

全域運動補償之 指向性多目標追蹤

Referring Multi-Object Tracking aided by Global Motion Compensation · (GMC-Link)

指導教授
朱威達

專題成員
白昕宸 · 王駿愷 · 彭以呈

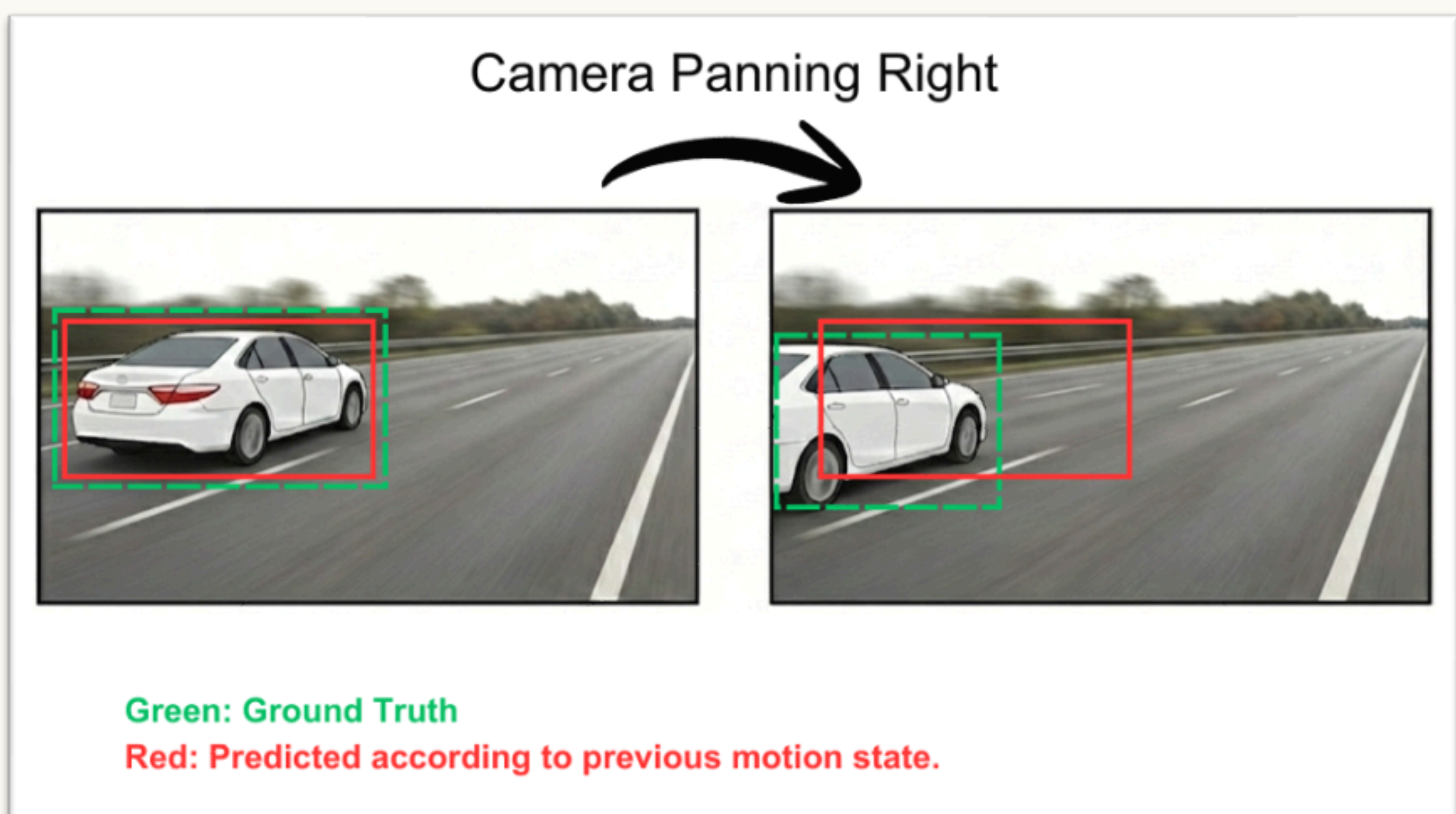
測試環境
Ubuntu 24.04 · RTX 3060 Ti

一、摘要

指向性多目標追蹤（**RMOT**）在輸入影片與一句自然語言描述後，找出畫面中所有符合描述的目標。然而車載攝影機 **自身的運動** 會使畫面位移混雜物體運動與相機運動，使涉及運動語意描述難以正確匹配。本專題提出 **GMC-Link**，以全域運動補償分離相機自我運動、萃取物體真實殘餘運動，與語言特徵對齊後以決策層加成融合至既有 RMOT 模型，**無須重新訓練**即可提升運動類別描述的追蹤表現。

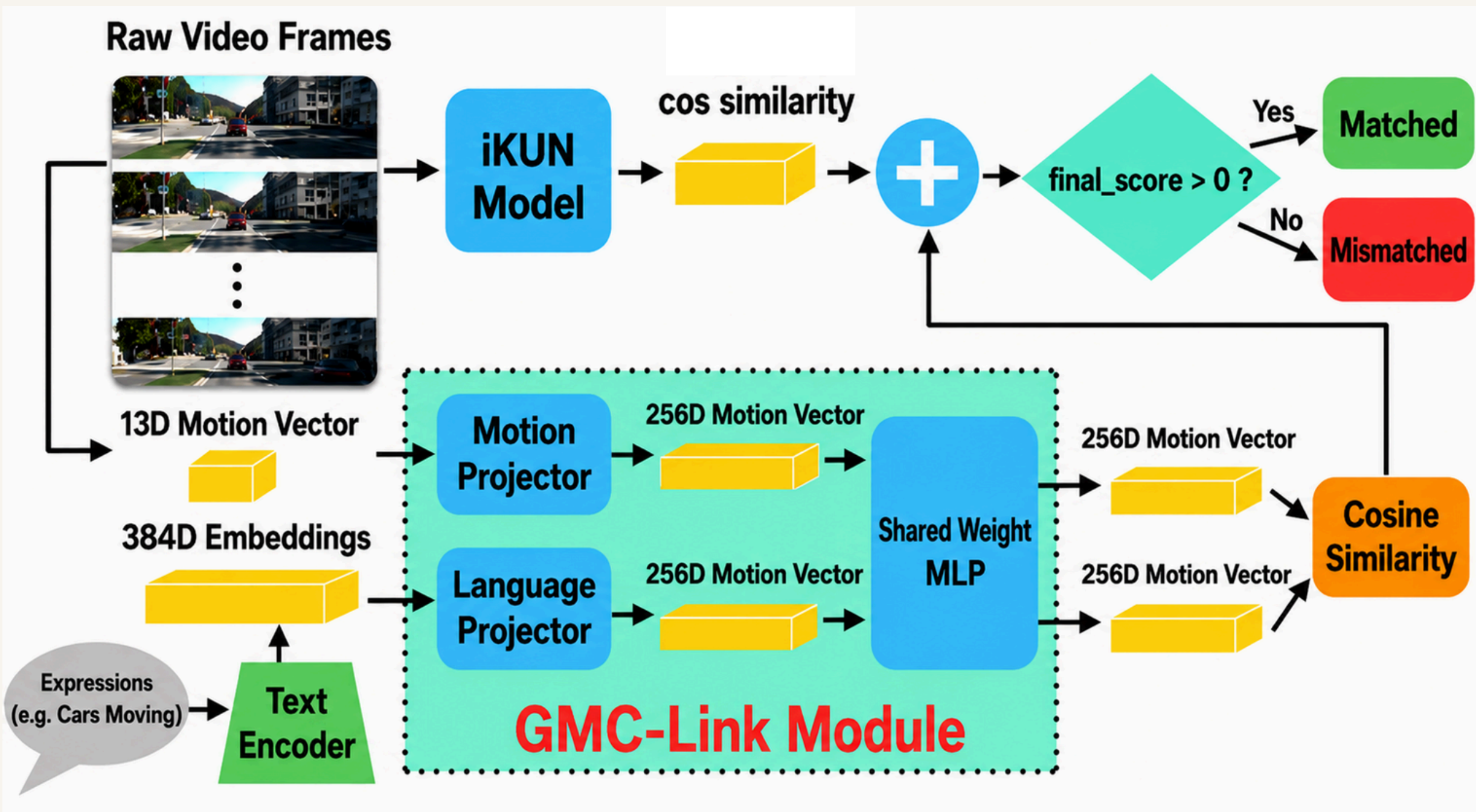
二、研究動機與目的

觀測位移 = 物體真實運動 - 相機自我運動 (*ego-motion*)



- 外觀導向方法對「運動」語意天生不敏感 (spatially-ignorant)
- 原始邊界框位移把相機運動與物體運動混為一談
- 研究目的：**補償相機運動**，正確判斷運動語意描述

三、系統架構與核心技術



（圖一）GMC-Link 架構 — 模組以決策層加成方式接入不更動原模型。

A · GMC 全域運動補償

ORB 特徵 → BFMatcher (Hamming + Lowe's 0.7) → RANSAC → Homography；前景遮罩排除物件。

B · 13D 多尺度殘餘運動

殘餘速度 = 物體速度 - 相機速度；輸出 13 維運動向量。

C · Align 運動-語言對齊

共享 MLP，以InfoNCE loss 驅動對齊運動-語言特徵
語言端 SentenceTransformer · all-MiniLM-L6-v2

D · Fuse 決策層融合

$final = \text{logit} + \alpha \cdot \text{sim_score}$ 運動／外觀兩軸分別校準。

五、實作技術與環境

語言與框架

Python

PyTorch

OpenCV

all-MiniLM-L6-v2

基礎架構

iKUN

FlexHook

Refer-KITTI

Refer-KITTI v2

資料集

六、創新與優勢

幾何補償降低相機晃動影響 — 利用 ORB + 單應性矩陣分離相機與物體運動
即插即用、免重訓 — 作為決策模組加入現有追蹤框架即可使用。

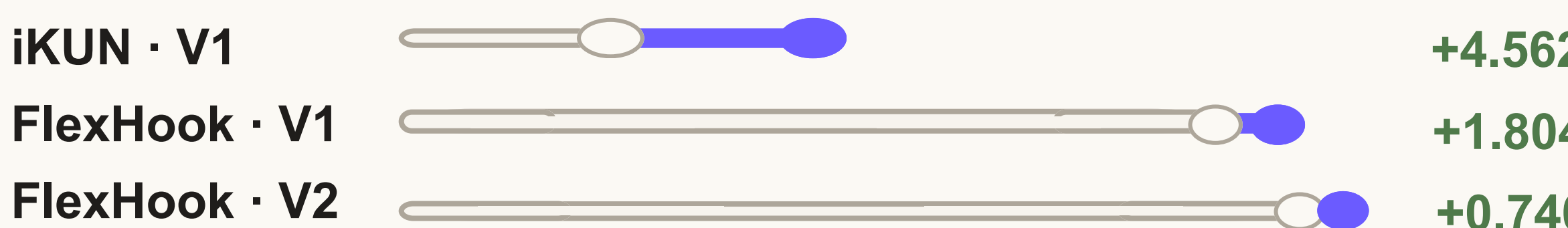
逐類別關聯性調節 — 運動／外觀兩軸分別校準，抑制純外觀描述上的運動雜訊。

七、成果展示

（表一）整體 *pooled HOTA* — 融合前後比較

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

（圖二）僅運動類別描述：原模型 → +GMC-Link



註：x軸為HOTA值範圍 22–52

運動類別描述 · 最大相對成長

+17.9%

iKUN · V1 (pooled HOTA 25.531 → 30.093)

三種架構在運動類別描述上全數提升。GMC-Link 在 原本最弱於運動語意的 iKUN 收益最大，正驗證「補償相機運動」的設計初衷。



（圖三）Query "moving cars" — baseline 全數漏追，+GMC-Link 正確召回移動中車輛（綠框）。

八、結論與未來展望

結論 · CONCLUSION

GMC-Link 補上既有 RMOT 對運動語意的盲點，**不需重訓** 即於三種架構皆獲增益，並且**超越原論文**，運動類別描述最高相對提升 **+17.9%**。

未來展望 · FUTURE WORK

運動估計與語言對齊合併為單一可學習系統，一起訓練。

加速 GMC 以邁向即時推論。

延伸至更多框架與資料集。



DEMO 影片
youtu.be/CRn4tVLbUg