AI CUP 2024 春季賽

AI驅動出行未來:跨相機多目標車輛追蹤競賽

隊伍:TEAM_5179

隊員:邱啟翰、石旻翰、胡雅晴 (隊長)、王政邦

Private leaderboard:0.927524993 / Rank 24

壹、環境

本研究的跨相機多目標車輛追蹤模型在以下硬體和軟體環境下進行開發訓練:

一、作業系統: 本次實驗使用的作業系統為 Ubuntu 22.04 LTS

二、硬體規格:

1. **CPU**:Intel Core i7-9700F, 擁有8個核心和8個執行緒, 基本頻率為3.00GHz, 最大頻率為4.70GHz。
2. **GPU**:NVIDIA GeForce RTX 4090, 擁有24 GB GDDR6X 記憶體, 提供16384 CUDA 核心。
3. 記憶體:系統配備32 GB DDR4 RAM。

三、程式語言與套件版本:

1. 程式語言:Python 3.7.12
2. 主要函式庫:

套件	版本	套件	版本
numpy	1.21.6	opencv-python	4.7.0.72
loguru	0.7.2	scikit-image	0.19.3
scikit-learn	1.0.2	tqdm	4.66.2
torchvision	0.12.0+cu113	Pillow	9.5.0
thop	0.1.1.post2209072238	ninja	1.11.1.1
tabulate	0.9.0	tensorboard	2.11.2
lap	0.4.0	filterpy	1.4.5
h5py	3.8.0	matplotlib	3.5.3
scipy	1.7.3	seaborn	0.12.2
prettytable	3.7.0	easydict	1.13
pyyaml	6.0.1	yacs	0.1.8
termcolor	2.3.0	gdown	4.7.3
onnx	1.8.1	onnxruntime	1.8.0
ultralytics	8.0.145	xmltodict	0.13.0
onnx-simplifier	0.3.5		

四、預訓練模型：本研究中使用了以下預訓練模型作為基礎進行微調：

1. ByteTrack
2. YOLOX
3. ReID
4. YOLOv7
5. YOLOv8

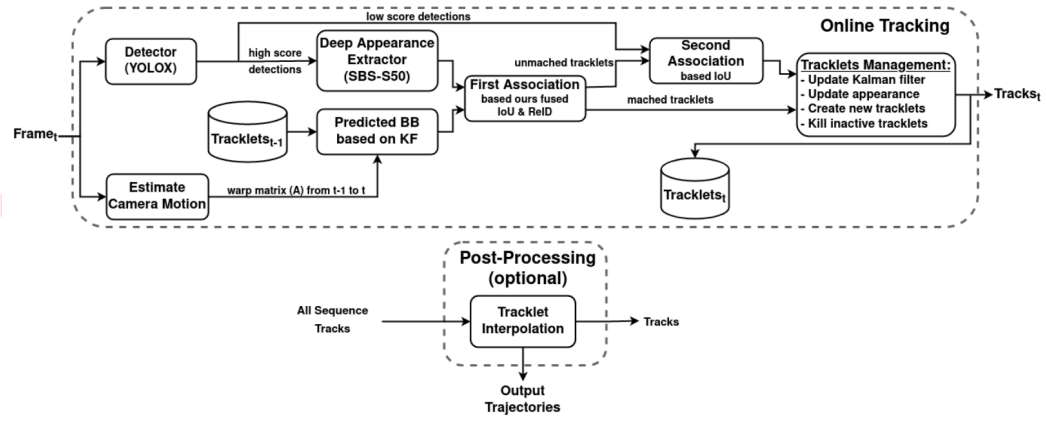
五、額外資料集：無

貳、演算方法與模型架構

一、BOTSort

BOTSort 是一種基於 SORT（Simple Online and Realtime Tracking）的多目標追蹤方法。該方法著重於計算效率和即時性，適合於資源有限的場景中應用。在 SORT 演算法的基礎上，BOTSort 所作的主要修改和改進包含：

1. 物體運動建模：BOTSort在 SORT 的基礎上更動了卡爾曼濾波器的向量定義以及衡定速度假設。由於直接估計邊界框的寬度和高度的效果較長寬比效果好，狀態向量從七元調整為八元，並且依此更改了過程噪聲協方差矩陣及測量噪聲協方差矩陣。
2. 攝相機運動補償：為了減少攝相機晃動對預測結果造成的影響，BOTSort 採用了 OpenCV 使用的 global motion compensation (GMC) 技術來處理這一問題。透過稀疏的圖像配准技術，可以更準確地估計背景運動，從而忽略場景中的動態物體。
3. **IoU-ReID** 融合：BOTSort綜合了 IoU的空間資訊和 ReID 的外觀資訊來進行相似性匹配步驟。

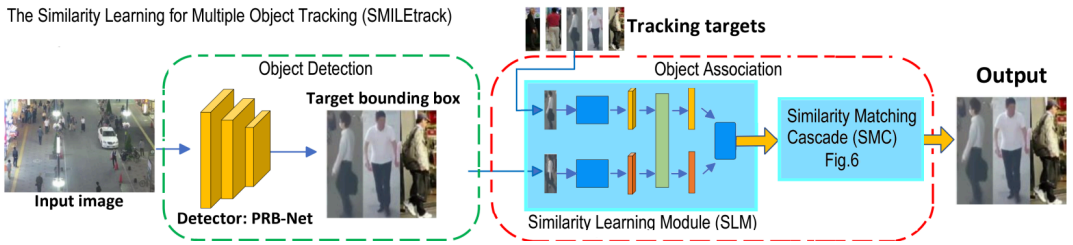


圖一：BoT-SORT 模型架構[1]

二、SMILETrack

相較於 BOTSort，SMILETrack 在多目標追蹤技術上進行了多方面的創新和改進使其在複雜和動態變化的場景中表現更加優異。該演算法採用分離檢測和嵌入（SDE）策略，融合了高效物體檢測器和基於Siamese網絡的相似性學習模塊。

1. 物體檢測：SMILETrack 採用了更為先進的 YOLOv7 模型進行物體檢測。相比於 BOTSort 可能使用的早期 YOLO 版本，YOLOv7 提供了更高的檢測準確性和速度。這史的 SMILETrack 能更精準的檢測結果，直接提高了追蹤的基礎數據質量。
2. 相似性學習模塊（**SLM**）：BOTSort 僅依賴基於物理位置的匹配，而 SMILETrack 增加了相似性學習模塊。SLM通過Siamese網絡計算兩個目標物體的外觀相似性。該模塊包含一個受視覺Transformer啟發的Patch Self-Attention（PSA）區塊，生成可靠的特徵進行相似性匹配。PSA區塊利用自注意力機制來生成每個物體的特徵向量，從而在處理遮擋和外觀相似的情況下表現出色。
3. 相似性匹配級聯（**SMC**）模塊：SMILETrack 在相似性匹配步驟中加入了 SMC 模塊。SMC模塊是一個使用新穎GATE函數來進行連續視頻幀之間的物體匹配。SMC模塊通過匈牙利算法來解決線性分配問題，從而提高匹配的準確性和穩定性。



圖二：SMILETrack 模型架構[2]

本次競賽中，我們以 BOT-SORT 與 SMILETRACK 兩架構為核心同步進行研究，藉由調整訓練參數以及嘗試不同的 detection 模型來優化模型效果。

參、創新性

本次作品主要創新的點其一為我們將 BOT-SORT 演算法進行拆分，並藉由嘗試將多種檢測模型橫向比較來找出最優的模型組合，包括 YOLOv7、YOLOv8 和 YOLOX。我們深入分析了每個檢測模型的優勢和劣勢，並通過實驗驗證來確定哪個模型在特定情況下能夠提供最佳的性能。這樣的橫向比較不僅提高了檢測精度，還優化了整體系統的效率和可靠性。

其二，我們在閱讀大量相關論文後，選擇並嘗試實驗了數個目前在多目標追蹤（MOT）任務中表現非常突出的架構，這些架構包括 SMILETrack、BoostTrack 和 UCMCTrack。我們經過討論後，考慮到後兩者對於 tracking 的改動甚多，且由於此次比賽的資料集為低幀率，連 Kalman Filter 也無法達成預期效果，因此我們最後決定同時也對 SMILETrack 進行研究。

我們的工作不僅限於單一方法的改進，更是將多種先進技術融合，通過創新的組合和優化實現了性能的顯著提升。這些創新點不僅為我們在此次比賽中取得了不錯的成績，還為未來相關領域的研究提供了新的思路和方法。通過拆分 BOT-SORT 演算法並結合最優檢測模型，以及應用先進的 MOT 架構，我們在多目標追蹤任務中實現了突破。

肆、資料處理

在此次比賽中，我們面臨的資料集幀率極低，這對於多目標追蹤任務提出了額外的挑戰。低幀率資料集通常會導致目標之間的運動模糊，進而影響追蹤的準確性。我們擔心在追蹤端引入其餘的車輛 MOT 資料集進行訓練，可能會導致在實際推斷過程中結果不如預期。

在檢測端，我們選用了性能優異的檢測模型。這些模型在多目標檢測任務中表現出色，我們認為它們已經足夠應對此次競賽中的檢測需求。經過與此次競賽提供的標籤進行比對，我們發現競賽所需的檢測精確度相對較低。因此，我們決定不在檢測任務中使用額外的資料進行訓練，這樣可以避免引入不必要的變數，保持模型的簡潔和高效。

雖然我們沒有實質使用額外的資料集進行訓練，但我們仍然採取了一些策略來增加模型的泛用度。首先，我們使用了 fast-reid 模型，這是一個強大的重新識別工具。該模型從其餘 VehicleID 任務的預訓練模型進行微調，在 validation set 上取得了一些分數提升，這讓我們對模型在實際應用中的表現更有信心。通過這種方法，我們可以減少模型過擬合（overfit）的風險，確保其在不同場景下具有較好的泛化能力。

針對此次資料集的特殊性，我們進行了大量的影像增強（image augmentation）操作。由於資料集中存在不少車輛重疊並被遮去部分面積的情況，這些增強技術在 reidentification 時尤為重要。當車輛部分被遮擋時，角度和顏色差異會使得模型難以準確識別同一車輛。為了應對這些挑戰，我們選用了多種影像增強方法，包括 random erase augmentation、random affine、flip 及 auto augmentation。

- **Random Erase Augmentation**：這種方法通過隨機擦除影像的一部分來模擬遮擋情況，使得模型能夠學習在部分遮擋下仍能準確識別車輛。這對於應對實際場景中車輛被其他物體部分遮擋的情況非常有幫助。

- **Random Affine**：這種方法對影像進行隨機仿射變換，包括旋轉、縮放和平移。通過這種方式，我們能夠模擬不同的拍攝角度和距離，提升模型對各種角度和視角下車輛的識別能力。
- **Flip**：隨機翻轉影像是另一種簡單而有效的增強方法。它可以讓模型學習到車輛在不同方位的外觀特徵，增加模型的穩健性。
- **Auto Augmentation**：這是一種自動化的增強策略，通過預定義的一系列增強操作（如亮度調整、對比度調整等）隨機組合應用於影像上。這種方法可以大幅度增加訓練資料的多樣性，進而提高模型的泛化能力。

這些影像增強技術的應用，使得我們的模型在面對資料集中多變的環境和各種遮擋情況時，仍能保持較高的識別準確度。雖然在我們的測試中，使用如此技巧僅讓 validation set 提升 0.01 - 0.02 的 mota + idf1 分數，然而這樣的泛化性在 private testset 就有非常明顯的體現，讓該成績相較 public testset 不僅沒有 overflow，還讓我們名次因此提升了超過 10 名。

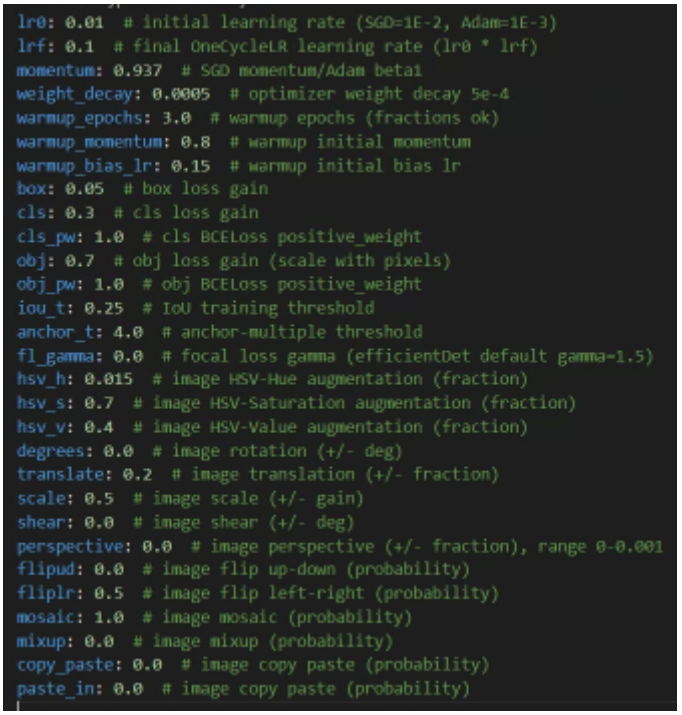
伍、訓練方式

除了擴增資料外，我們在各個實驗的實驗參數大部分都使用預設參數。由於個實驗都有相當多參數，且有些模型的參數預設值我們也沒有每個都進行研究，因此此處僅列出兩組。

1. **Fast-Reid:**

   BIAS_LR_FACTOR: 1.
   IMS_PER_BATCH: 256
   MAX_EPOCH: 60
   STEPS: [30, 50]
   CHECKPOINT_PERIOD: 1
   OPT: Adam, BASE_LR: 0.00035
   WEIGHT_DECAY: 0.0005
   WEIGHT_DECAY_NORM: 0.0005
   IMS_PER_BATCH: 64
   SCHED: MultiStepLR
   GAMMA: 0.1
   WARMUP_FACTOR: 0.1
   WARMUP_ITERS: 2000

2. **Detection:**



```yaml
lr0: 0.01  # initial learning rate (SGD=1E-2, Adam=1E-3)
lrf: 0.1  # final OneCycleLR learning rate (lr0 * lrf)
momentum: 0.937  # SGD momentum/Adam beta1
weight_decay: 0.0005  # optimizer weight decay 5e-4
warmup_epochs: 3.0  # warmup epochs (fractions ok)
warmup_momentum: 0.8  # warmup initial momentum
warmup_bias_lr: 0.15  # warmup initial bias lr
box: 0.05  # box loss gain
cls: 0.3  # cls loss gain
cls_pw: 1.0  # cls BCELoss positive_weight
obj: 0.7  # obj loss gain (scale with pixels)
obj_pw: 1.0  # obj BCELoss positive_weight
iou_t: 0.25  # IoU training threshold
anchor_t: 4.0  # anchor-multiple threshold
fl_gamma: 0.0  # focal loss gamma (efficientDet default gamma=1.5)
hsv_h: 0.015  # image HSV-Hue augmentation (fraction)
hsv_s: 0.7  # image HSV-Saturation augmentation (fraction)
hsv_v: 0.4  # image HSV-Value augmentation (fraction)
degrees: 0.0  # image rotation (+/- deg)
translate: 0.2  # image translation (+/- fraction)
scale: 0.5  # image scale (+/- gain)
shear: 0.0  # image shear (+/- deg)
perspective: 0.0  # image perspective (+/- fraction), range 0-0.001
flipud: 0.0  # image flip up-down (probability)
fliplr: 0.5  # image flip left-right (probability)
mosaic: 1.0  # image mosaic (probability)
mixup: 0.0  # image mixup (probability)
copy_paste: 0.0  # image copy paste (probability)
paste_in: 0.0  # image copy paste (probability)
```

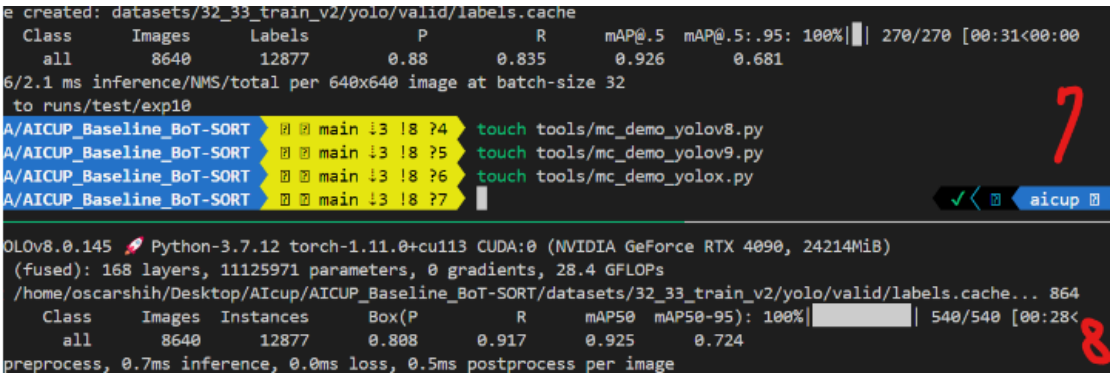
圖三：yolov7 訓練參數

除了訓練參數，我們也使用了 wandb 與 tensorboard 這兩個視覺化工具來追蹤訓練過程。

陸、分析與結論

由於此次比賽任務要求在非常低幀率下進行多物件多鏡頭偵測，因此許多傳統的追蹤技術無法有效應用。我們針對 SMILETrack 進行了一些調參處理後，發現其結果都未能達到 baseline。我們推測這是因為 SMILETrack 的追蹤前處理較為複雜，因此在面對低幀率時未能充分發揮其特色，導致其性能未達到預期。因此，在本節中，我們將重點分析和探討我們最終版本的解決方案，即 BOTSORT 結合 fast-reid-augment 和 YOLOv8 的方法。

在實施 YOLOv8 之後，我們對 YOLOv7 和 YOLOv8 的檢測能力進行了詳細的對比分析。我們選用了 YOLOv8l 和 YOLOv7e6e 進行訓練和測試。結果顯示，兩者在 mAP50 指標上均表現優異，達到了 0.925 和 0.926 的高水準。然而，在更具挑戰性的 mAP50-95 和 recall 指標上，YOLOv8 顯示出明顯的優勢，大幅領先 YOLOv7。



圖四：yolov7 與 yolov8 detection 結果

首先，在 mAP50-95 指標方面，YOLOv8 取得了顯著的領先。這表明 YOLOv8 在不同的 IoU 閾值下，都能保持較高的準確度，這對於多物件檢測任務尤為重要。YOLOv8 的優異表現可以歸因於其改進的網絡結構和更先進的特徵提取技術，使其在各種複雜場景下均能準確識別目標。

其次，在 recall 指標上，YOLOv8 同樣表現突出。高 recall 意味著模型能夠檢測到更多的真實目標，這對於低幀率的視頻分析特別重要。由於低幀率下物體移動較快，容易出現目標遺漏的情況，因此高 recall 能有效降低這類問題的發生，提高整體系統的穩健性和可靠性。

為了進一步驗證我們的方法，我們將 YOLOv8 detection model 與 fast-reid 結合，並將 baseline 的 1.17 分數一舉提升到 1.40。在手動比對 label 與我們的預測結果後，我們發現 label 其實少辨識了非常多車輛，因此我們藉由手動調整 detection threshold 到 0.6 來接近該 dataset 的標示精確度。最終我們在 train dataset 的分數提高到 1.48。

結論來說我們這次主要還是藉由調整 detection model 與 fastreid augmentation 來提收我們的正確率。實際上我們仍有許多想實作的實驗最終都來不及完成，未來希望能加入更多如 boostTrack 與 UCMCTrack 架構來比較 performance。此外，關於關掉 Kalman Filter 後的 tracking 補償也是我們未特別著墨的點。

柒、程式碼

GitHub連結：https://github.com/oscar-shih/AICUP-2024---MCMOT

捌、使用的外部資源與參考文獻

一、參考文獻

[1] Aharon, N., Orfaig, R., & Bobrovsky, B.-Z. (2022). BoT-SORT: Robust Associations Multi-Pedestrian Tracking (Version 2). arXiv. https://doi.org/10.48550/arXiv.2206.14651v2

[2] Wang, Y.-H., Hsieh, J.-W., Chen, P.-Y., Chang, M.-C., So, H. H., & Li, X. (2024). SMILEtrack: SiMIlarity LEarning for Occlusion-Aware Multiple Object Tracking (Version 4). arXiv. https://doi.org/10.48550/arXiv.2211.08824v4

報告作者聯絡資料表

隊伍名稱	TEAM_5179	Private Leaderboard 成績	0.927524993	Private Leaderboard 名次	Rank 24
身分 (隊長/隊員)	姓名 (中英皆需填寫) (英文寫法為名, 姓, 例:Xiao－Ming, Wu, 名須加連字號, 姓前須加逗號)	學校＋系所中文全稱 (請填寫完整全名, 勿縮寫)	學校＋系所英文中文全稱 (請填寫完整全名, 勿縮寫)	電話	E-mail
隊長	胡雅晴 Ya-Chin, Hu	國立台灣大學電機工程學系	National Taiwan University, Department of Electrical Engineering	0955-415-299	jessicahu1012@gmail.com
隊員1	邱啟翰 Chi-Han, Chiu	國立台灣大學電機工程學系	National Taiwan University, Department of Electrical Engineering	0910-246-080	ericcm509mh@gmail.com
隊員2	石旻翰 Min-Han, Shih	國立台灣大學電機工程學系	National Taiwan University, Department of Electrical Engineering	0966-418-657	mhshih0405@gmail.com
隊員3	王政邦 Jeng-Bang, Wang	國立台灣大學電機工程學系	National Taiwan University, Department of Electrical Engineering	0984-380-925	r12943022@ntu.edu.tw
指導教授資料					
每隊伍至多可填寫兩名	指導教授 中文姓名	指導教授 英文姓名 (英文寫法為名,姓, 例:Xiao－Ming, Wu, 名須加連字號, 姓前須加逗號)	任職學校＋系所 中文全稱 (請填寫完整全名, 勿縮寫)	任職學校＋系所 英文全稱 (請填寫完整全名, 勿縮寫)	E-mail
教授 1					
教授 2					

★註1:請確認上述資料與AI CUP報名系統中填寫之內容相同。自2023年起，獎狀製作將依據報名系統中填寫內容為準，有特殊狀況需修正者，請主動於報告繳交期限內來信moe.ai.ncu@gmail.com。，報告繳交截止時間後將不予修改。

★註2:繳交程式碼檔案與報告，請Email至:ccumvllab@gmail.com，並同時副本至:t_brain@trendmicro.com與moe.ai.ncu@gmail.com。缺一不可。