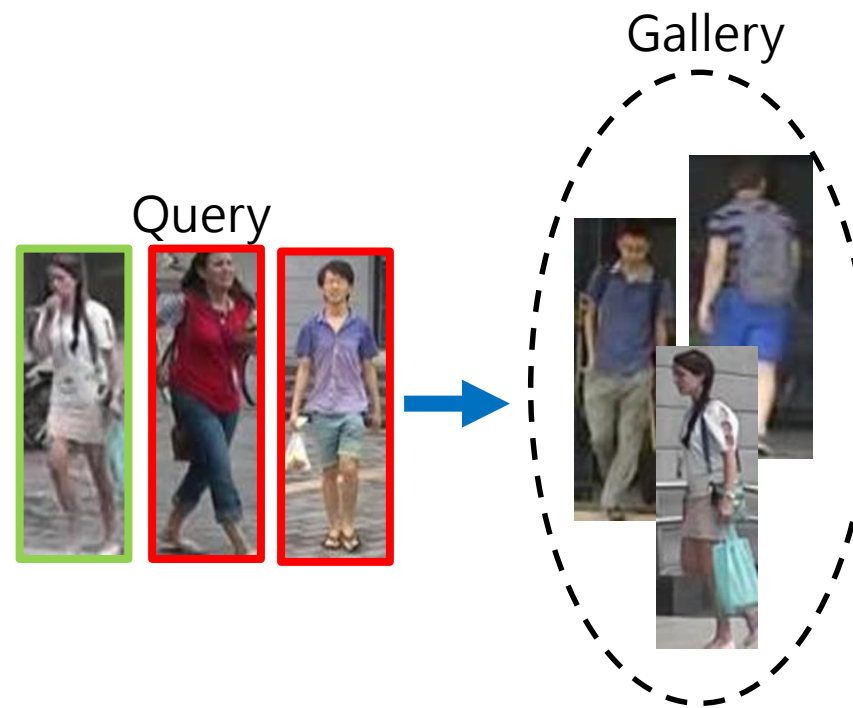
封閉式 (Close-world)

- 比對圖庫 (Gallery) **一定包含** 所有查詢圖庫 (Query) 的行人。

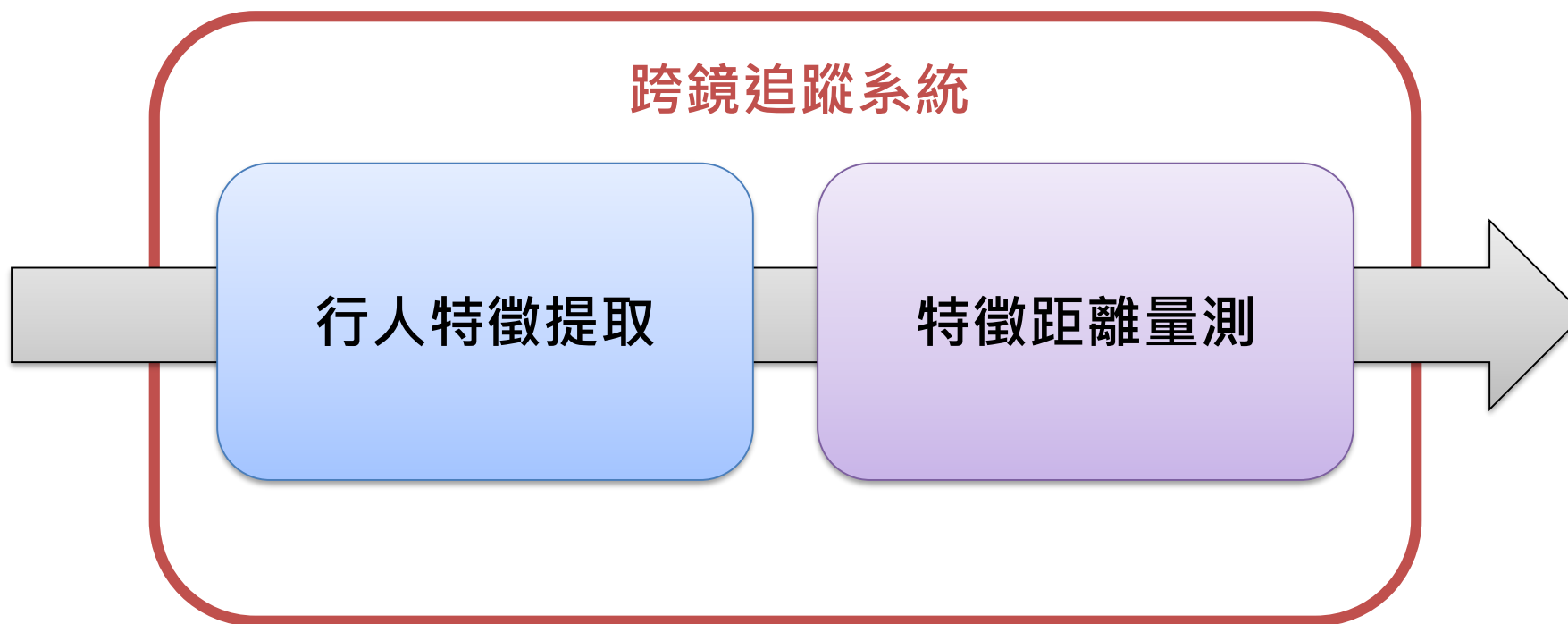


開放式 (Open-world)

- 比對圖庫 (Gallery) **不一定包含** 所有查詢圖庫 (Query) 的行人。



Person Re-ID 作法摘要

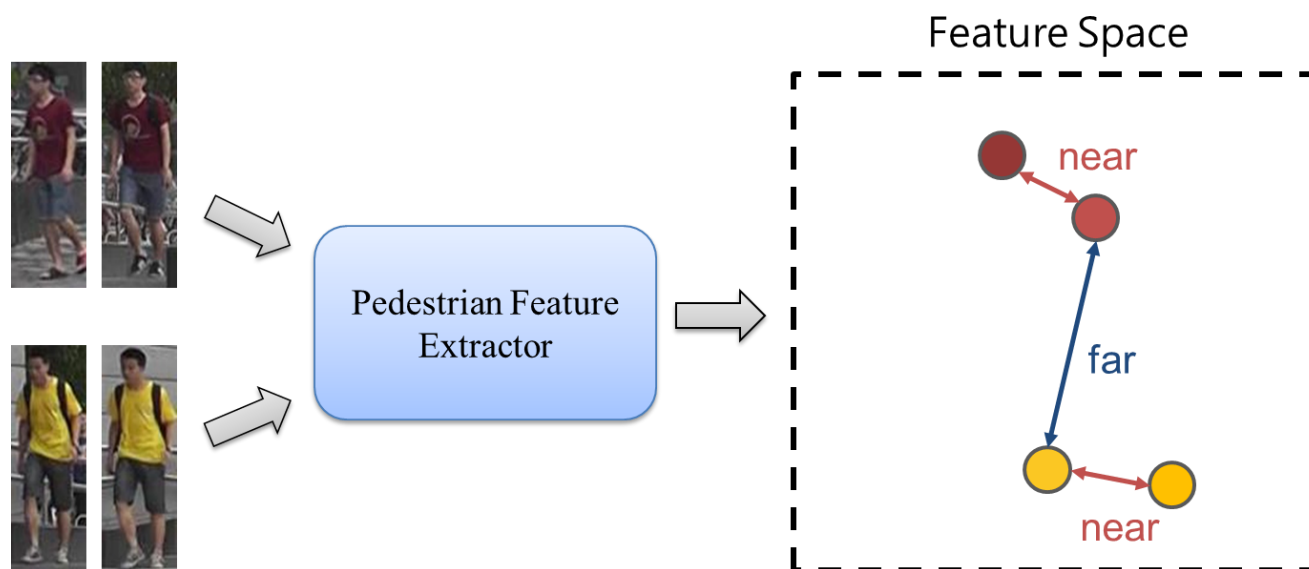


Y. Sun, L. Zheng, Y. Yang, Q. Tian, and S. Wang, "Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline)," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 480-496.

Person Re-ID 作法摘要

1. 行人特徵提取

- 學習出一個特徵提取網路，使提取出的特徵具有相同人距離越近，不同人距離越遠的特性。



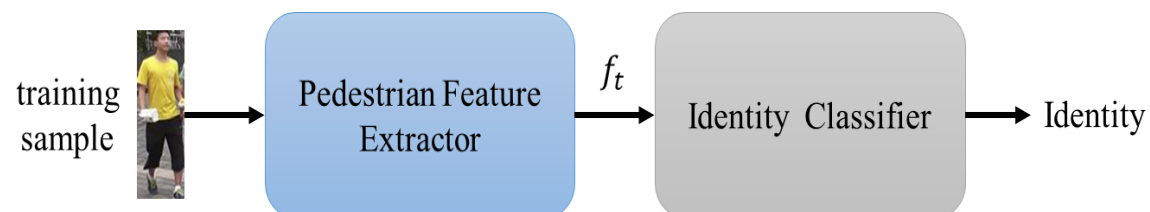
2. 特徵距離量測

- 量化特徵表示之間的距離大小，通常使用歐式距離或是餘弦距離

特徵擷取器學習

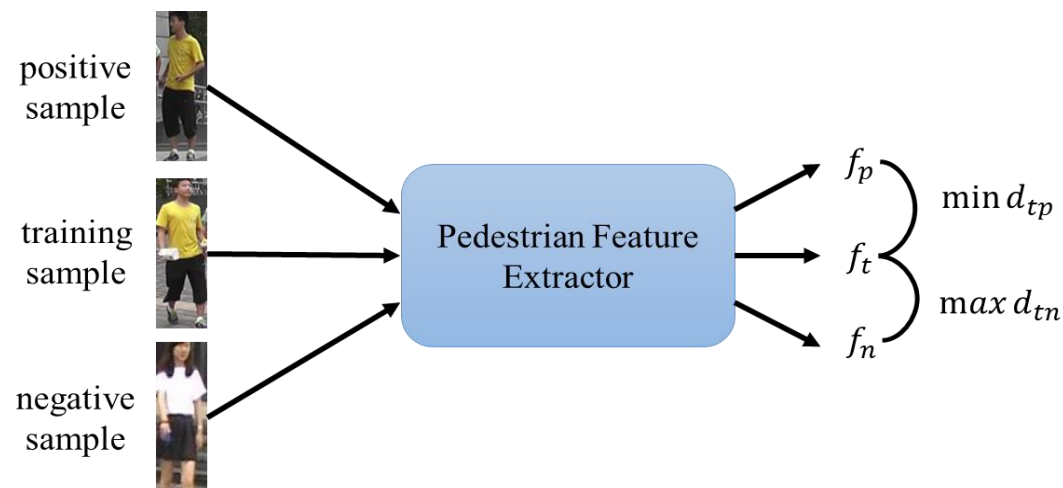
1. 身分學習 (Identity Learning)

- 串接一個行人身分的分類網路

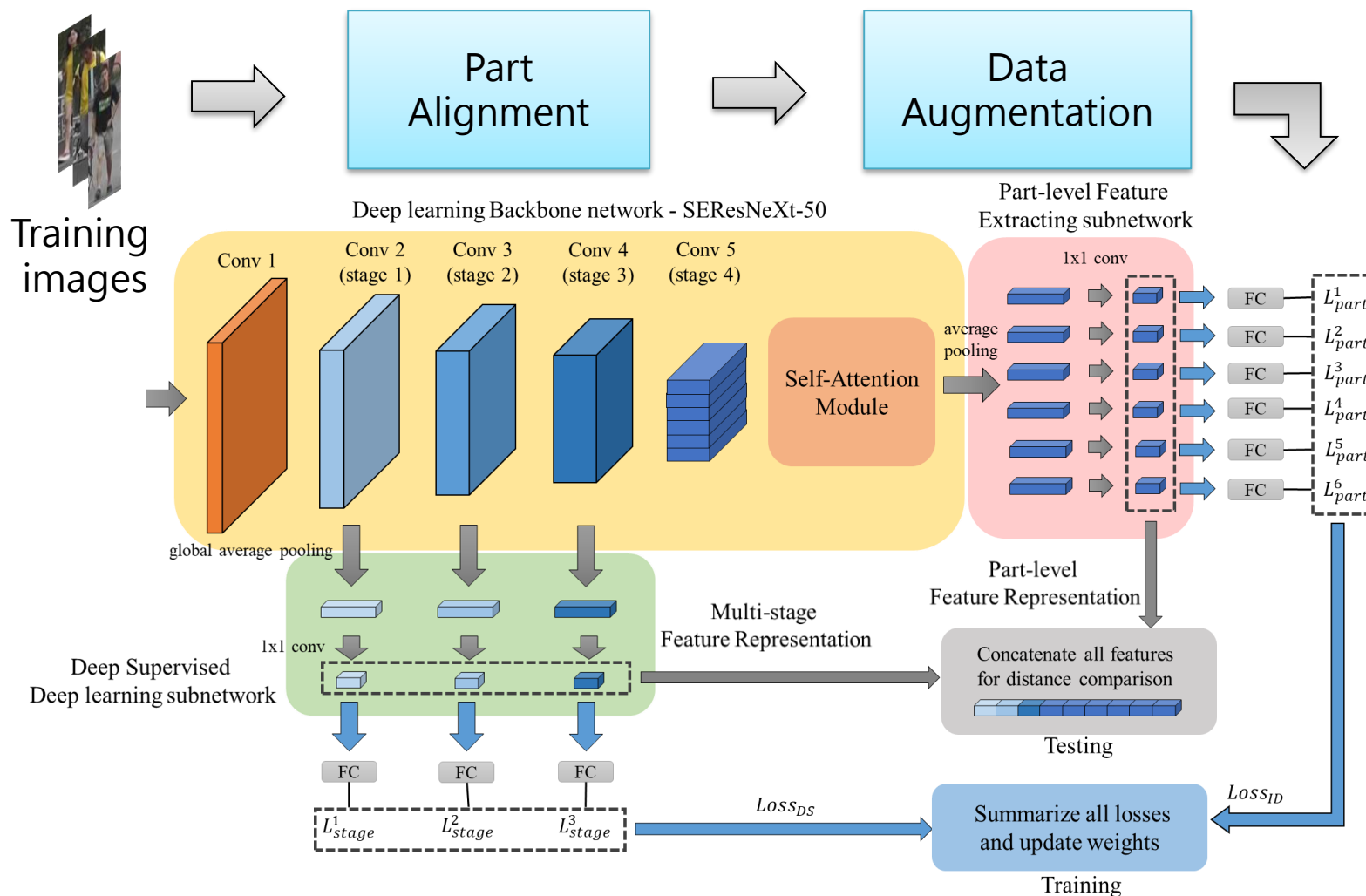


2. 度量學習 (Metric Learning)

- 學習最小化相同人的特徵間距離，最大化不同人的特徵間距離



網路訓練流程



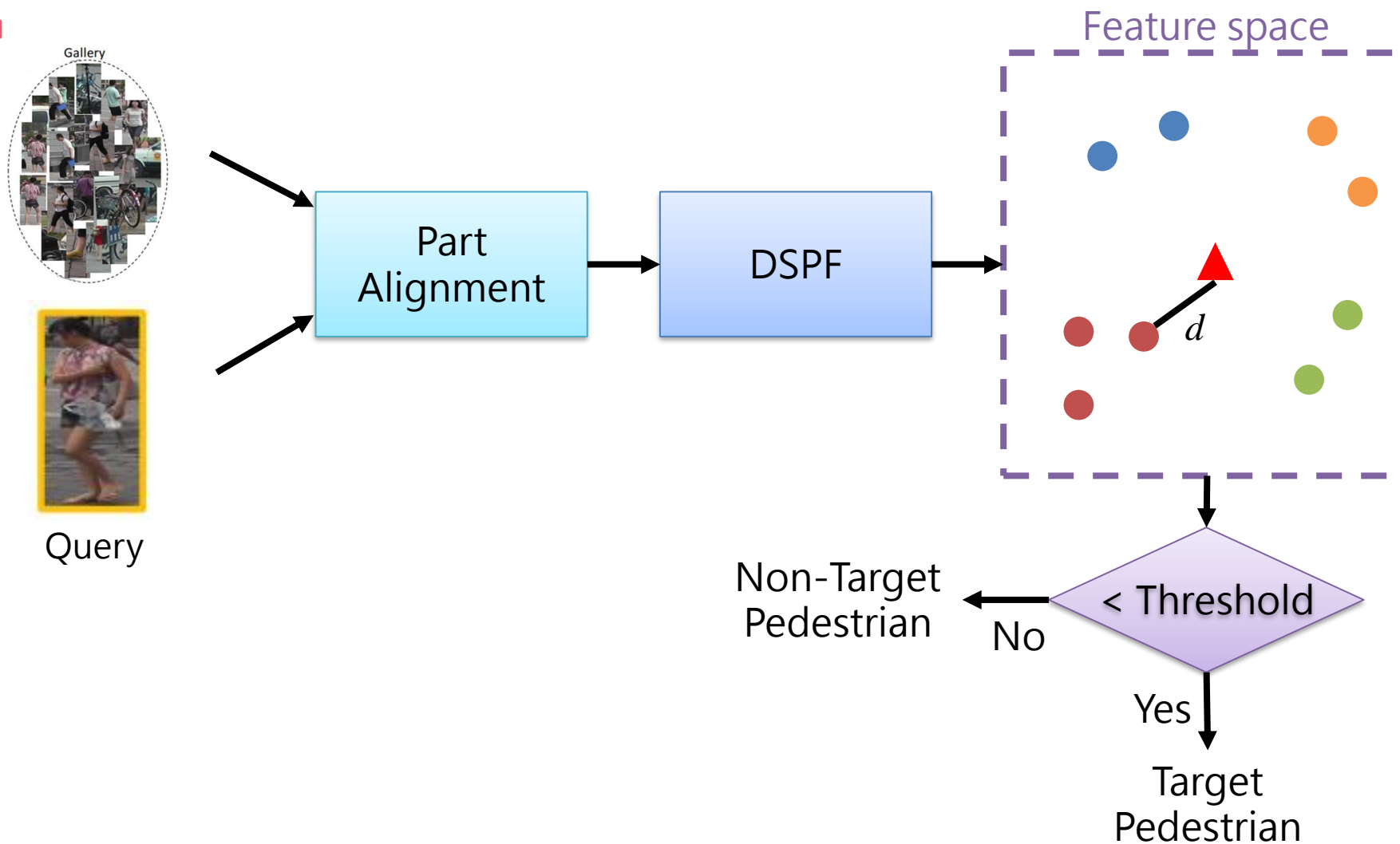
網路測試流程

A. 資料集與效能評估指標

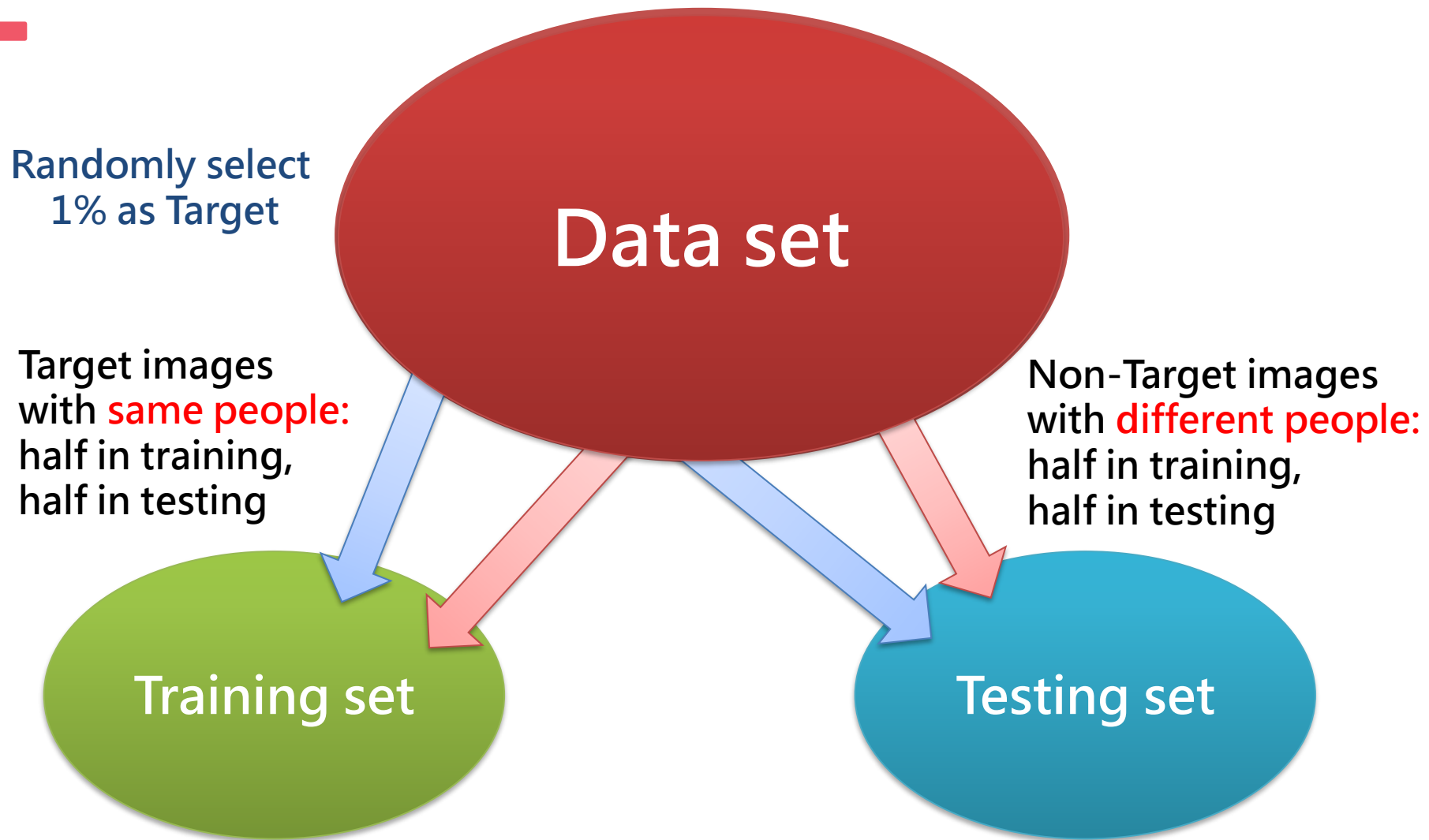
B. 訓練與測試流程

C. 效能展示與比較

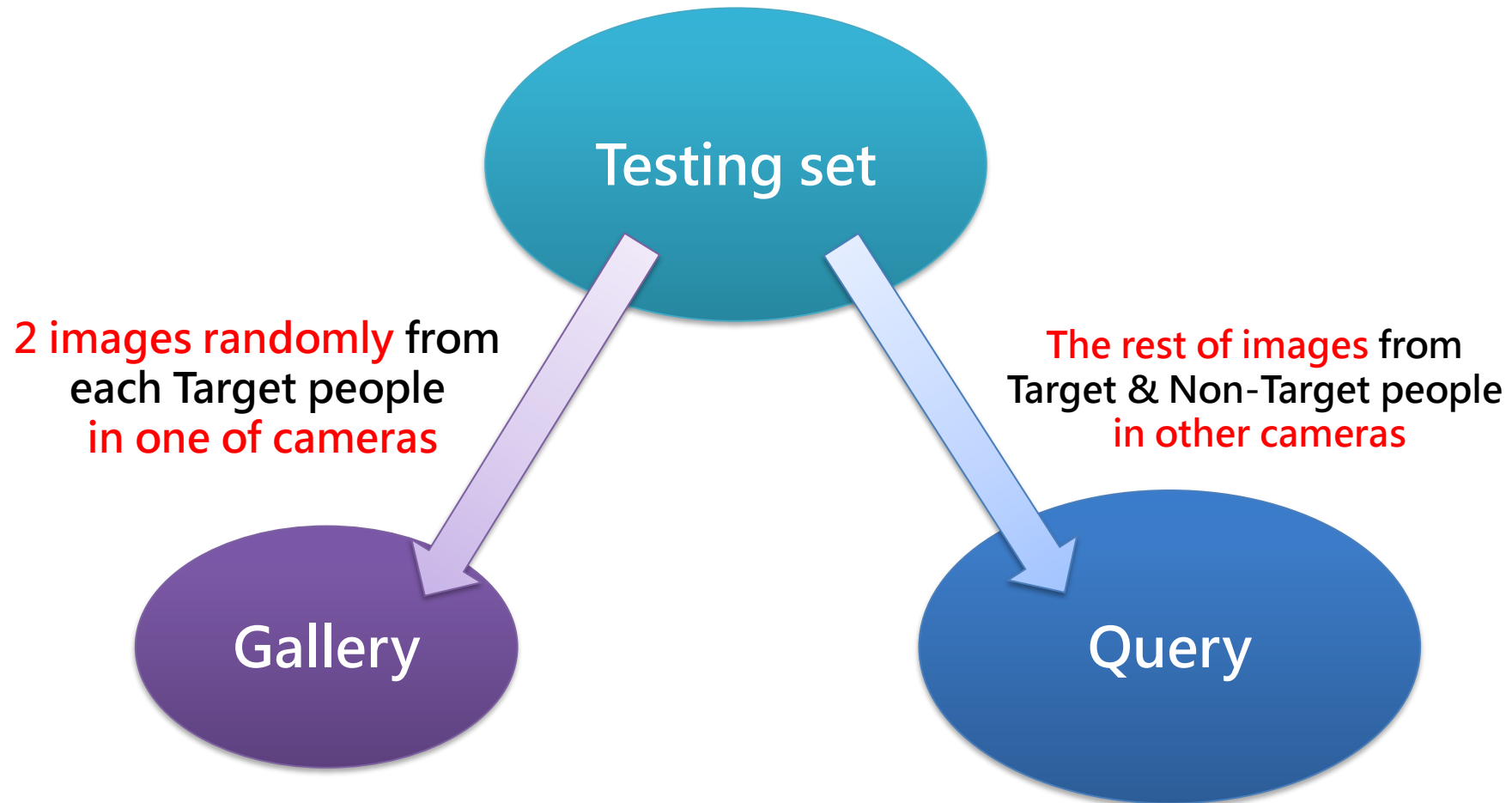
D. 電資學院3樓實際測試環境



一般協定的資料集分配流程(1/2)



一般協定的資料集分配流程(2/2)



All steps repeat for 10 times

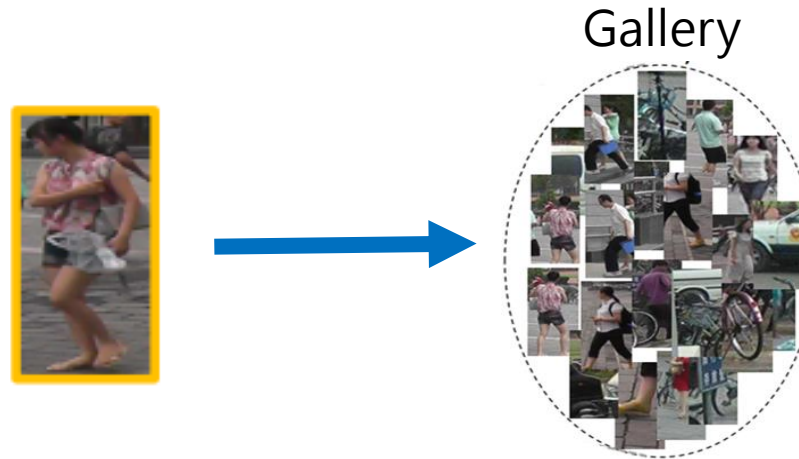
評測指標: TTR & FTR

- A. True Target Rate (TTR)
- B. False Target Rate (FTR)

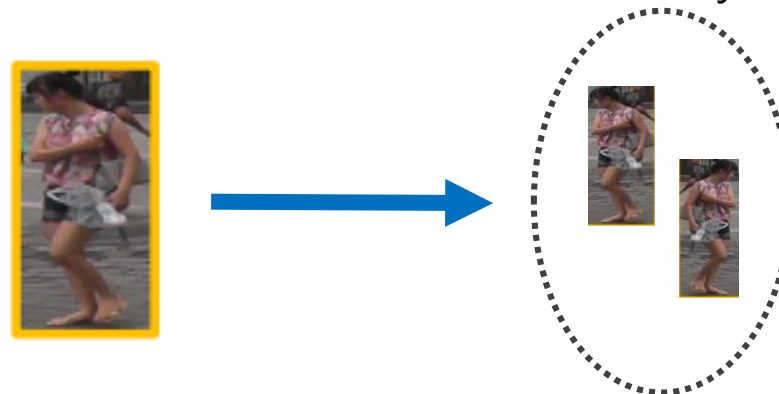
$$TTR = \frac{\text{Number of Target images verified as one of Target people}}{\text{Total of Target images}}$$
$$FTR = \frac{\text{Number of Non-Target images verified as one of Target people}}{\text{Total of Non-Target images}}$$

評測項目: 集合驗證(SV) & 個別驗證(IV)

- Set Verification (SV)



- Individual Verification (IV)



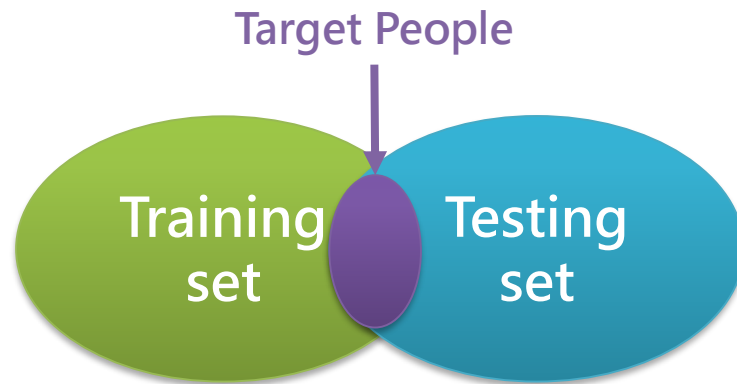
B. 訓練與測試流程

C. 效能展示與比較

D. 電資學院3樓實際測試環境

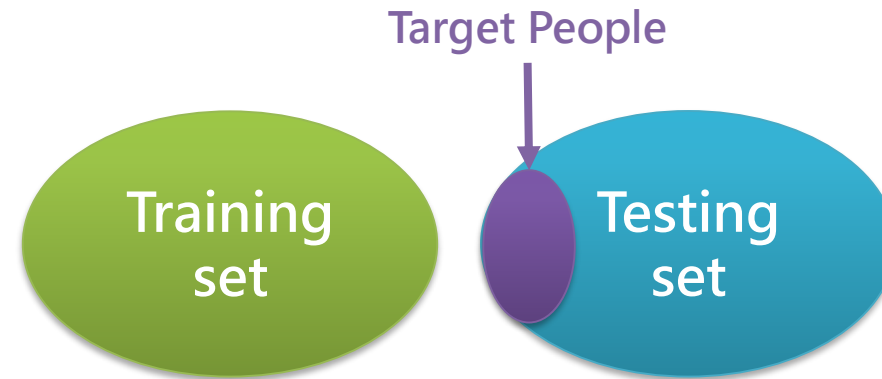
公正的資料集分配方法—盲測式

- 兩種資料分配主要的差異：
預訓練過(文獻協定)



目標行人身分重疊

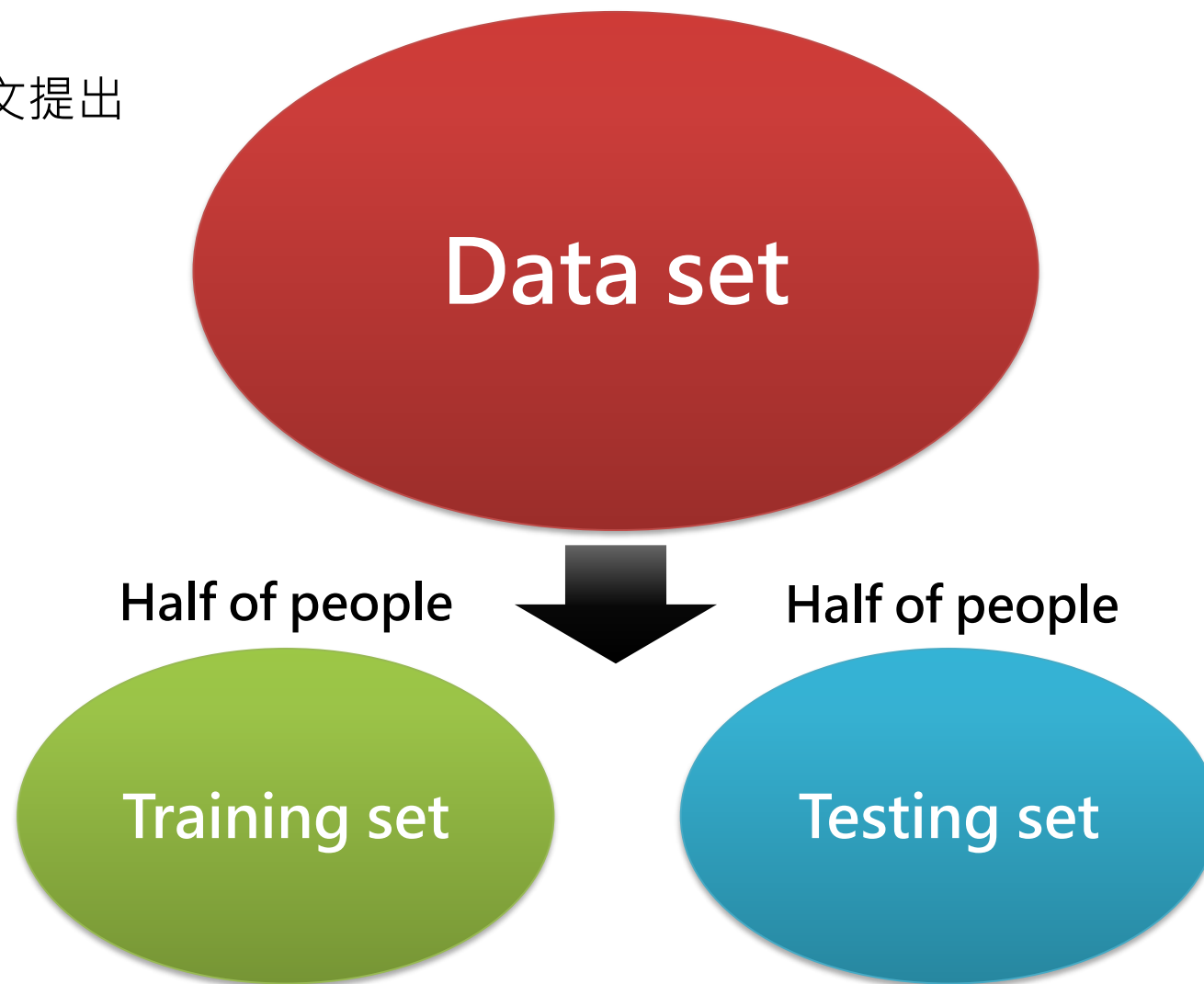
盲測式 (本論文提出)



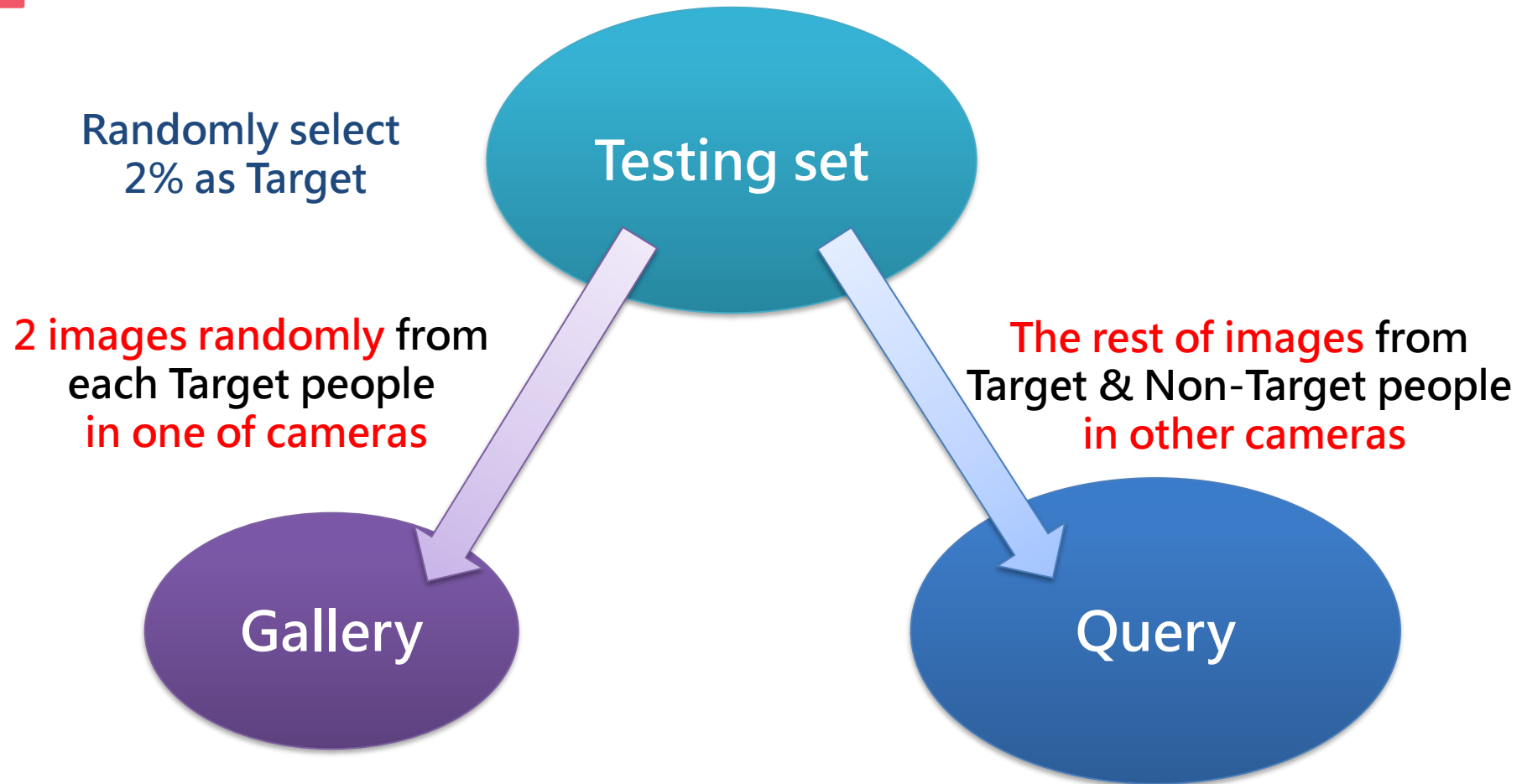
行人身分不重疊

盲測式資料分配流程(1/2)

- 由本論文提出



盲測式資料分配流程(2/2)



All steps repeat for 10 times

資料集參數

B. 訓練與測試流程

C. 效能展示與比較

D. 電資學院3樓實際測試環境

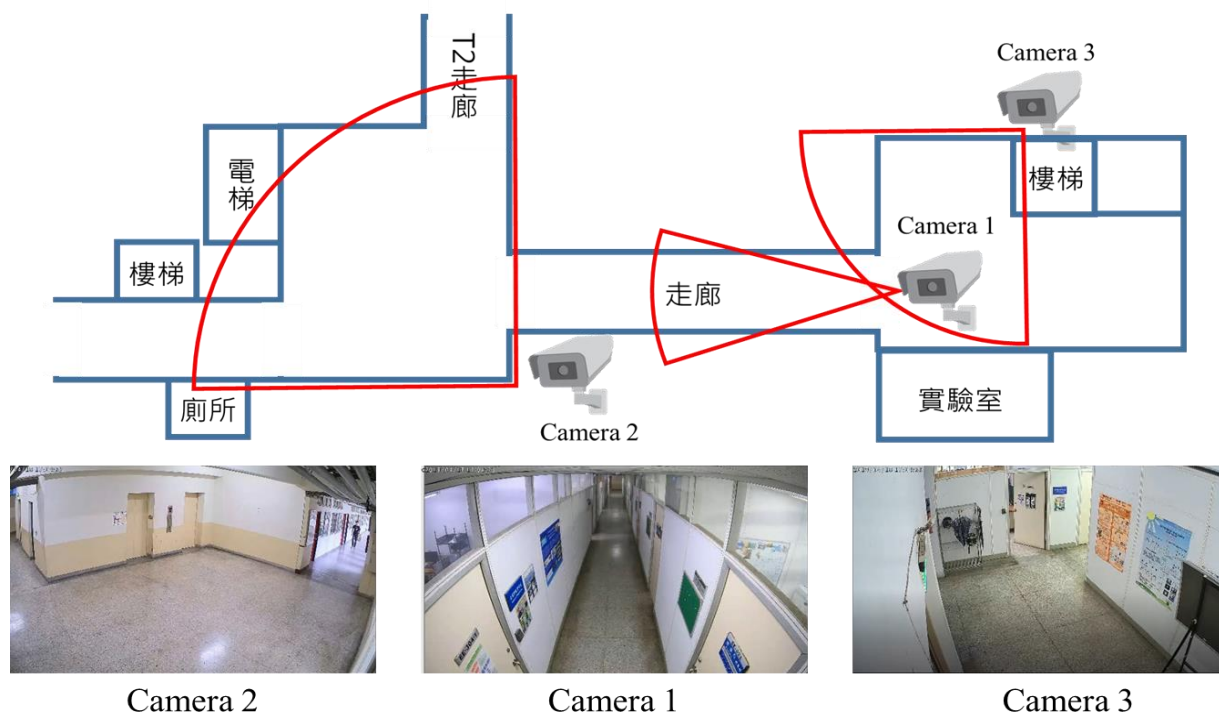
資料集 (釋出年分)	Market-1501 (2015)	DukeMTMC-ReID (2017)
攝影機數量	6	8
行人類別數	1,501	1,812 (1404 + 408)
圖像總數	32,668	36,411
行人偵測器	DPM	人工
特色	穿著多為短袖、短褲 框選不準的情形較多	穿著多為外套、長褲 存在干擾作用的行人

盲測式 & 預訓練過資料分配效能比較

Dataset	Market-1501						DukeMTMC-ReID					
FTR (%)	0.10	1	5	10	20	30	0.10	1	5	10	20	30
Evaluation	Set Verification (SV)											
盲測式	40.96	72.75	90.00	94.75	97.44	98.05	38.54	67.29	80.80	85.86	89.33	90.92
預訓練過	47.65	86.25	96.55	98.56	99.48	99.83	43.75	77.60	90.48	94.39	97.26	98.21
Evaluation	Individual Verification (IV)											
盲測式	79.22	96.61	99.06	99.5	99.70	99.81	74.68	90.34	95.38	96.45	98.13	99.06
預訓練過	87.74	97.54	99.33	99.63	99.94	99.94	81.83	95.80	98.25	98.75	99.43	99.89

論文貢獻3： 架設實際應用場景並建立EE3F資料集

- 架設跨3台攝影機的小型室內應用場景，收集行人圖像並進行標註，建立資料集為後續研究奠定基礎。



電資學院3樓 - 小型監測網

訓練集與測試集分配

- 使用預訓練過分配方式

Dataset	Training set	Testing set	
		Gallery	Query
Camera	1, 2, 3	3	1, 2, 3
Number of Identity	395	31	395
Total images	12,806	62	12,420
Target People	31	31	31
Non-Target People	364	0	364

EE3F 資料集的測試效果展示

● 預訓練過分配

Dataset	EE3F					
FTR (%)	0.10	1	5	10	20	30
Evaluation	Set Verification (SV)					
DSPF	60.90	81.70	92.77	96.78	98.39	99.59
Evaluation	Individual Verification (IV)					
DSPF	89.25	95.39	98.77	99.84	100	100

發現待測試的問題 EE3F Data Set 盲測沒做

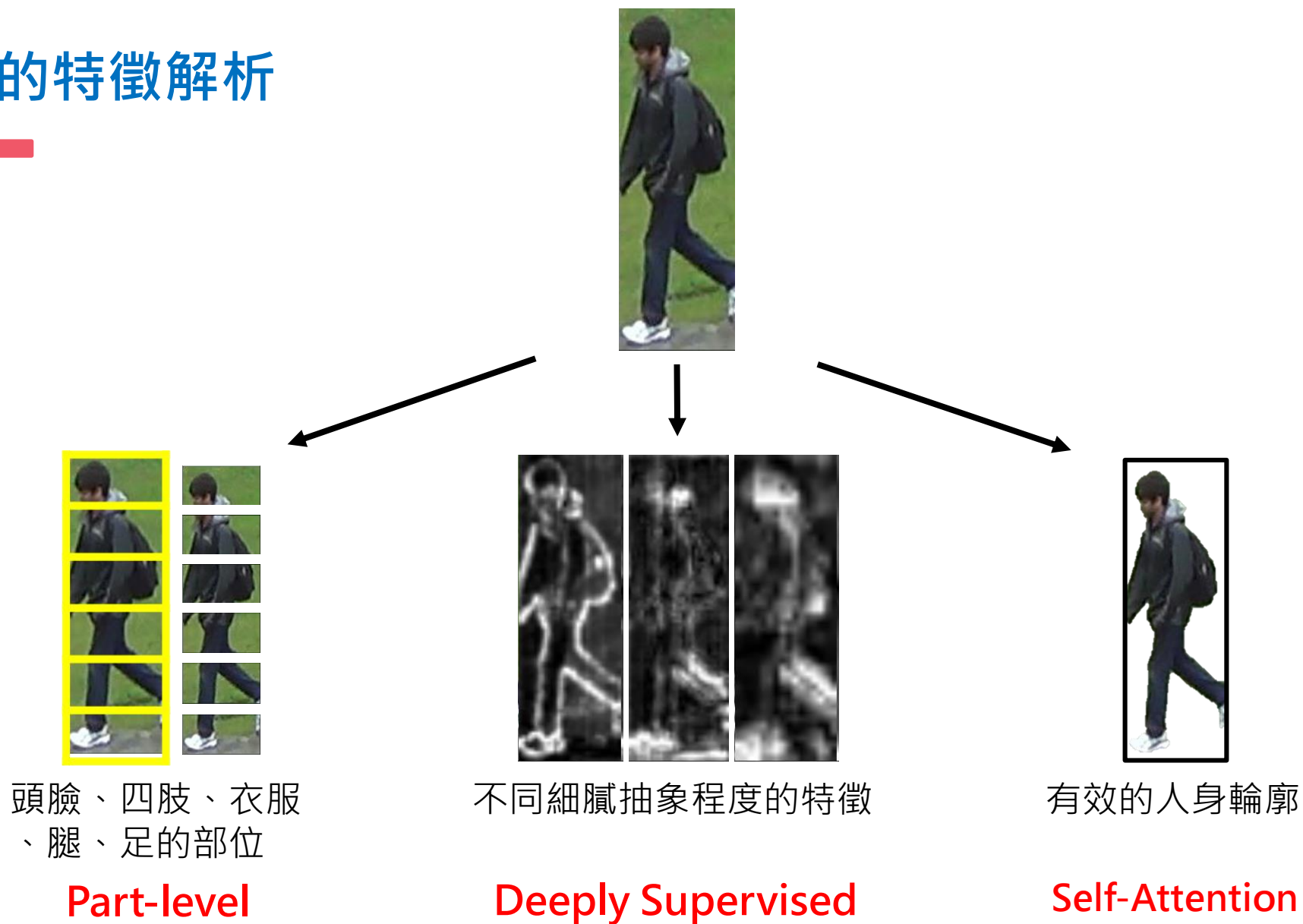
2實驗目標：電資學院三樓跨鏡追蹤

- 在固定時間區間內，比對出曾經出現的人，並記錄時間
- 專題生所作的功能
 - 針對輸入影片框選人，自動分群後，選擇目標人為Gallery，自動對Query的來源做編號分類。
- 攝影機畫面取得方式，連上EE303-5G內網後，有以下方法：

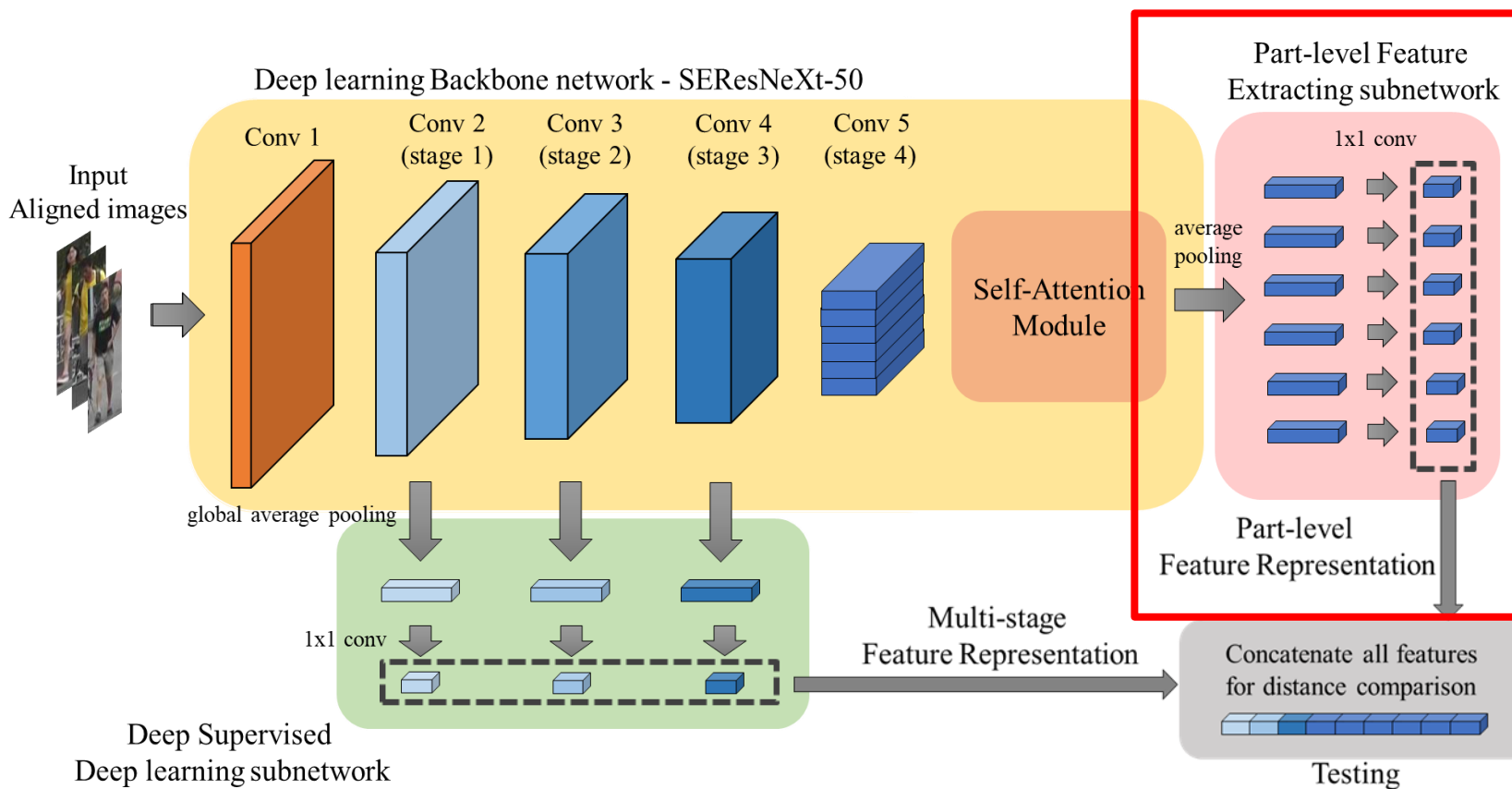
通訊方式 HTTP			
攝影機編號	解析度	FPS	URL
1	1920*1080	15	http://root:EE5040701@192.168.1.101/video3.mjpg
2			http://root:EE5040701@192.168.1.102/video3.mjpg
3			http://root:EE5040701@192.168.1.103/video3.mjpg

通訊方式 RTSP			
攝影機編號	解析度	FPS	URL
1	1920*1080	30	rtsp://root:EE5040701@192.168.1.101:6006/video1.sdp
2			rtsp://root:EE5040701@192.168.1.102:6006/video1.sdp
3			rtsp://root:EE5040701@192.168.1.103:6006/video1.sdp

行人圖像的特徵解析

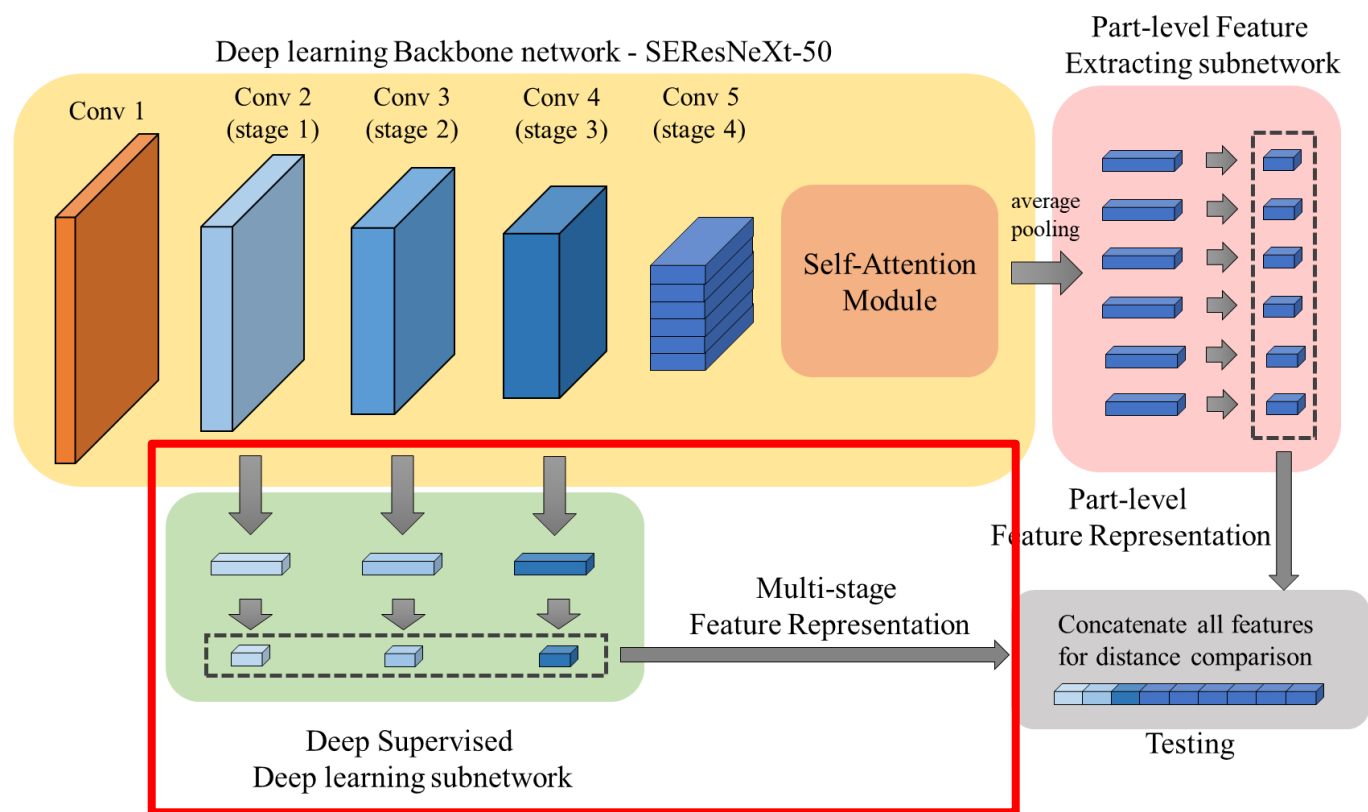


論文貢獻1： 深度監督部位特徵 (DSPF) 之深度網路架構



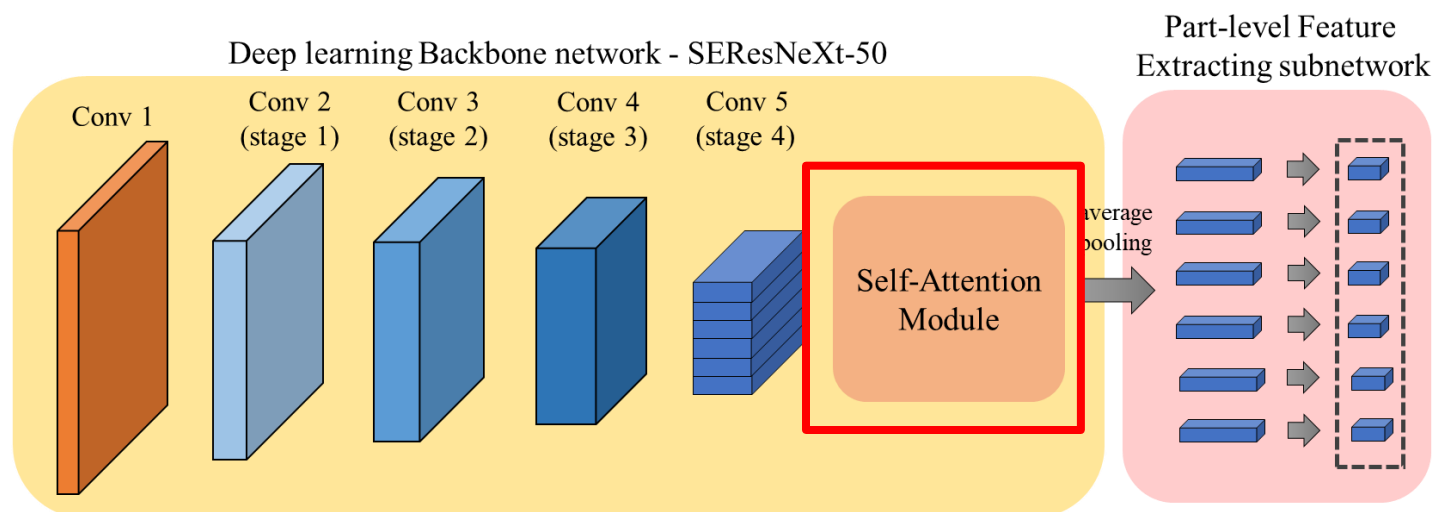
論文貢獻1： 整合網路多階段輸出的特徵

- 使用深度監督的概念加強深度學習網路學習的效率，並提取各階段輸出的不同尺度特徵，提升特徵的識別能力。



論文貢獻1： 引入自我注意力機制

- 利用整合空間與通道的注意力機制，強調行人的重點部位特徵，進一步改善骨幹網路特徵提取的能力。



與目前文獻上最好的APN的效能比，不相上下，但整體效能是更好！

3世超架構與問題及目前採用架構比較

- 1 土超留下來的系統亦是針對預錄的影片運作，且需要事先標好Ground Truth，無法針對即時畫面運作。
- 2 OpenPose在同一電腦上運作速度緩慢，不適合作為Real-time系統。
- 3 程式碼需要整理理解，輸出特徵的function code不熟悉使用。
- 4 發現待測試的問題，EE3F DataSet 盲測沒做。
- 5 目前所採用的網路架構-OSNet
 - 與土超的解決方案比較，採用原因的分析：
 - OSNet使用Torchreid library，Torchreid是採用Pytorch framework選寫，Pytorch API說明豐富易安裝及使用，可以利用Pytorch自行修改網路架構，熟悉後對後續修改自己的網路架構有幫助。
 - 已完成可以輸入任一Person即輸出feature，並計算與其他Person Bounding Boxes的相似度如何。
 - 有了每個Person Bounding Boxes的相似度關係就可以做分群，實做後續系統。
 - 可以做為baseline幫助後續研究。

3世超架構與問題及目前採用架構比較

● OSNet

- Omni-Scale Feature Learning for Person Re-Identification
- Kaiyang Zhou, Yongxin Yang, Andrea Cavallaro, Tao Xiang.
- In ICCV'19.

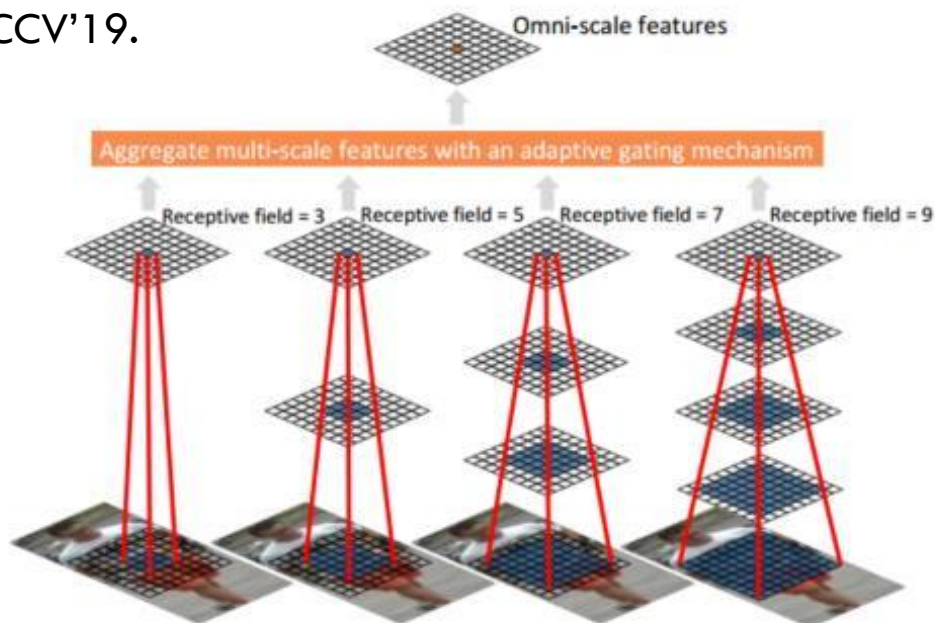


Figure 2. A schematic of the proposed building block for OSNet. The layer depth is incremented across different streams to achieve different scales. The multi-scale features are dynamically aggregated to generate omni-scale features.

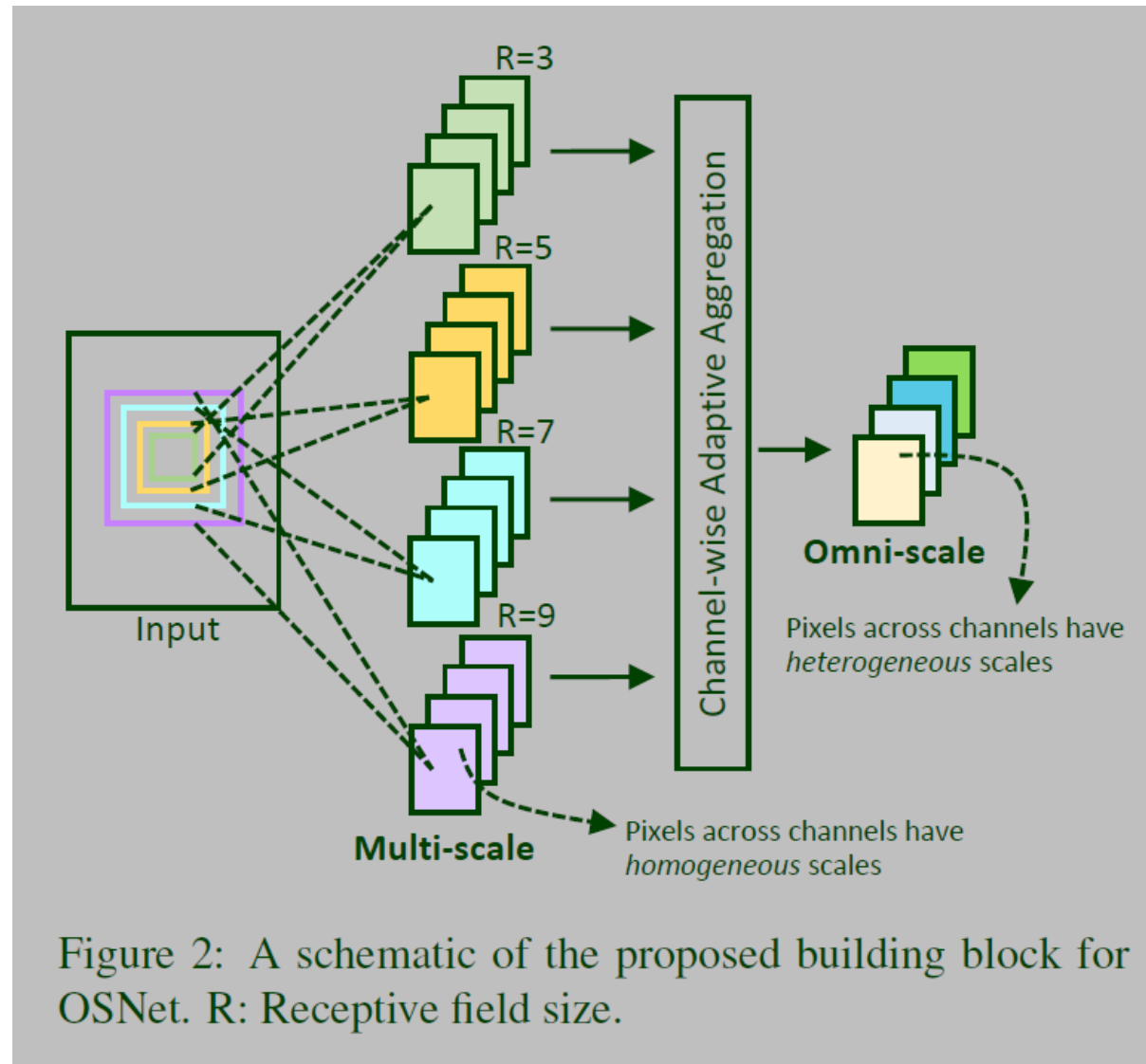


Figure 2: A schematic of the proposed building block for OSNet. R: Receptive field size.

OSNet

stage	output	OSNet
conv1	128×64, 64 64×32, 64	7×7 conv, stride 2 3×3 max pool, stride 2
conv2	64×32, 256	bottleneck × 2
transition	64×32, 256 32×16, 256	1×1 conv 2×2 average pool, stride 2
conv3	32×16, 384	bottleneck × 2
transition	32×16, 384 16×8, 384	1×1 conv 2×2 average pool, stride 2
conv4	16×8, 512	bottleneck × 2
conv5	16×8, 512	1×1 conv
gap	1×1, 512	global average pool
fc	1×1, 512	fc
# params		2.2M
Mult-Adds		978.9M

Table 1: Architecture of OSNet with input image size 256×128 .

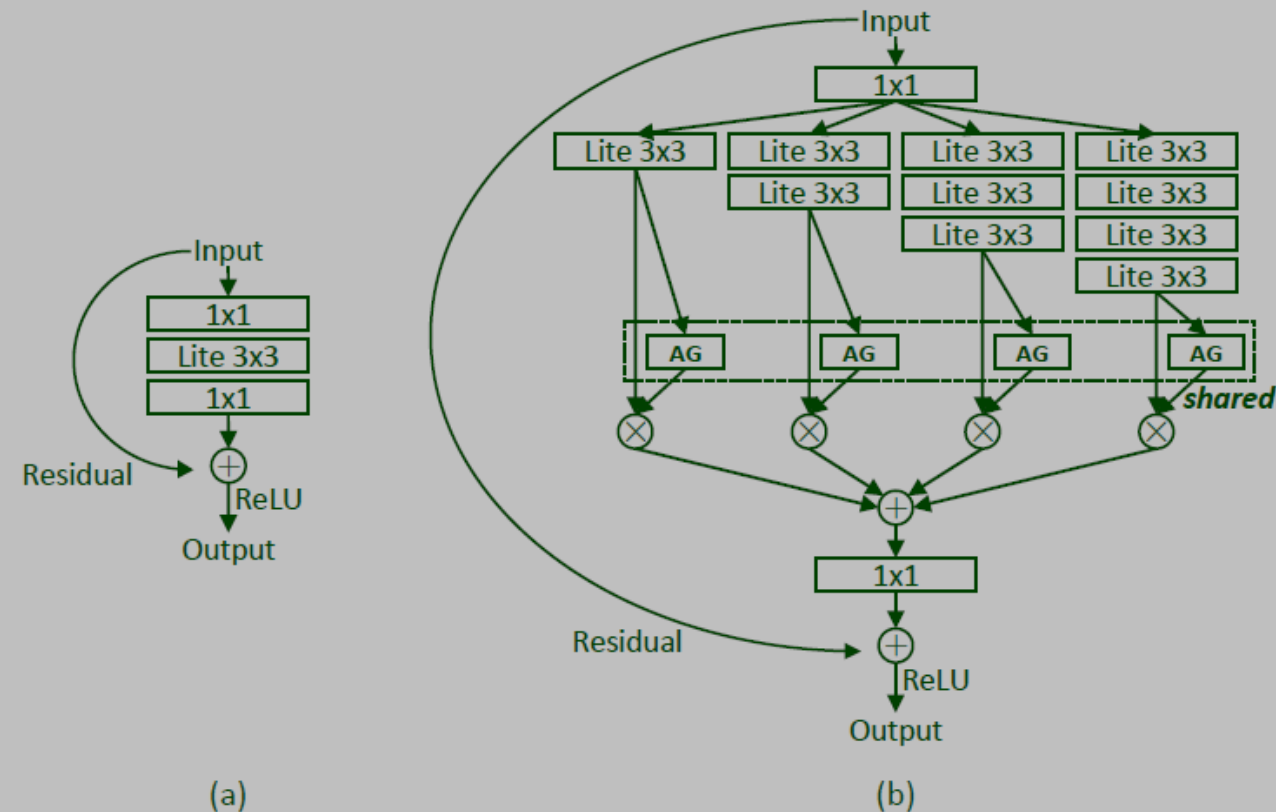
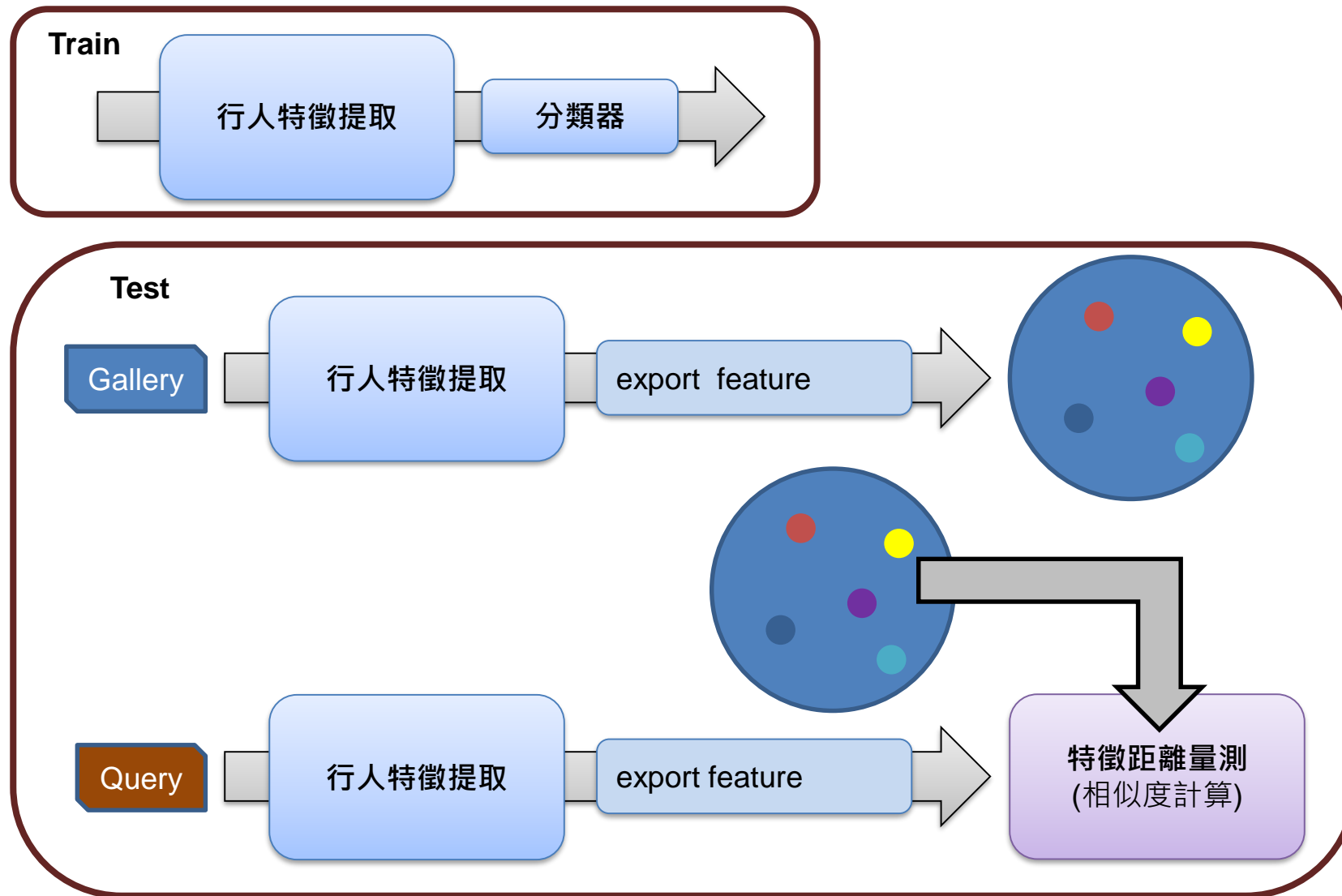


Figure 4: (a) Baseline bottleneck. (b) Proposed bottleneck. AG: Aggregation Gate. The first/last 1×1 layers are used to reduce/restore feature dimension.

Method	Publication	Backbone	Market1501		CUHK03		Duke		MSMT17	
			R1	mAP	R1	mAP	R1	mAP	R1	mAP
ShuffleNet ^{†‡} [78]	CVPR'18	ShuffleNet	84.8	65.0	38.4	37.2	71.6	49.9	41.5	19.9
MobileNetV2 ^{†‡} [43]	CVPR'18	MobileNetV2	87.0	69.5	46.5	46.0	75.2	55.8	50.9	27.0
BraidNet [†] [63]	CVPR'18	BraidNet	83.7	69.5	-	-	76.4	59.5	-	-
HAN [†] [29]	CVPR'18	Inception	91.2	75.7	41.7	38.6	80.5	63.8	-	-
OSNet [†] (ours)	ICCV'19	OSNet	93.6	81.0	57.1	54.2	84.7	68.6	71.0	43.3
DaRe [64]	CVPR'18	DenseNet	89.0	76.0	63.3	59.0	80.2	64.5	-	-
PNGAN [39]	ECCV'18	ResNet	89.4	72.6	-	-	73.6	53.2	-	-
KPM [46]	CVPR'18	ResNet	90.1	75.3	-	-	80.3	63.2	-	-
MLFN [2]	CVPR'18	ResNeXt	90.0	74.3	52.8	47.8	81.0	62.8	-	-
FDGAN [11]	NeurIPS'18	ResNet	90.5	77.7	-	-	80.0	64.5	-	-
DuATM [47]	CVPR'18	DenseNet	91.4	76.6	-	-	81.8	64.6	-	-
Bilinear [52]	ECCV'18	Inception	91.7	79.6	-	-	84.4	69.3	-	-
G2G [44]	CVPR'18	ResNet	92.7	82.5	-	-	80.7	66.4	-	-
DeepCRF [3]	CVPR'18	ResNet	93.5	81.6	-	-	84.9	69.5	-	-
PCB [53]	ECCV'18	ResNet	93.8	81.6	63.7	57.5	83.3	69.2	68.2	40.4
SGGNN [45]	ECCV'18	ResNet	92.3	82.8	-	-	81.1	68.2	-	-
Manacs [60]	ECCV'18	ResNet	93.1	82.3	65.5	60.5	84.9	71.8	-	-
AANet [56]	CVPR'19	ResNet	93.9	83.4	-	-	87.7	74.3	-	-
CAMA [71]	CVPR'19	ResNet	94.7	84.5	66.6	64.2	85.8	72.9	-	-
IANet [17]	CVPR'19	ResNet	94.4	83.1	-	-	87.1	73.4	75.5	46.8
DGNet [84]	CVPR'19	ResNet	94.8	86.0	-	-	86.6	74.8	77.2	52.3
OSNet (ours)	ICCV'19	OSNet	94.8	84.9	72.3	67.8	88.6	73.5	78.7	52.9

Table 3: Results (%) on big re-ID datasets. It is clear that OSNet achieves state-of-the-art performance on all datasets, surpassing most published methods by a clear margin. It is noteworthy that *OSNet has only 2.2 million parameters*, which are far less than the current best-performing ResNet-based methods. -: not available. †: model trained from scratch. ‡: reproduced by us. (Best and second best results in red and blue respectively)



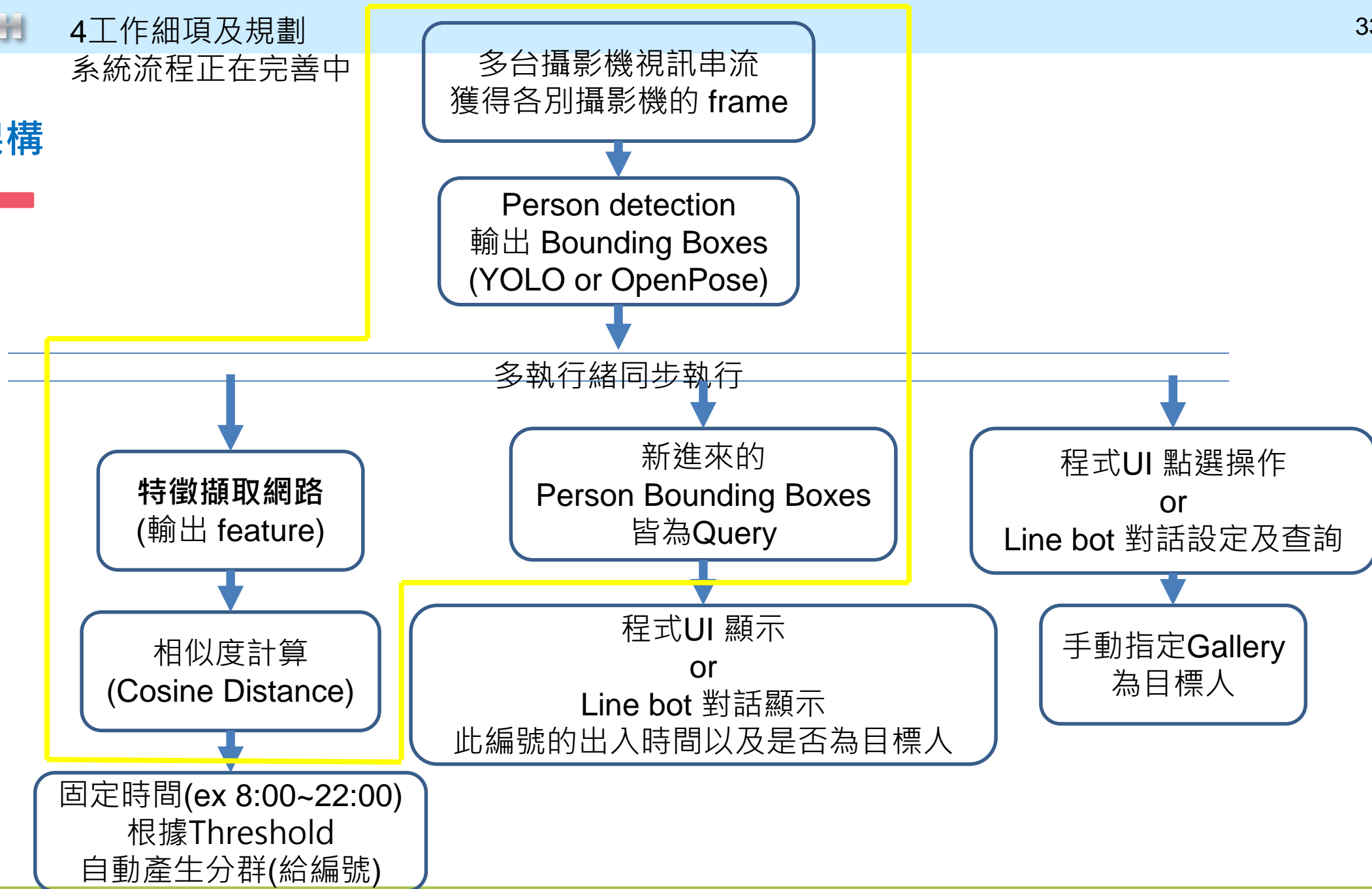
4工作細項及規劃

1. Person detection	Object detection (YOLO) Openpose	Done cont.
2. 特徵擷取網路架構	Torchreid (A Library for Deep Learning Person Re-Identification in Pytorch) backbone: OSNet Tensorflow -> Pytorch 世超-DSPF (backbone: SE-ResNeXt)	Done cont.
3. 動態每天分群給編號 4. 使用3的分群編號， 去做Query識別	Feature Extractor Compute distance matrix Real-time system (design : client 、 server 、 multithreading)	Done Done cont.
5. 人眼確認編號的確實身份	結合Line bod應用	cont.

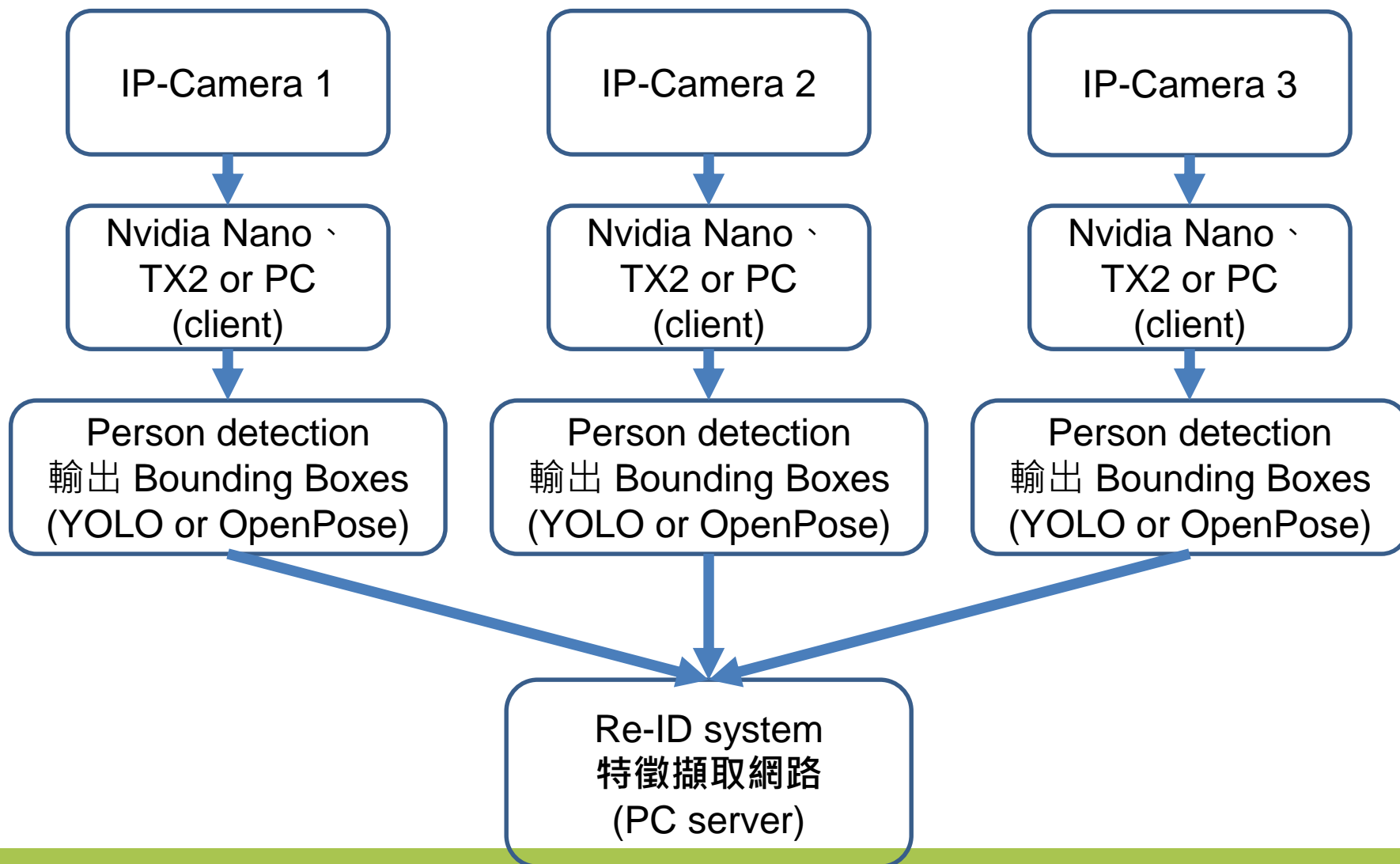
我和專題生皆使用OSNet完成了針對行人的特徵提取以及相似度計算，透過程式設計可以讓Person Bounding Boxes分群(很像的人為同一群，同一編號)。

4工作細項及規劃
系統流程正在完善中

整體系統架構

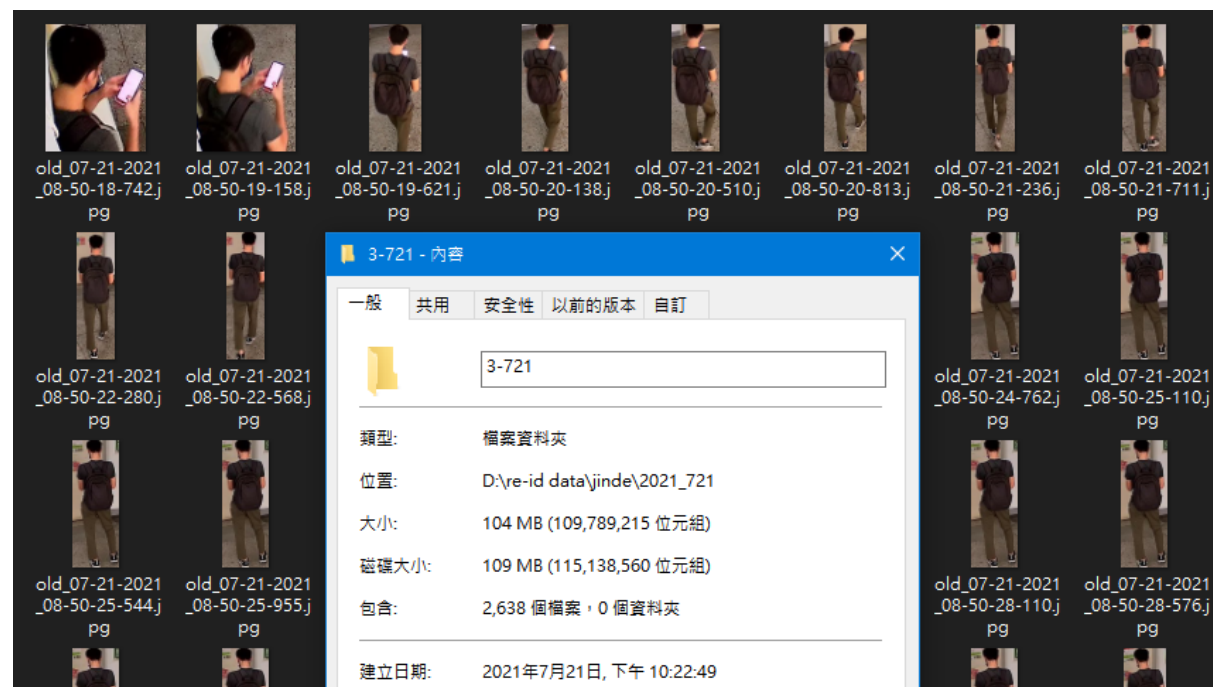
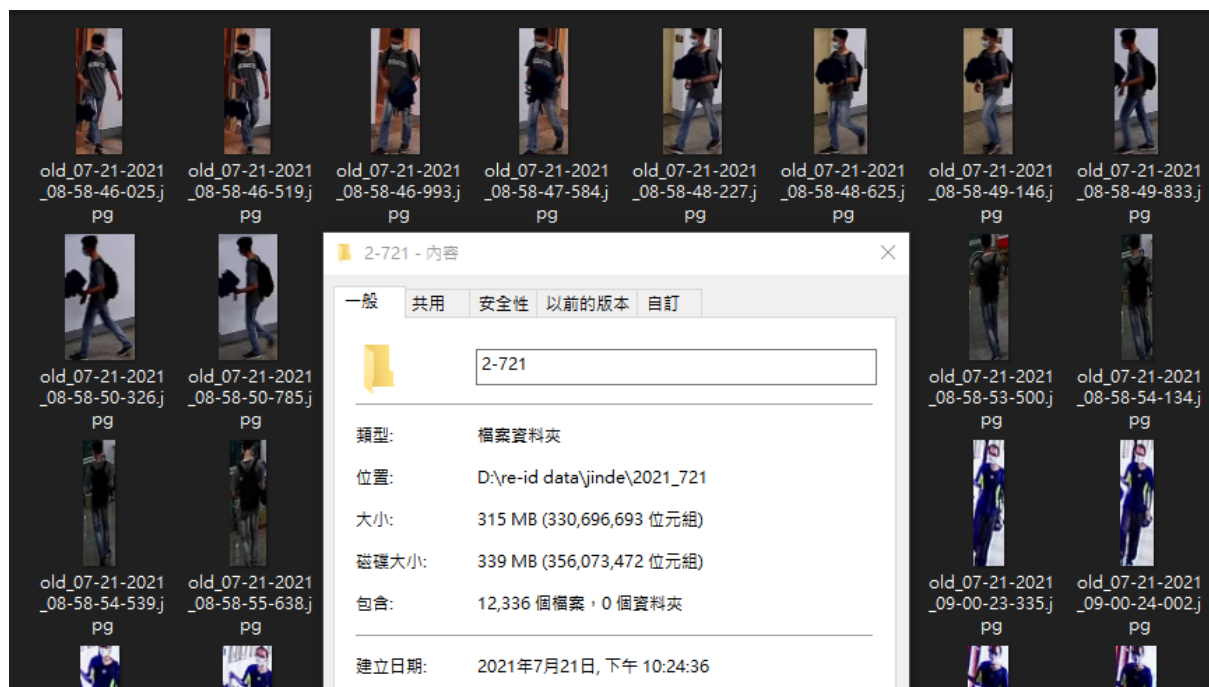


整體系統架構 (利用Edge computing 設計 client and server 解決 Real-time 延遲問題)



目前收集的data正在分類中

2021/7/21 08:50～22:20	Total images	Number of Identity
Ip-cam 2	12336	?
Ip-cam 3	2638	?



在同一台電腦上，2台IP-Cam同時運作YOLO
FPS約3~5

整合系統完成後，想辦法提高準度，有以下方法：

- 1 利用EE3F data set等更多data set，加入訓練OSNet，提高Generalization及辨識能力。
- 2 根據士超-DSPF的邏輯架構，使用Pytorch-Torchreid重現，取得Person的feature。
- 3 改良OSNet架構，設計事後Re Ranking等。
- 4 結合1、2、3，整合OSNet及士超-DSPF網路架構的邏輯，重新設計網路。