

SAIoU: SORT ベースの物体追跡における ID Swap 抑制に向けた改良型 IoU 指標の提案

吉田龍人† 伊藤天詞† 藤井純一郎† 山下隆義‡

†八千代エンジニアリング株式会社 ‡中部大学

E-mail: ry-yoshida@yachiyo-eng.co.jp

1. はじめに

動画内の物体を時系列で特定する Multiple Object Tracking (MOT)は実空間をデータ化するための基盤技術の 1 つである. 中でも SORT[1]を始めとする Heuristic な Track by Detection のアプローチは簡潔な枠組みにより広く利用されている. Track by Detection では, 物体の位置を特定する Detection と, Bounding Box (Box)を対応付ける Association の工程で同一物体を識別する. Association のエラーは 2 つに大別される. 1 つは同一物体に異なる ID を割り振るエラー (以降, Split と称する)で, もう 1 つは異なる物体に同一 ID を割り振るエラー (以降, Swap と称する)である. Split・Swap は一般にトレードオフの関係を有する. 例えば同一である確信性が高いペアだけを対応付けた場合は Split が増加する. 反対に確信性が低いペアまで対応付けた場合は Split の減少に伴って Swap が増加する.

従来の MOT の性能評価において Split・Swap は等価なエラーとして評価される. しかし, 都市空間内における歩行者の経路選択など, 軌跡の一貫性が重要となる分析を MOT 解析結果に基づいて行う場合, Swap が致命的なエラーとなる. Swap は 1 つの軌跡情報に 2 つの物体の情報が混入するデータ汚染を引き起こすため, 解析データの信頼性を著しく損なう. これに対して Split が招く事態は情報の欠損に留まり, 分割した個々の断片は同一性が保持されるため部分的なデータの信頼性が担保される. さらに Split は後処理によって欠損箇所を結合できる余地があり, データの再利用性が高い利点もある. 以上の理由から個体の移動情報を正確に抽出・分析するためには Split よりも Swap を抑制することが望まれる.

本稿では SORT をはじめとする多くの Tracker で Association 指標となっている Intersection over Union (IoU)の改良に焦点を当てて, Swap 抑制を目指す. 本研究の貢献は以下の 2 点である.

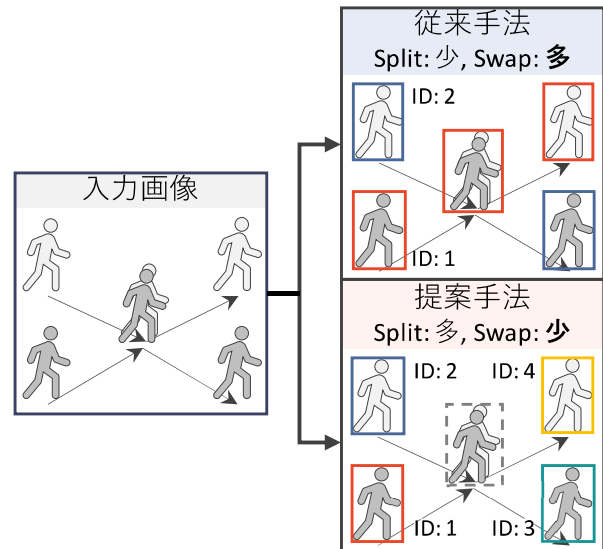


図-1 提案手法のコンセプト. 各 ID に所属する物体の同一性を担保するために, Split を許容して Swap を抑制することを目的とする.

- Swap 抑制に特化した新指標 Soft Assignment IoU (SAIoU)の提案: 重畳する Box の配置に基づいて Swap 発生リスクを Ambiguity として定量化し, これを動的なペナルティとして組み込む手法を提案した.
- Swap 抑制効果の性能実証: MOT17 を用いた実験により, 提案手法が既存指標と比較して AssPr を向上させ, Swap を抑制する効果を持つことを定量的に示した.

2. 関連研究

2.1. MOT

MOT は Track by Detection と End-to-End のアプローチに大別される[2]. 前者はアルゴリズムで同一物体を判別する Heuristic ベースの手法と学習済みモデルで判定するモデルベースの手法に区分される. Heuristic ベースの手法を代表する SORT は精度と計算量に優れ, 様々な派生手法が提案されている. 本研究は SORT に焦点を当てるものである.

2.2. SORT と IoU 評価指標

SORTはKalman Filter (KF)によって, Tracklet における真の Box の現在位置と将来位置を推定しつつ, 新たに得た Box と既往の Tracklet を Association する. Association の実施基準にはコストとして $1 - IoU$ を採用し, ハンガリアンアルゴリズムによって全割当コストの総和が最小化する組合せを決定する. 負の IoU を取るため, IoU の総和を最大化させることと同義である. よって基本的には IoU の高い Box ペアが優先して対応付けられる. IoU とは(1)式で示す 2 つの Box 間の Intersection を Union で除した値であり, 値域は $[0, 1]$ をとる.

$$IoU = \frac{Intersection}{Union} \quad (1)$$

IoU は Intersection の影響で, 2 つの Box が非重畳の場合に一律で 0 となる性質を持つ. そこで Simple Track[3]では, 非重畳の場合でも距離に基づいた評価を可能にするために Generalized IoU (GIoU)[4]を採用した. GIoU は(2)式に示す IoU に Box 領域の Convex Hull を用いたペナルティ項を導入したものである. ペナルティ項は Box 間のユークリッド距離に負の相関を持つ.

$$GIoU = IoU - \frac{ConvexHull - Union}{ConvexHull} \quad (2)$$

C-BIoU Tracker[5]では Box を一定の比率で拡張する Buffered IoU (BIoU)を提案し, Frame 間で安定的に重なりを生じさせる手法を示した. BIoU は top-left-w-h 形式の Box (x, y, w, h) が与えられたときに拡張率 b によって拡張した Box_b を(3)式で算出する.

$$Box_b = (x - bw, y - bh, w + 2bw, h + 2bh) \quad (3)$$

さらに BIoU の性能評価実験の比較対象には, (4)式に示す Distance IoU (DIoU)[6]が同時に採用された.

$$DIoU = IoU - \frac{\|c_i - c_j\|_2}{d^2} \quad (4)$$

ここで c は Box の中心座標, d は最小外接矩形の対角線の長さである. DIoU は GIoU と同様に IoU とペナルティ項で構成される. ペナルティ項によって非重畳の Box 間における距離に基づいた評価と, 包含関係にある Box 間の中心座標との距離に基づいた評価を両立する.

これまでに提案された IoU の改善手法は, 非重畳 Box 間でも評価値の差異を表現することに焦点を置いている. すなわち, これらは検出位置が乖離した場

合でも対応付けを成立させるための改良であり, Split を防止する効果が高いと解釈される. 一方で Swap に対してペナルティを与える効果を有する指標とはなっていない.

2.3. 評価指標

MOT では性能評価に Higher Order Tracking Accuracy (HOTA)[7]を利用する. HOTA は(5)式に示す Detection Accuracy (DetA) と Association Accuracy (AssA) の幾何平均である.

$$HOTA = \sqrt{DetA \cdot AssA} \quad (5)$$

ここで DetA は物体検出の総合的な精度を表す指標である. MOT ではこれと同時に検出漏れの少なさを表す Detection Recall (DetRe)や過検出の少なさを表す Detection Precision (DetPr)を評価する. 同一 Detector を使用した場合でも, Association 手法によって若干の変動が生じる点に留意が必要である.

AssA は Association の精度を総合的に評価する指標であり, 本研究で着目する Split・Swap の多寡が直接的に影響する. AssA は True Positive Association (TPA), False Negative Association (FNA), False Positive Association (FPA)を用いて(6)式で定義される.

$$AssA = \frac{|TPA|}{|TPA| + |FNA| + |FPA|} \quad (6)$$

なお TPA とは Box の位置と ID の双方で正しく対応した Ground Truth (GT)と Prediction ペアの数である. FNA は同一物体の対応が得られなかったペアの数, FPA は非同一物体との対応が生じたペアの数である. さらに AssA は(7)式に示す Association Recall (AssRe)と(8)式に示す Association Precision (AssPr)を用いた(9)式で表現できる.

$$AssRe = \frac{|TPA|}{|TPA| + |FNA|} \quad (7)$$

$$AssPr = \frac{|TPA|}{|TPA| + |FPA|} \quad (8)$$

$$AssA = \frac{AssRe \cdot AssPr}{AssRe + AssPr - AssRe \cdot AssPr} \quad (9)$$

よって, Split が発生した場合は, GT Box と対応する Box が減少するため, FNAが増加し, AssReが低下する. 反対に Swap ではこれらに加えて, GT Box と誤って対応する Box が Tracklet 内で増加するため, FPAが増加し, AssPrが低下する.

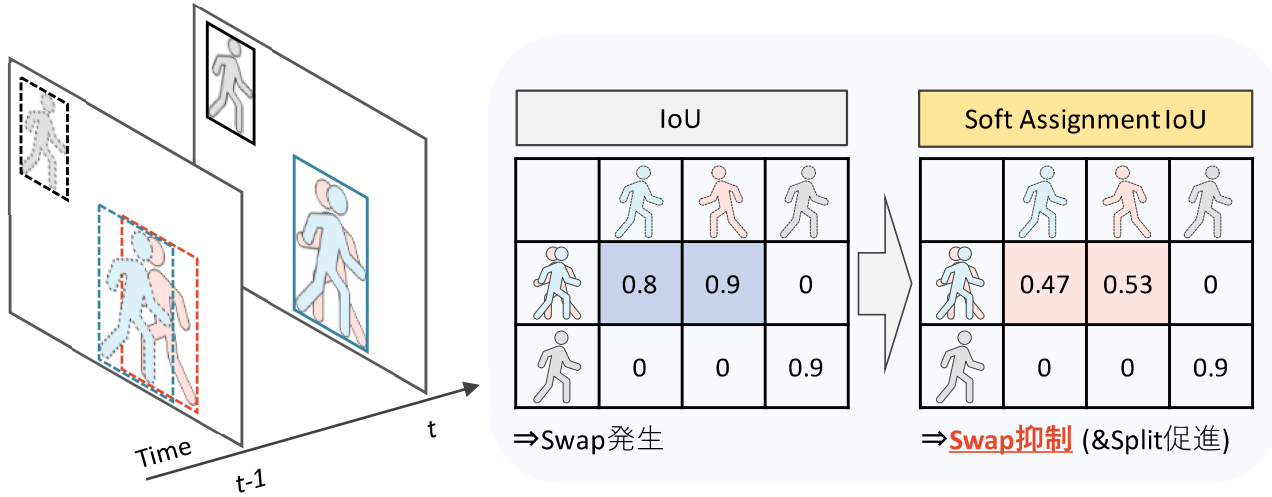


図-2 提案手法の概要図. 従来の IoU は Ambiguity の高い状況下において Swap が発生するリスクがある. 一方で Ambiguity に応じてペナルティを課す SAIoU は, Swap を抑制する効果を有する.

以上を踏まえて, 本研究では包括的な Tracking 性能を表す HOTA にも着目した上で, AssRe に比べて AssPr を向上させる手法を明らかにする.

3. 提案手法

3.1. 問題設定

SORT では IoU に基づくハンガリアンアルゴリズムによって同一物体を識別する. したがって Swap が発生する主たる要因は, Occlusion などにより複数の物体が近接する状況において, ネガティブペアがポジティブペアよりも高い IoU を有する状況が形成される点にある. IoU は Box 間の重なりを一对一で局所的に評価する指標である. さらに, ハンガリアンアルゴリズムでは IoU だけでペアを Association する. つまり高い IoU を示す Box が複数存在するという状況が Swap を誘引する.

Swap の対策には割当閾値を小さく設定して IoU の大きい Box のみを対応付ける方法が想定される. しかしこの処理は Hard Threshold として機能するため, 移動速度の大きい物体の追跡時などにおいて Split を著しく増加し, Tracking 精度そのものを低下させる副作用がある.

したがって, ハンガリアンアルゴリズムを使用する前提のもとで Swap を抑制しつつ Tracking 精度を維持するためには, 一律な閾値調整に依存するのではなく, Box の配置に応じて評価値を動的に変動させられる新たなアプローチが求められる.

3.2. Soft Assignment IoU

Swap 発生リスクの高い状況として, 特定の Detection や Tracklet に対して高い IoU を持つ割当候補が複数存在する状況が挙げられる. このような状況を表す指標を Ambiguity とし, Ambiguity を反映することのできる新たな IoU 指標である SAIoU を提案する. SAIoU の概要図を図-2 に示す.

Tracklet と Box の集合をそれぞれ $T = \{t_1, \dots, t_N\}$, $D = \{d_1, \dots, d_M\}$ とし, これらに基づく IoU 行列 $IoU \in [0, 1]^{N \times M}$ の (i, j) 成分を $IoU_{i,j}$ とする. このとき Tracklet t_i と Box d_j 間の SAIoU を(10)式で算出する.

$$SAIoU_{i,j} = w_{i,j} \cdot IoU_{i,j} \quad (10)$$

$$w_{i,j} = \frac{IoU_{i,j}}{A_{i,j} + \epsilon} \quad (11)$$

$$A_{i,j} = \left(\sum_{k=1}^N IoU_{k,j} \right) + \left(\sum_{l=1}^M IoU_{i,l} \right) - IoU_{i,j} \quad (12)$$

ここで ϵ はゼロ除算を防止する微小値である. 重み項 w は値域を $[0, 1]$ とする Ambiguity を逆数的に表現した信頼度係数として機能する. 割當時に競合する Box の量が少なくなるほど 1 に近似する. $A_{i,j}$ は Tracklet t_i と検知物体 d_j が他の候補と共有する IoU の総和を表し, 競合が少ない時ほど IoU に近似する.

従って IoU に Ambiguity に応じたペナルティ係数をかけ合わせる SAIoU は, 誤割当を発生させるリスクの高い組合せに対してのみペナルティを付与す

ることができるため、Swap 抑制に効果がある。ただし、本手法は Split を促進する方向に作用するリスクがある点には留意が必要である。

4. 実験

4.1. 実験方法

SAIoU の有効性を検証するため、IoU, GIoU, BIoU, DIoU との性能比較実験を行った。具体的にはハンガリアンアルゴリズムのコスト関数である $1 - \text{IoU}$ の IoU 項を各 IoU 指標に置換し、性能差を評価した。Tracker には SORT を使用し、BIoU の Box の拡張率は $b=15\%$ とした。

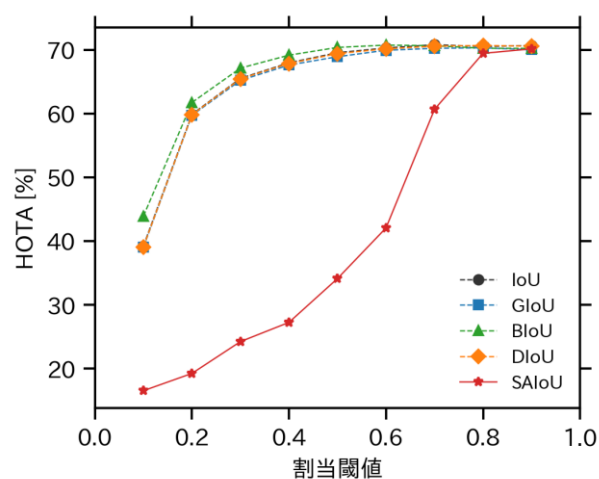
IoU 指標ごとに Association の最適な割当閾値は変動することが想定されるため、グリッドサーチで 0.1 から 0.9 まで 0.1 刻みで評価した。なおこの閾値は小さいときほど、IoU 指標の大きいペアだけを対応付ける条件となる。

データセットには GT が公開されている MOT17[8] の Train set を用いた。Detector には YOLOX[9] を用いた。YOLOX のモデルには Byte Track[10] の公式 GitHub にて公開されている、MOT17 用学習済みモデルを使用した。

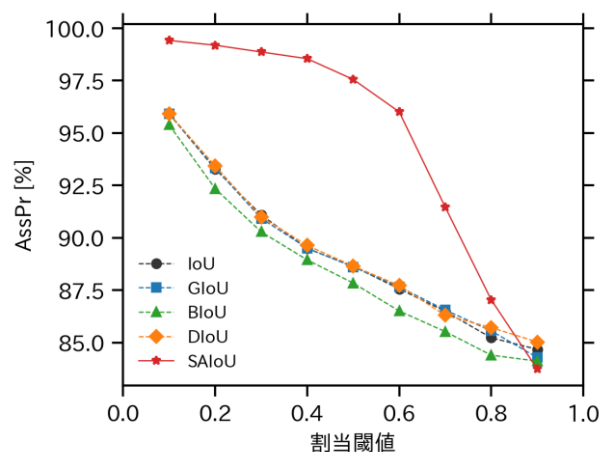
4.2. 実験結果

各 IoU 指標について、横軸に各割当閾値、縦軸に HOTA, AssPr, AssRe を示したグラフを図-3 に示す。既存手法は概ね類似した挙動を示した。閾値の上昇に伴って HOTA が向上し、0.5 以降で飽和する傾向が見られた。対して SAIoU はこれらと明確に異なる挙動を示した。具体的には 0.8 以下の閾値領域において既存手法よりも AssRe が低く推移する一方で、AssPr は顕著に高い値を維持した。SAIoU は全般的に優れた AssPr を示すが、割当閾値 0.9 の条件下では例外的に他の手法を下回る結果となった。

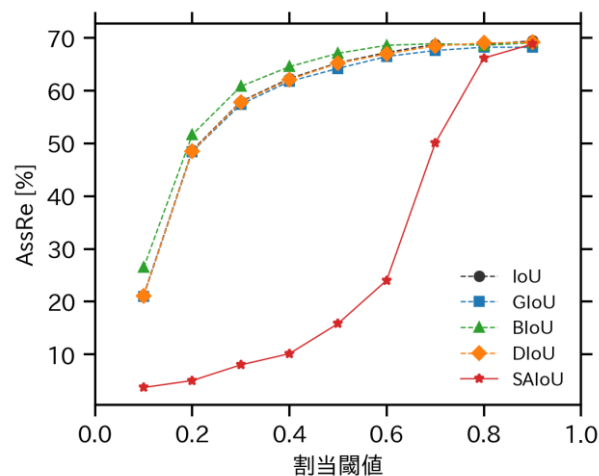
表-1 に、各指標において HOTA が最大となったケースの評価値を抜粋する。いずれの実験も割当閾値が 0.7~0.9 という大きい閾値で最大値を示した。HOTA は IoU を用いた場合に最大値の 70.79, SAIoU を用いた場合に最小値の 70.14 となり、差は僅少であった。実験では同一の Detector を使用したため、DetA, DetRe, DetPr といった Detection 関連の性能はほぼ同等の値を示した。



(1) HOTA



(2) AssPr



(3) AssRe

図-3 割当閾値と HOTA/AssPr/AssRe

表-1 各 IoU 指標の HOTA 最大時の性能評価結果

metric	割当閾値	HOTA	DetA	DetRe	DetPr	AssA	AssRe	AssPr
IoU	0.7	70.79	77.22	83.15	86.39	65.37	68.82	86.49
GIoU	0.8	70.30	76.99	83.25	85.99	64.67	68.26	85.55
DIoU	0.8	70.67	77.04	83.24	86.05	65.31	69.05	85.71
BloU	0.7	70.63	76.95	83.33	85.86	65.28	68.93	85.52
SAIoU	0.9	70.14	76.85	83.45	85.58	64.49	68.86	83.74

5. 考察

5.1. 従来 IoU 指標の性能比較

表-1 より, SAIoU を含む各 IoU 指標に対する最大 HOTA の差は 0.65 以内に留まり, 最大 HOTA に対する有意な差は確認されなかった.

既存 IoU 指標の中でも IoU, GIoU, DIoU は特に性能差が乏しく, 各割当閾値に対する変化もほぼ同等であった. GIoU は Box 間距離に応じたペナルティを, DIoU はそれに加えて包含関係にある Box 間にもペナルティを付与する機構を有する. しかし本実験のような高 FPS 動画を対象とした SORT Tracker の解析では, これらのペナルティが寄与する場面が限定的であったと推察される. フレーム間の Box の移動量が小さく, Positive ペアの多数が重畳する状況下では, 非重畳や包含関係に基づくペナルティの影響以上に, 基本的な重なりの有無が支配的であった可能性が高い.

一方で BloU は低い割当閾値において高い性能を示す傾向が見られた. この結果は Box を拡張することで実質的な IoU が増加し, 擬似的に高い閾値で解析した場合と同等の効果が早期に得られたためと推察される. これはピーク時の性能においては他指標との有意差は確認されなかった点からも裏付けられる. Box を拡張した場合でも候補間の相対的な割当順位自体は基本的に変化しないため, フレーム間で重畳状態が担保されやすい MOT17 データセットでの BloU の効果は限定的だったと示唆される. ただし, フレーム間で物体が大きく移動し, 重畳領域が消失しやすい低 FPS 映像においては, BloU の拡張処理が有効に機能する可能性がある.

以上の結果から, SORT アルゴリズムを用いた高 FPS 映像の追跡タスクにおいて, 通常の IoU が既存指標の中で最も合理的な IoU 指標であると結論付け

る. IoU はペナルティ項や Box の拡張に関する計算が不要であるため計算効率に優れる.

一方で FPS や対象物の移動速度に変化が生じた場合はこの限りではなく, Box を疑似的に重畳させられる BloU などにも一定の効果が期待される.

5.2. SAIoU の特性

図-3 より SAIoU は低い割当閾値適用時に AssPr において他の手法よりも顕著に高い値を示した. これは(11)式で定義する Ambiguity に基づく重み項 w が, 効果的に作用したことを示唆している. 問題設定で提起した特定の Box・Tracklet に割当候補が多数存在する状況に対して, ペナルティを付与する処理が Swap を抑制する結果をもたらした. 特に割当閾値を 0.1 に設定した際には 99.41 という極めて高い AssPr を示した.

SAIoU におけるペナルティ付与の特性を踏まえると, 低い割当閾値を設定した条件の下で発生した FPA は, 1 対 1 の対応関係を持つネガティブ Box ペアによるものである可能性が極めて高い. 一方で IoU ベースの同一物体識別だけで高い性能を維持しつつ, これらの FPA の発生を抑制することは現実的に困難である. よって AssPr を 100% に近付けるためには, DeepSORT[11] で採用された Re-Identification (Re-ID) 特徴量など, 異なる尺度を用いた同一物体識別が有効であると推察される.

提案手法の唯一の問題点として, 割当閾値を 0.9 まで高めた際に AssA および AssPr が低下する現象が発生したことが挙げられる. これは(12)式に示したペナルティに起因していると推察される. 本手法では多くの Box と重なり領域を有する Box ほどペナルティが増大する. 例えば横に並んだ 3 つの Box と重畳した場合, 中心の Box との SAIoU は両端の Box に比べて極めて小さくなるリスクがある. このような影響によってネガティブペアが誤って対応付けられた

可能性がある。ただし 0.9 という閾値設定は本手法が目的とする Swap 抑制という方針に反するため、解析上の問題はほとんどないと想定する。

以上より、SAIoU は Split を許容して Swap を抑制するという本研究の目的に適う手法であることが確認された。追跡結果の同一性が高く担保されるため、歩行者の経路選択といった個体の移動分析タスクにおいて、より信頼性の高いデータが提供できる。SAIoU は従来の IoU 指標に比べて同一である確信度の高いものを対応付けるため、ByteTrack に代表される多段階割当をする Tracker の初段割当処理に適用することで、HOTA を維持しつつ Swap を劇的に低減できる可能性がある。

6. おわりに

6.1. まとめ

本研究では MOT における Tracklet の Swap 低減を目的に、重畳する Box 配置に基づく Ambiguity に着目した改良型 IoU 評価指標の SAIoU を提案した。SORT アルゴリズムをベースに MOT17 データセットで既存の IoU 評価指標との比較実験を行った。Swap と相関を持つ Association 評価指標である AssPr において、提案手法が高い性能を示すことを明らかにした。

6.2. 今後の課題

提案手法は AssPr の向上には極めて有効であるが、AssRe が低いため、HOTA 自体は抑制される傾向にある。今後は、既存手法との統合による HOTA の向上が課題である。

DeepSORT に代表される Re-ID 特徴量を用いたコスト関数や、ByteTrack に代表される多段階割当処理などを導入することで、より高い HOTA が実現すると推察される。さらに GIAOTracker[12]で提案された Global Link に代表される、Split した Tracklet を後処理で再結合させるような処理とも高い相性を持つと期待される。これらの技術を併用することで、より実用的なデータ取得システムの構築を目指す。

参考文献

[1] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, ‘Simple Online and Realtime Tracking’, *Int. Conf. Image Process.*, pp.

3464–3468, Sep. 2016.
[2] M. Adžemović, ‘Deep Learning-Based Multi-Object Tracking: A Comprehensive Survey from Foundations to State-of-the-Art’, Jun. 2025, arXiv:2506.13457.
[3] J. Li, Y. Ding, and H. Wei, ‘SimpleTrack: Rethinking and Improving the JDE Approach for Multi-Object Tracking’, Mar. 2022, arXiv:2203.03985.
[4] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, ‘Generalized Intersection over Union: A Metric and A Loss for Bounding Box Regression’, Apr. 2019, arXiv:1902.09630.
[5] F. Yang, S. Odashima, S. Masui, and S. Jiang, ‘Hard to Track Objects with Irregular Motions and Similar Appearances? Make It Easier by Buffering the Matching Space’, Nov. 2023, arXiv:2211.14317.
[6] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, and D. Ren, ‘Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression’, Nov. 2019, arXiv:1911.08287.
[7] J. Luiten *et al.*, ‘HOTA: A Higher Order Metric for Evaluating Multi-Object Tracking’, *Int. J. Comput. Vis.*, vol. 129, no. 2, pp. 548–578, Feb. 2021.
[8] P. Dendorfer *et al.*, ‘MOTChallenge: A Benchmark for Single-Camera Multiple Target Tracking’, Dec. 2020, arXiv:2010.07548.
[9] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, ‘YOLOX: Exceeding YOLO Series in 2021’, Aug. 2021, arXiv:2107.08430.
[10] Y. Zhang *et al.*, ‘ByteTrack: Multi-Object Tracking by Associating Every Detection Box’, Apr. 2022, arXiv:2110.06864.
[11] N. Wojke, A. Bewley, and D. Paulus, ‘Simple Online and Realtime Tracking with a Deep Association Metric’, Mar. 2017, arXiv:1703.07402.
[12] Y. Du, J. Wan, Y. Zhao, B. Zhang, Z. Tong,

and J. Dong, ‘GIAOTracker: A comprehensive framework for MCMOT with global information and optimizing strategies in VisDrone 2021’, Feb. 2022, arXiv:2202.11983.