

B-2

GMC-Link

全域運動補償之 指向性多目標追蹤

*Referring Multi-Object Tracking
aided by Global Motion Compensation*

01 任務背景 · THE TASK

指向性多目標追蹤 (RMOT)

給一句**動作或外型**的描述，再給一段**影片**，模型須從畫面中眾多目標裡，找出**所有**符合描述 的物體並持續追蹤。



實現Artificial General Intelligence

是自動駕駛、智慧監控與影片檢索的共同基礎能力。

02 研究動機及缺口 · THE GAP

現有方法忽略了相機運動

A

仰賴視覺外觀特徵

以顏色、形狀等外觀線索匹配描述

對「運動」這類語意天生不敏感

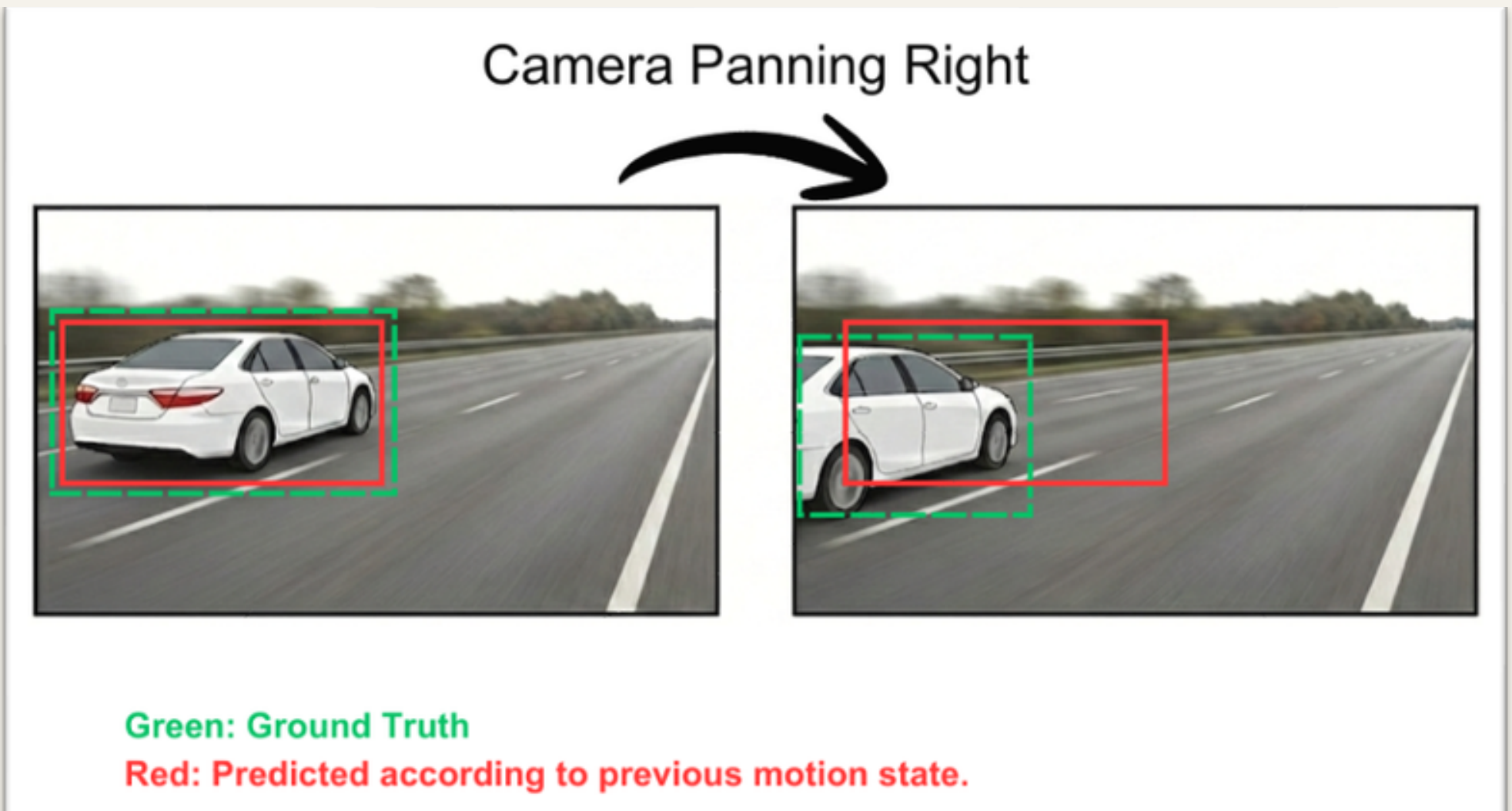
B

僅用原始邊界框位移

直接以 bounding box 的位移當運動線索

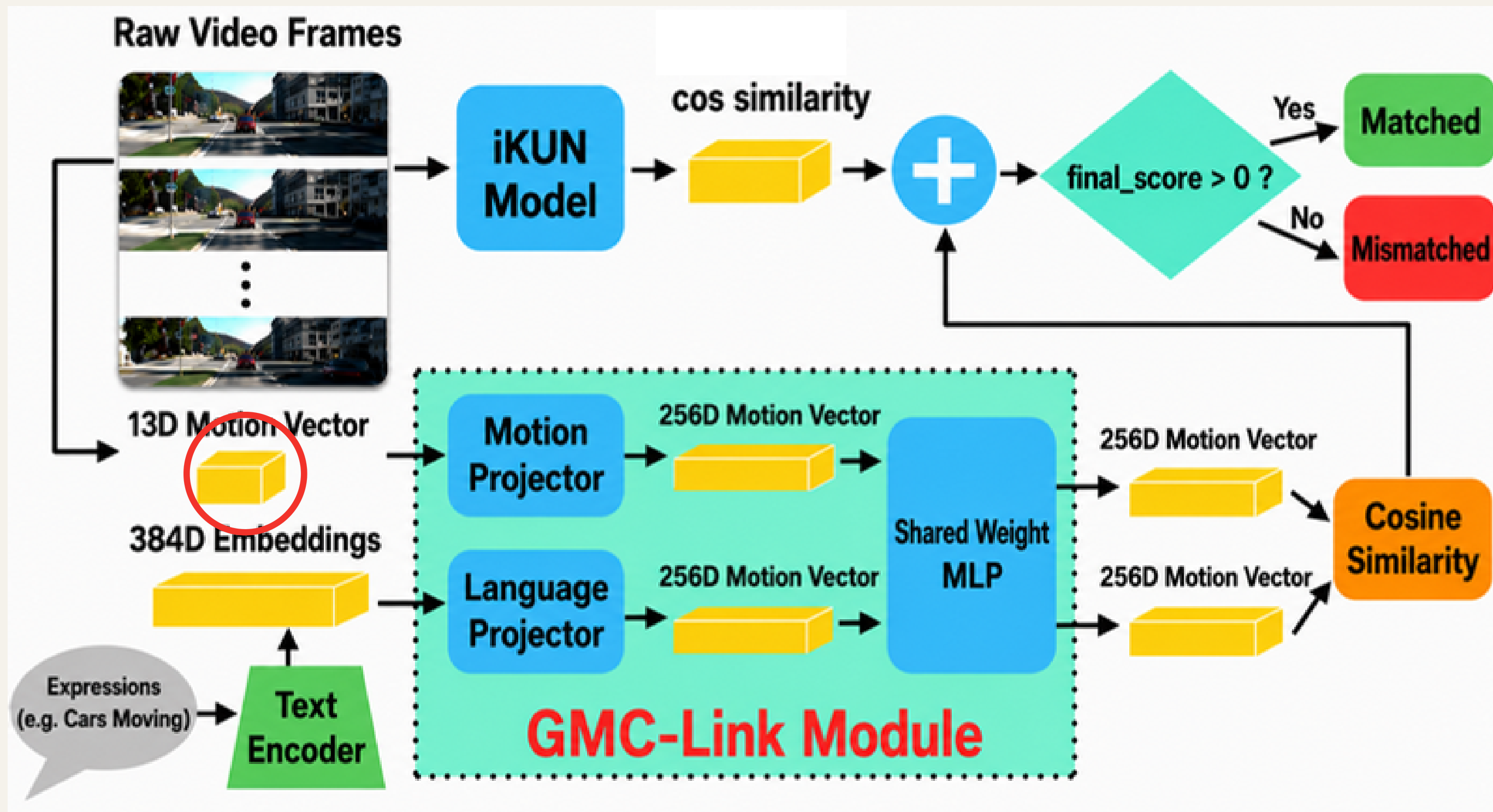
卻把相機運動與物體運動混在一起。

觀測位移 = 物體真實運動 + 相機自我運動 (*ego-motion*)



03 方法總覽 · GMC-LINK

GMC-Link 系統架構



(圖一) GMC-Link 系統架構 — GMC 模組以決策層加成方式接入，不更動原模型。

04 方法一 · GMC

全域運動補償：先把相機「扣掉」

$$\text{殘餘速度} = \text{物體原始速度} - \text{相機速度}$$

01 ORB 特徵點

於相鄰影格擷取背景特徵點。



02 BFMatcher 比對

Hamming 距離 + Lowe's ratio 0.7 篩選對應點。



03 RANSAC

穩健估計、剔除離群對應。



04 Homography

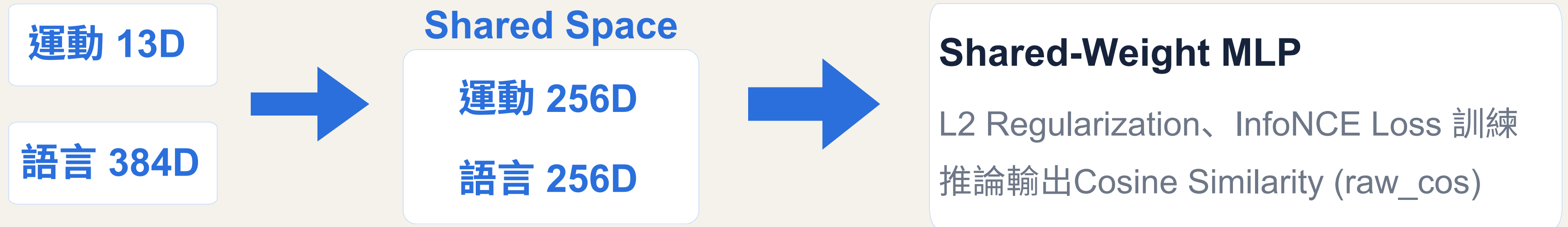
建立影格間的背景對映矩陣。

保存物件**原始未變形**的中心座標，以累積複合單應性一次性校正

輸出含多尺度位移、寬高變化、中心座標、尺寸與訊噪比的 **13 維運動向量**。

05 方法三 · ALIGNMENT & FUSION

運動—語言對齊與決策融合



語言端以 SentenceTransformer (all-MiniLM-L6-v2, 384D) 編碼描述句。

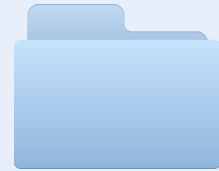
$$\text{final score} = \text{model_logit} + \alpha \cdot \text{raw_cos}$$

決策層線性加成 — 將 GMC 運動分數加到原模型輸出，依運動／外觀兩軸分別校準。

06 實驗設定 · SETUP

Refer-KITTI × 三種追蹤架構

2



資料集

Refer-KITTI

Refer-KITTIv2

2



基礎架構

iKUN

FlexHook

HOTA



評估指標

High-Order Tracking
Accuracy (HOTA)

Plug and Play 以決策層加成接入，**不重訓**基礎模型，比較融合前後表現。

07 結果一 · OVERALL

整體 Pooled HOTA 表現

架構 · 資料集	原模型	+ GMC-Link	較原模型
iKUN · V1	44.224	44.634	+0.410
FlexHook · V1	53.110	53.526	+0.416
FlexHook · V2	42.526	42.807	+0.281

(表一) 融合前後 Pooled HOTA 比較。

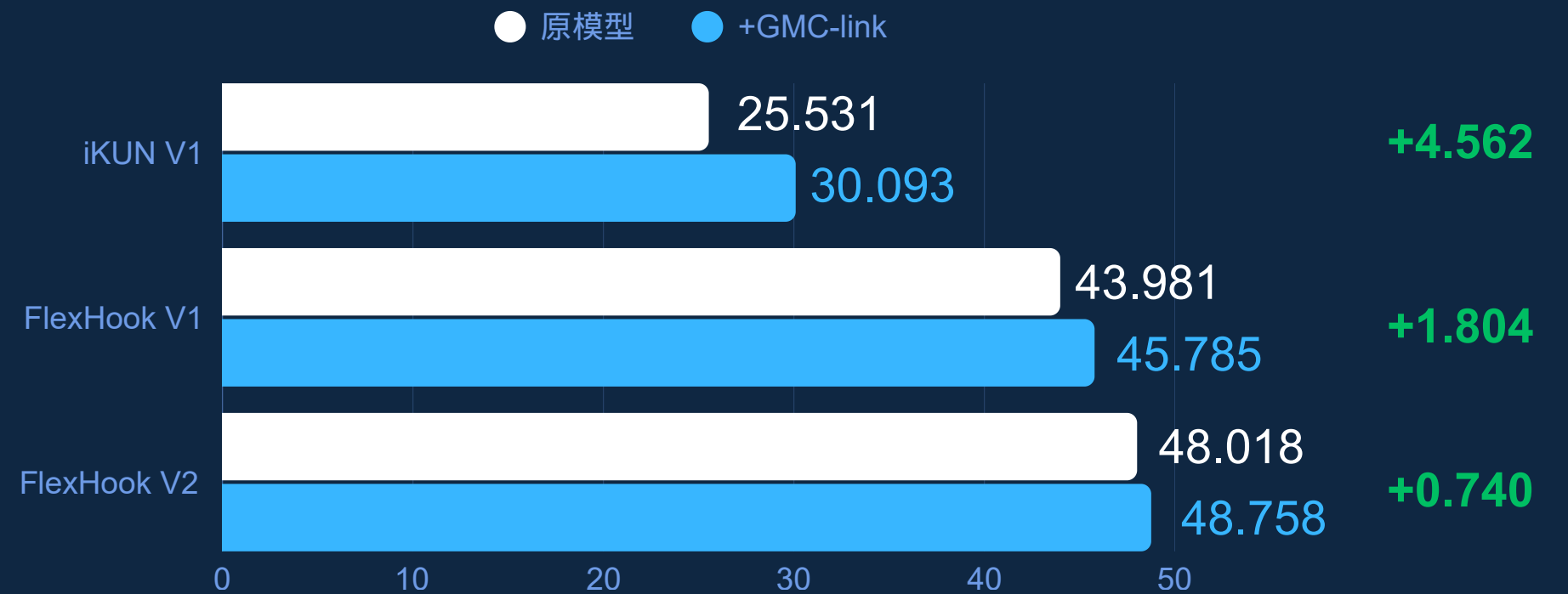
Conclusion 三種架構皆獲得整體增益，且超越原模型表現。

08 結果二 · KEY RESULT

GMC-link顯著改善運動描述的追蹤效能

+17.9%

iKUN 在運動描述的 pooled HOTA 提升 (25.531 → 30.093)



為什麼提升最大

iKUN缺乏運動建模能力
對相機運動與物體運動區分不足

GMC-link 做了什麼

- 消除相機運動(Ego-Motion)干擾
- 強化真實物體運動描述
- 提升運動場景下的匹配精準度

結果

- 運動類別提升最高達+17.9%
- 有效縮小與靜止場景性能差距
- 對運動描述弱的模型幫助最大

哪個設計最關鍵？

1.Ego-Motion的補償是決定性元件。

相較直接採用原始邊界框位移

ORB + 單應性能有效分離相機運動，是運動描述能否正確匹配的關鍵。

2.多尺度時速度貢獻最大

同時捕捉短／中／長期運動模式，是對齊分離度提升幅度最大的單一設計。

3. 決策層的融合

決策層加成能保留原模型能力，亦可保留在不同模型中的彈性調整。

10 結論 · CONTRIBUTIONS

結論與貢獻

提出GMC-Link

以全域運動補償萃取真實殘值運動
補上既有RMOT對運動語意的盲點

即插即用不需重訓

以決策層加成接入，可泛化至多種對運動不敏感
(spatially-ignorant)的視覺-語言追蹤框架

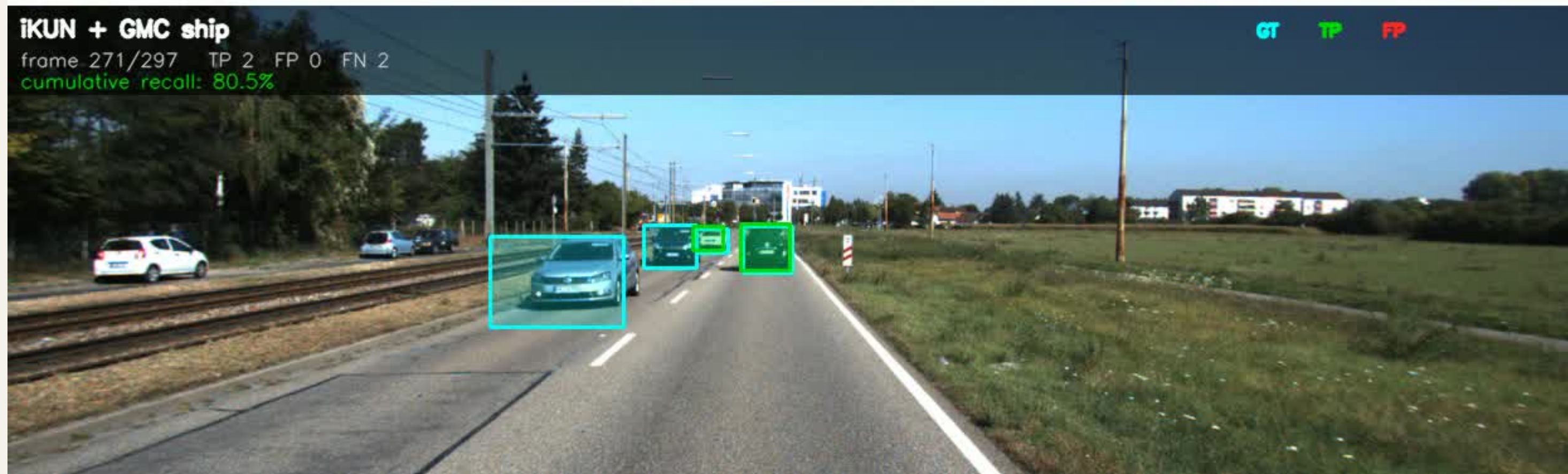
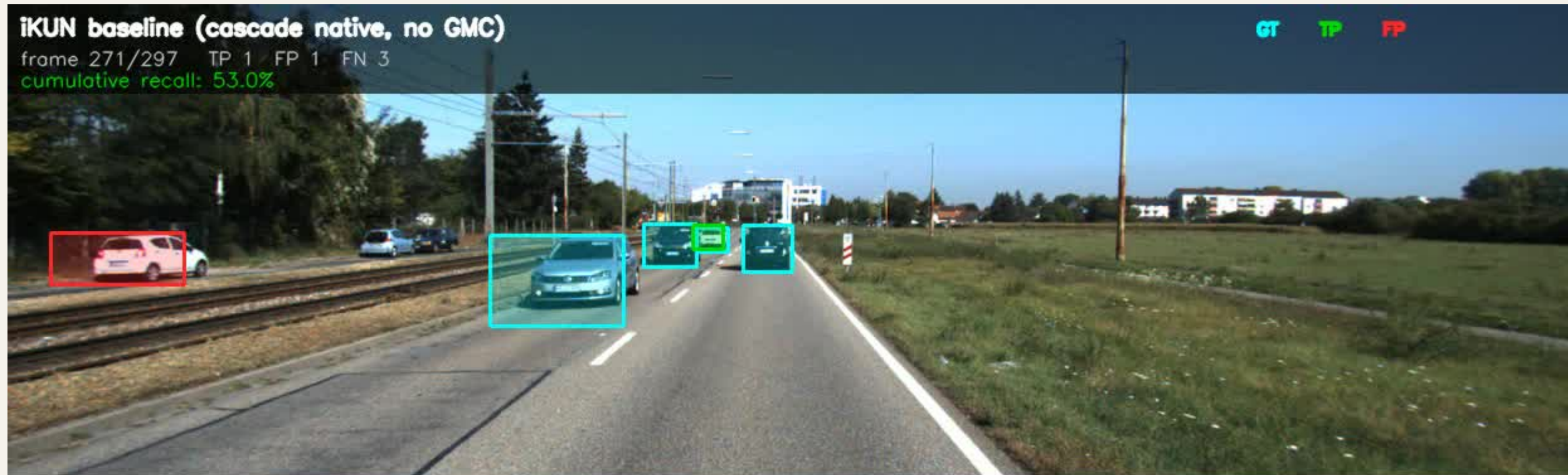
逐類別關聯性調節

運動、外觀兩軸採不同校準尺度
抑制純外觀描述(ex: black-cars)上的運動雜訊

完整增強流程

從感知、對齊、融合到判定
形成一致的端到端增益鏈

Demo Video



— THANK YOU · Q & A

GMC-Link

Referring Multi-Object Tracking aided by Global Motion Compensation

指導教授 朱威達

專題成員 白昕宸 · 王駿愷 · 彭以呈