

Occlusion-Aware SORT: Observing Occlusion for Robust Multi-Object Tracking

Supplementary Material

7. Reliability Analysis and Pseudo code

7.1. Occlusion-Aware Module

7.1.1. Reliability Analysis

Occlusion in 2D image space follows two primary spatial cues:

- **Depth ordering:** In pedestrian or vehicle scenes, 2D overlap typically indicates physical proximity, and bounding boxes with significantly lower bottom-edge y-coordinates (*i.e.*, larger b , Fig. 3) are typically nearer to the camera and thus more likely to occlude others.
- **Intersection area:** For objects with intersecting bounding boxes (non-zero IoU), the overlapping region correlates with the extent of physical occlusion.

OAM leverages both cues and constructs a logical occlusion relationship matrix:

$$C_{IoU}(\mathcal{D}_i, \mathcal{D}_j) > 0 \wedge b_j - b_i > thre_{occ} \quad (12)$$

where $C_{IoU}(\mathcal{D}_i, \mathcal{D}_j)$ ensures that only **geometrically consistent occlusion hypotheses** are considered; $thre_{occ}$ aims to reduce **false depth ordering brought by fluctuations and strong competition at the bottom edge**.

In terms of GM (Eq. (5)), for each bounding box i , σ_i^x and σ_i^y are defined by:

$$\sigma_i^x = \frac{w_i}{k_x}, \quad \sigma_i^y = \frac{h_i}{k_y}. \quad (13)$$

The GM aims to reflect the relative importance of pixels affected by occlusion. All in all, OAM applies a soft probabilistic prior that reflects the confidence that multiple trajectories should be associated. This design complements hard IoU logic from the trajectory view, emphasizing the association of unoccluded trajectories.

7.1.2. Pseudocode

$\mathcal{D} = [l, t, r, b]$ indicates detected bounding boxes, where l, t, r, b denote the left, top, right, and bottom of the bounding box. Herein, we assume that N bounding boxes ($\mathcal{D} \in \mathbb{R}^{N \times 4}$), need to be associated with high-score detections, where the width and height of the image are W and H . The Pseudo-code about OAM is shown as follows, where \leftarrow indicates assignment. Algorithm 1 implements three key components:

- Lines 5-9: Eq. (5);
- Lines 10-35: Eq. (2) and Eq. (4);
- To improve computational efficiency, lines 2-3 and line 6 are used for data filtering.

Moreover, it is worth noting that to avoid fluctuations in estimation or detection, we set a threshold ($thre_{occ}$) for comparing the bottom edges (line 12 of the algorithm). When the difference value exceeds $thre_{occ}$, it will be considered an occlusion. The $thre_{occ}$ for the experiment in our work is 5. This can also alleviate the corresponding issues caused by incorrect bottom-edge depth ordering. Moreover, Gaussian scaling factors k_x and k_y are used to adapt to the proportion of background pixels within each detected bounding box. Their values are discussed in the Sec. 8.1.

7.2. Occlusion-Aware Offset

7.2.1. Reliability Analysis

Occlusion-Aware Offset (OAO) addresses the limitation of traditional IoU-based association in occluded scenarios, and this situation is analyzed in Sec. 3. Under occlusion, the spatial relationship between detections and tracks becomes unreliable due to:

- **Unreliable prediction of lost trajectory:** Multiple trajectories competing for the same detection region due to overlapping occlusions, as shown in Fig. 8 (in Sec. 8.5).
- **Positional ambiguity:** Occluded objects experience centroid shifts, bounding box distortions or strong competition, making IoU an inconsistent measure of true spatial proximity. This situation leads to cost confusion between detected bounding box and trajectories, causing ID switch as shown in Fig. 9 (in Sec. 8.5).

OAO mitigates these issues by incorporating occlusion awareness into the association cost matrix from the view of trajectory, effectively enhancing the discriminative power of the cost.

7.2.2. Pseudocode

We assume that the N detections $\mathcal{D} \in \mathbb{R}^{N \times 4}$ need to be associated with M estimations \mathbf{X} . The pseudo-code for Occlusion-Aware Offset (OAO) is shown Algorithm 2. The spatial consistency score about detections and estimations will be updated. Notably, the initial calculations of the spatial consistency score are different in different trackers. For example, Hybrid-SORT [58] uses $HMIoU$; PD-SORT [52] uses $DVIOU$ that is a 3D IoU. Herein, C_{IoU} and IoU are used uniformly.

7.3. Bias-Aware Momentum

7.3.1. Reliability Analysis

In theory, occlusion-aware tracking requires dynamic adjustment of both the process noise \mathbf{Q} and measurement noise \mathbf{R} . However, real-time estimation of these parameters presents significant challenges:

Algorithm 1 Occlusion-Aware Module ($\hat{O}c \leftarrow OAM(\cdot)$)

Input: Detected bounding boxes $\mathcal{D} \in \mathbb{R}^{N \times 4}$
Parameter: Image dimensions (W, H), Gaussian scaling factors k_x, k_y , occlusion trigger threshold $thre_{occ}$
Output: Refined occlusion coefficients $\hat{O}c \in \mathbb{R}^N$

- 1: Initialize $\hat{O}c \leftarrow \mathbf{0}^N$ {Initialize occlusion coefficients}
- 2: $IoU_{\mathcal{D}} \in \mathbb{R}^{N \times N} \leftarrow C_{IoU}(\mathcal{D}, \mathcal{D})$ {Compute IoU matrix}
- 3: $\text{diag}(IoU_{\mathcal{D}}) \leftarrow 0$ {Ignore self-relation}
- # **Spatial consistency determination**
- 4: **if** $\max(IoU_{\mathcal{D}}) > 0$ **then**
- 5: $GM \in \mathbb{R}^{H \times W} \leftarrow \mathbf{0}^{H \times W}$ {Initialize Gauss-Map}
- 6: **for each** \mathcal{D}_i in $\{\mathcal{D}_i | \max(IoU_{\mathcal{D}}[i]) > 0\}$ **do**
- 7: $\sigma_i^x \leftarrow w_i/k_x, \sigma_i^y \leftarrow h_i/k_y$ {Adaptive sigmas}
- 8: $GM[t_i : b_i, l_i : r_i] \leftarrow \max(GM[t_i : b_i, l_i : r_i], \text{drawGaussian}(\mathcal{D}_i, \sigma_i^x, \sigma_i^y))$
- 9: **end for**
- 10: $\text{bottoms} \leftarrow \mathcal{D}[:, 3]$ {Bottom y-coordinates (b)}
- 11: $\text{areas} \leftarrow (\mathcal{D}[:, 2] - \mathcal{D}[:, 0]) \times (\mathcal{D}[:, 3] - \mathcal{D}[:, 1])$ {Calculate Bounding-box areas}
- 12: $\text{validMask} \leftarrow (\text{bottoms}[:, \text{None}] - \text{bottoms}[\text{None}, :]) \leq -thre_{occ} \wedge (IoU_{\mathcal{D}} > 0)$ {The occlusion relationship matrix.}
- 13: $\text{all}_L, \text{all}_T, \text{all}_R, \text{all}_B \leftarrow \mathcal{D}$ {Vectorized coordinates}
- # **Obtaining occlusion coefficient**
- 14: **for each** \mathcal{D}_i and i in \mathcal{D} **do**
- 15: $js \leftarrow \{j | \text{validMask}[i, j] = \text{TRUE}\}$ {Select objects that occludes object i .}
- 16: **if** $js = \emptyset$ **then**
- 17: **continue**
- 18: **end if**
- 19: $l_i, t_i, r_i, b_i \leftarrow \mathcal{D}_i$
- 20: $\text{localGM} \in \mathbb{R}^{h_i \times w_i} \leftarrow GM[t_i : b_i, l_i : r_i]$ {Crop local GM from GM : Gaussian heatmap about \mathcal{D}_i }
- 21: $T \in \mathbb{R}^{\text{len}(js)} \leftarrow \max(t_i, \text{all}_T[js])$ { $\text{len}(js)$ is array length of js }
- 22: $B \in \mathbb{R}^{\text{len}(js)} \leftarrow \min(b_i, \text{all}_B[js])$
- 23: $L \in \mathbb{R}^{\text{len}(js)} \leftarrow \max(l_i, \text{all}_L[js])$
- 24: $R \in \mathbb{R}^{\text{len}(js)} \leftarrow \min(r_i, \text{all}_R[js])$
- 25: $tClip \in \mathbb{R}^{\text{len}(js)} \leftarrow \max(0, T - t_i)$ {Coordinate transformation}
- 26: $bClip \in \mathbb{R}^{\text{len}(js)} \leftarrow \min(h_i, B - b_i)$
- 27: $lClip \in \mathbb{R}^{\text{len}(js)} \leftarrow \max(0, L - l_i)$
- 28: $rClip \in \mathbb{R}^{\text{len}(js)} \leftarrow \min(w_i, R - r_i)$
- 29: $\text{occlusionMap} \leftarrow \mathbf{0}^{h_i \times w_i}$ {Boolean occlusion map: occlusion region}
- 30: **for each** j in js **do**
- 31: $\text{occlusionMap}[tClip[j] : bClip[j], lClip[j] : rClip[j]] \leftarrow 1$ {Obtain the overlapping region}
- 32: **end for**
- 33: $\hat{O}c_i \leftarrow \frac{\sum(\text{localGM} \odot \text{occlusionMap})}{\text{areas}_i}$ { \odot indicates Hadamard (element-wise) product.}
- 34: **end for**
- 35: **end if**
- 36: **return** $\hat{O}c$

Algorithm 2 Occlusion-Aware Offset

$S \leftarrow OAO(\mathbf{X}, IoU_{\mathcal{D}, \mathbf{X}})$

Input: estimations $\mathbf{X} \in \mathbb{R}^{M \times 4}$ from tracked trajectories; $IoU_{\mathcal{D}, \mathbf{X}} \in \mathbb{R}^{N \times M}$ between \mathcal{D} and \mathbf{X}

Parameter: Coefficient τ that is used to balance IoU and $\hat{O}c$

Output: Spatial consistency S

- 1: $\hat{O}c^{\mathbf{X}} \in \mathbb{R}^{1 \times M} \leftarrow OAM(\mathbf{X})$
- 2: $S \in \mathbb{R}^{N \times M} \leftarrow \tau \cdot (1 - \hat{O}c^{\mathbf{X}})[\text{None}, :] + (1 - \tau) \cdot IoU_{\mathcal{D}, \mathbf{X}}$ {This describes Eq. (8).}
- 3: **return** S

Refine K	Dataset	MOTA	HOTA	AssA	IDF1
✓	DanceTrack	89.73	61.49	48.03	63.65
		89.75	60.81	46.96	62.74
✓	MOT17	76.29	67.06	68.74	78.19
		76.11	66.98	68.67	77.74

Table 7. The influence of refining **K** through BAM (OA-SORT).

- **Measurement noise estimation:** The error variance of the detection model under varying occlusion intensities is difficult to characterize. Occlusion-induced errors—including feature confusion and information loss—are varied and difficult to model accurately.
- **Process noise estimation:** Motion model errors stem from natural scene dynamics and model residuals, which are inherently unobservable without ground truth. Furthermore, MOT in 2D images involves discrete observations of inherently nonlinear and irregular target motions. Any predictor necessarily exhibits lag during motion transitions, representing an unavoidable inherent error.

Herein, the additional experiments for BAM are conducted, as shown in Tab. 7, for Kalman gain (**K**) that can reflect and consider **Q** and **R**.

The result shows that BAM implements an instantaneous and implicit measurement reliability adjustment for low-confidence detections by directly modifying observation. This approach attempts to regulate the influence of new observations during the update step, eliminating the need for explicit error modeling or covariance recomputation—aligning well with both practical constraints and robustness objectives in real-world MOT. According to Eq. (10) and Eq. (11), the state update in Eq. (11) can be reformulated as:

$$\mathbf{X}_{t|t} = \mathbf{X}_{t|t-1} + \text{BAM} \cdot \mathbf{K}_t(\mathbf{Z}_t - \mathbf{H}_t \mathbf{X}_{t|t-1}). \quad (14)$$

7.3.2. Pseudocode

We assume that a trajectory is association with a low-score detection, $d \in \mathbb{R}^4$; the trajectory’s estimation $\mathbf{x}_{t|t-1} \in \mathbb{R}^4$.

Algorithm 3 Bias-Aware Momentum ($\mathbf{x}_{t|t} = \text{BAM}(d, \mathbf{x}_{t|t-1}, \hat{O}^{\mathbf{z}})$)

Input: det that is detected bounding box; $\mathbf{x}_{t|t-1}$; $\hat{O}^{\mathbf{z}}$
Output: Optimized $\mathbf{x}_{t|t}$

- 1: $IoU_{det, \mathbf{x}_{t|t-1}} \in (0, 1] \leftarrow C_{IoU}(det, \mathbf{x}_{t|t-1})$
- 2: $BAM \leftarrow IoU_{det, \mathbf{x}_{t|t-1}} \cdot (1 - \hat{O}^{\mathbf{z}})$ {This describes Eq. (9).}
- 3: $z' \leftarrow det$ { z' denotes observations.}
- 4: $\mathbf{x}_{t|t} \leftarrow \mathbf{x}_{t|t-1} + BAM \cdot \mathbf{K}_t(z' - \mathbf{H}_t \mathbf{x}_{t|t-1})$ {We omit unrelated Kalman filtering details for brevity. This describes Eq. (10) and Eq. (11).}
- 5: **return** $\mathbf{x}_{t|t}$

Algorithm 4 Association approach - high-score detections association

{This step can be mixed with the low-score association, such as TrackTrack [45]}

Input: $\mathcal{D}^{high}, \mathcal{T}$

- 1: $\mathcal{E} \leftarrow [t.estimation \text{ in } t \text{ for } \mathcal{T}]$ { \mathcal{E} indicate the estimation of trajectories.}
- 2: $IoU_{\mathcal{D}^{high}, \mathcal{E}} \leftarrow C_{IoU}(\mathcal{D}^{high}, \mathcal{E})$
- 3: $S \leftarrow OAO(\mathcal{E}, IoU_{\mathcal{D}^{high}, \mathcal{E}})$ {This describes Sec. 4.2.}
- 4: $S \leftarrow S + Other_{score}$ {Herein, other scores ($Other_{score}$) are used for allocation besides IoU. For example, Hybrid-SORT [58] and ByteTrack [62] use direction score and detection score. If using other scores, please calculate based on S . That's right here. Notably, this step can be mixed with the third line.}
- 5: $high_Match_{det}, high_Match_{tra} \leftarrow$ the Hungarian Algorithm uses S to allocation

OAM is used on the process of updating combining with the occlusion coefficient of the latest observation $\hat{O}^{\mathbf{z}}$ of the trajectory. The pseudocode is shown in Algorithm 3.

7.4. Association approach

Given detections \mathcal{D} , trajectories \mathcal{T} , the Pseudo-codes for high-score detections \mathcal{D}^{high} , low-score detections \mathcal{D}^{low} associations, and trajectory updating are shown as Algorithm 4 and Algorithm 5, where $\mathcal{D}^{high} \cup \mathcal{D}^{low} = \mathcal{D}$ and $\mathcal{D}^{high} \cap \mathcal{D}^{low} = \emptyset$.

8. Additional Experiments

8.1. σ^x and σ^y in GM

For object i , σ_i^x and σ_i^y in GM can be indicated as $\frac{w_i}{k_x}$ and $\frac{h_i}{k_y}$. The performance results for different k_x and k_y values are presented as shown in Tab. 8, Tab. 9 and Tab. 10, where the rows represent the same k_x value and the columns represent the same k_y value. As a result, the $(k_x, k_y) = (5, 3)$

Algorithm 5 Association approach - low-score detections association and trajectory updating

Input: $\mathcal{D}^{low}, \mathcal{T}^{un} \in \mathcal{T}$

- 1: $\mathcal{E}^{un} \leftarrow [t.estimation \text{ in } t \text{ for } \mathcal{T}^{un}]$ { \mathcal{E}^{un} indicate the estimation of unmatched trajectories.}
- 2: $IoU_{\mathcal{D}^{low}, \mathcal{E}^{un}} \leftarrow C_{IoU}(\mathcal{D}^{low}, \mathcal{E}^{un})$
- 3: $S \leftarrow IoU_{\mathcal{D}^{low}, \mathcal{E}^{un}} + Other_{score}$ {Herein, other scores ($Other_{score}$) used for allocation besides IoU. For example, Hybrid-SORT [58] and ByteTrack [62] use direction score and detection score. If using other scores, please calculate based on S . That's right here.}
- 4: $Low_Match_{det}, Low_Match_{tra} \leftarrow$ the Hungarian Algorithm uses S to allocation {Store index in \mathcal{D} and \mathcal{T} of matched detections and trajectories.}
- 5: $Match_{det} \leftarrow high_Match_{det} + Low_Match_{det}, Match_{tra} \leftarrow high_Match_{tra} + Low_Match_{tra}$ {If other matches exist, they are merged accordingly.}
- # **-----Trajectory updating-----**
- 6: **for** each d, t in $Match_{det}, Match_{tra}$ **do**
- 7: **if** $\mathcal{D}[d]$ is in \mathcal{D}^{low} **then**
- 8: $offset \leftarrow BAM(\mathcal{D}[d], \mathcal{E}[t], \mathcal{T}[t].\hat{O}^{\mathbf{z}})$ { BAM is for \mathcal{D}^{low} . This describes Eq. (10).}
- 9: **else**
- 10: $offset \leftarrow 1$
- 11: **end if**
- 12: $\mathcal{T}[t] \leftarrow$ update $\mathcal{T}[t]$ using $offset$ and $\mathcal{D}[d]$
- 13: **end for**
- 14: $\mathbf{Z} \leftarrow [t.last_observation \text{ for } t \text{ in } \mathcal{T}[Match_{tra}]]$
- 15: $\hat{O}^{\mathbf{z}} \leftarrow OAM(\mathbf{Z})$
- 16: $[\mathcal{T}[t].\hat{O}^{\mathbf{z}} \leftarrow \hat{O}^{\mathbf{z}}[t] \text{ for } \hat{O}^{\mathbf{z}}[t] \text{ in } \hat{O}^{\mathbf{z}}]$

	2	3	4	5	6
2	60.357	60.477	60.366	60.49	60.836
3	60.543	60.699	60.773	61.076	60.979
4	60.682	61.295	61.117	61.003	60.899
5	61.451	61.667	60.965	60.819	60.607
6	61.415	61.035	60.754	60.249	60.391

Table 8. HOTA (%) on the DanceTrack validation set under different (k_x, k_y) .

can achieve the best performance, i.e., $\sigma_i^x = \frac{w_i}{5}$ and $\sigma_i^y = \frac{h_i}{3}$.

The result primarily shows the approximate global optimum on the DanceTrack validation set. However, when k_x or k_y is larger, the Gaussian distribution might be prone to disorder as shown in Fig. 5. To avoid particularity of k_x and k_y , we attempted to reduce $k_x : k_y = 3 * \sqrt{2} : 3 = 1.414 : 1$, because the quadratic term $(\sigma_n^x)^2$ in Eq. (5) as well as $k_x \in (4, 5)$. The results in Tab. 11 agree with us that the

	2	3	4	5	6
2	46.171	46.326	46.268	46.351	47.023
3	46.494	46.794	46.902	47.374	47.363
4	46.732	47.731	47.418	47.318	47.167
5	47.932	48.257	47.211	47.074	46.719
6	47.926	47.379	47.022	46.207	46.472

Table 9. AssA (%) in DanceTrack under different k_x, k_y .

	2	3	4	5	6
2	61.712	61.848	62.176	62.249	62.924
3	62.031	62.312	62.83	63.348	63.037
4	62.198	63.373	63.182	63.069	62.818
5	63.444	63.789	63.074	62.778	62.507
6	63.38	62.943	62.682	61.938	62.292

Table 10. IDF1 (%) in DanceTrack under different k_x, k_y .

Dataset	(k_x, k_y)	HOTA	AssA	IDF1
test	-	62.20	47.40	63.00
	$(3 * \sqrt{2}, 3)$	63.13	48.53	64.17
	$(5, 3)$	62.93	48.23	64.23
val	-	59.39	44.90	60.67
	$(3 * \sqrt{2}, 3)$	61.49	48.03	63.65
	$(5, 3)$	61.67	48.26	63.79
Average	-	60.80	46.15	61.84
	$(3 * \sqrt{2}, 3)$	62.31	48.28	63.91
	$(5, 3)$	62.30	48.25	64.01

Table 11. The performance of OA-SORT under different k_x in DanceTrack. '-' indicates baseline Hybrid-SORT.

performance is better in the DanceTrack test dataset when $k_x = (3 * \sqrt{2}, 3)$. Finally, σ_n^x and σ_n^y are set to $\frac{w_n}{3 * \sqrt{2}}$ and $\frac{y_n}{3}$, respectively.

8.2. Ablation Study for MOT17

The additional ablations for MOT17 are executed as shown in Tab. 12. The results exhibit the performance improvement from different detection, including YOLOX and Public detection (FRCNN) provided by Hybrid-SORT and host, respectively. Overall, in Public detection with lower detection quality, the occlusion-aware framework can still ensure the performance. However, the benefits it brings are not as good as those from high-quality detection. This is due to missed detections and false positives as shown in Fig. 6, which are not the focus of the occlusion-aware framework.

Finally, the camera-motion compensation (CMC [37]) is integrated into OA-SORT under the detection provided by Hybrid-SORT. The results are reported in Tab. 13. Comparing enhancement brought by occlusion-awareness with

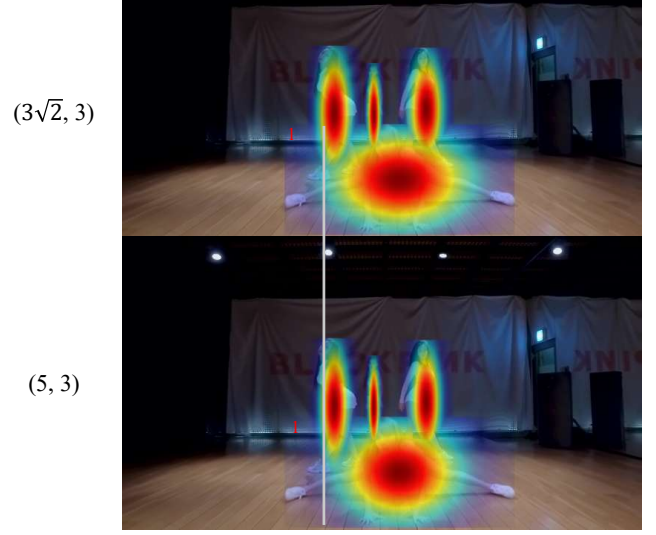
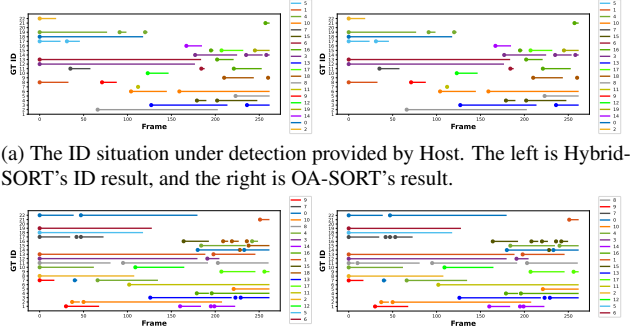


Figure 5. The example about different (k_x, k_y) .

Detection	Method	MOTA	HOTA	AssA	IDF1
YOLOX	Hybrid-SORT	75.65	66.75	68.36	77.64
	OA-SORT	76.29	67.06	68.74	78.19
Public	Hybrid-SORT	46.55	50.16	59.49	56.89
	OA-SORT	46.64	50.22	59.54	57.04

Table 12. Ablation under different detection on MOT17-val.



(a) The ID situation under detection provided by Host. The left is Hybrid-SORT's ID result, and the right is OA-SORT's result.

(b) The ID situation under detection provided by Hybrid-SORT. The left is Hybrid-SORT's ID result, and the right is OA-SORT's result.

Figure 6. The results on MOT17-09. The different color indicates different trajectory number and the dots on the line represent the starting point of trajectory interruption or identity change. GT ID represents the actual ID. The number in the legend represents the trajectory number rather than GT ID. The legend also reflects the total number of trajectories in the sequence.

or without CMC, we find that CMC brings benefits for occlusion-aware framework in HOTA. When CMC is not used, the HOTA and MOTA performance is improved by 0.31% and 0.64% from Hybrid-SORT to OA-SORT. In contrast, HOTA and MOTA enhancement with CMC is 0.39%

CMC	Method	MOTA	HOTA	AssA	IDF1	IDS
✓	Hybrid-SORT	75.65	66.75	68.36	77.64	250
	Hybrid-SORT	76.21	67.83	70.15	80.25	204
✓	OA-SORT	76.29 (+0.64)	67.06 (+0.31)	68.74	78.19	239
	OA-SORT	76.75 (+0.54)	68.22 (+0.39)	70.73	80.93	170

Table 13. The influence of CMC on MOT17-val.

and 0.54%.

8.3. Additional data visualization under Occlusion

According to Fig. 1, the additional data, IDF1 and MOTA, are provided as shown in Fig. 7. Although the HOTA in the #0026 sequence has decreased, both IDF1 and MOTA have improved, especially in terms of MOTA, which has increased by approximately 1.4%. Combing Fig. 1, the results show that the main reason for this situation may be that the proposed occlusion-aware framework focuses on instantaneous states rather than long-term, resulting in unstable tracking.

8.4. MOT20 test set

In Tab. 14, the results under MOT20 [10] are presented. OA-SORT maintains good performance. Compared with Hybrid-SORT, OA-SORT improves with 0.4 IDF1 and 0.3 AssR. The results indicate that OAO and BAM can still maintain performance in dense-occlusion scenarios.

8.5. Visualization

We analyze a video segment from the DanceTrack0005 under Hybrid-SORT and OA-SORT, as shown in Fig. 8 and Fig. 9. In frame 66 (in Fig. 8), Hybrid-SORT fails to maintain IDs of #3 and #2 due to position cost confusion brought by occlusion. In contrast, OA-SORT can stable their IDs utilizing OAO. However, occlusion-aware framework is difficult to handle severe inaccurate detection under strong competition, as shown for bounding box #4 and #5 in the Fig. 10.

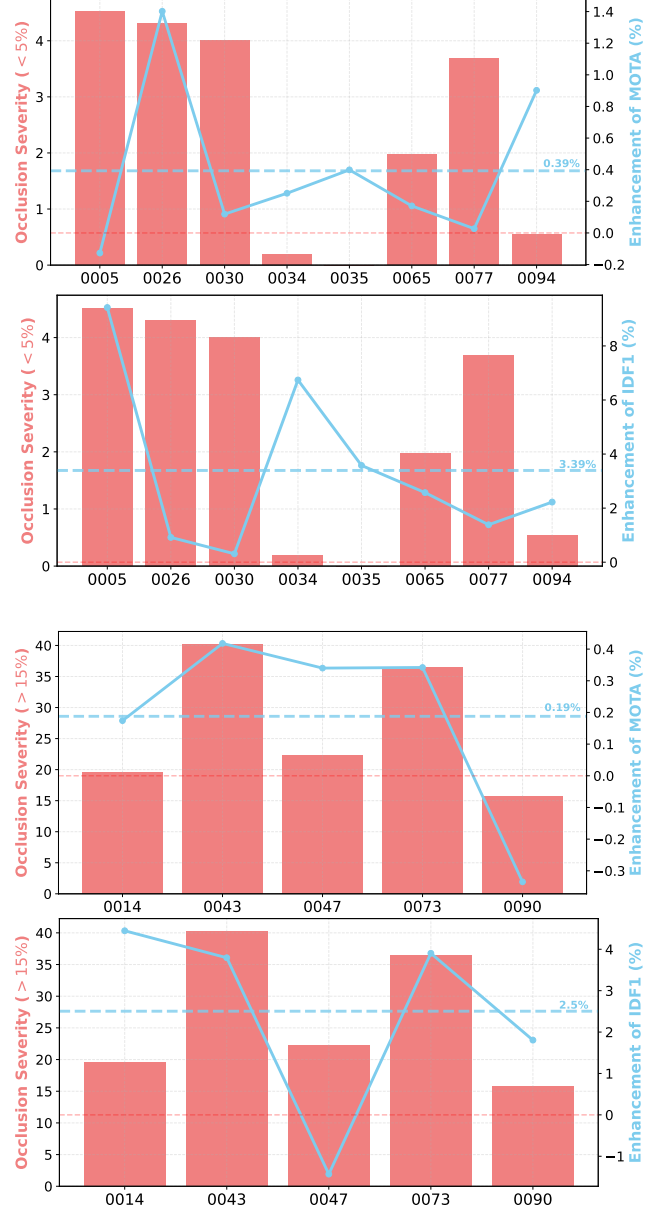
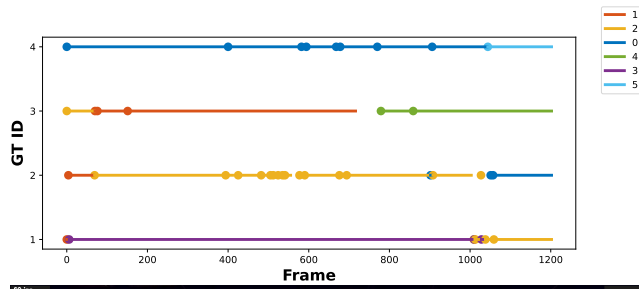


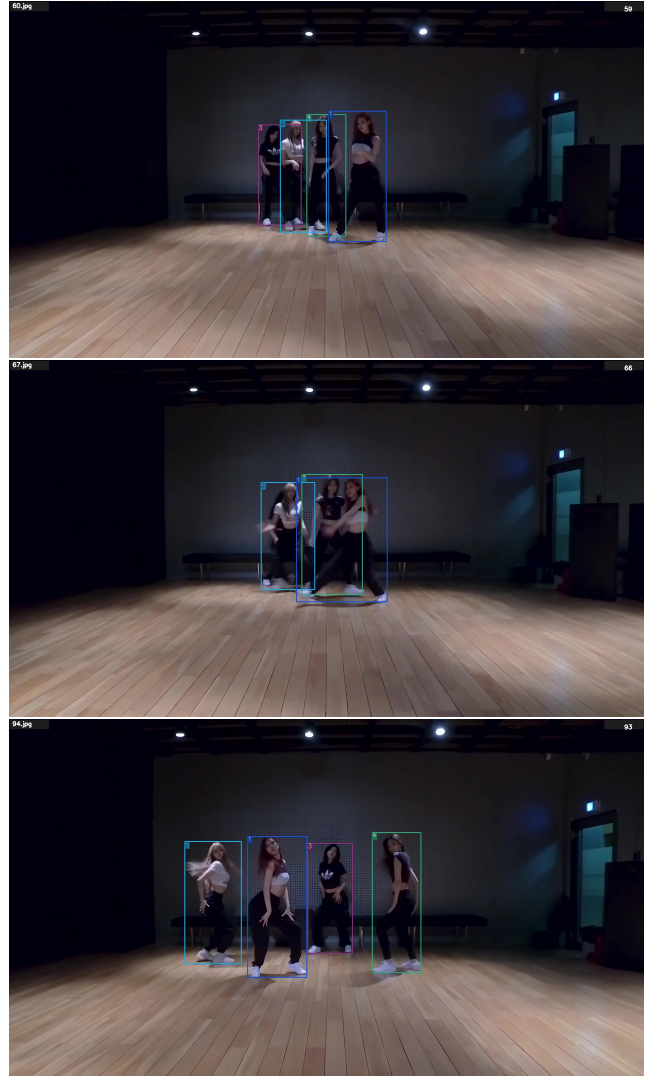
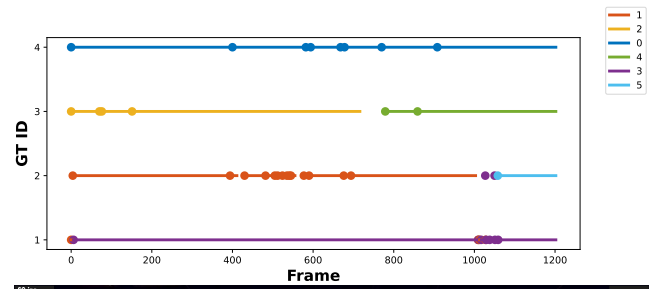
Figure 7. Performance improvement in MOTA and IDF1 over Baseline on the DanceTrack validation set under high and low occlusion severity.

Method	HOTA (%)↑	MOTA (%)↑	IDF1 (%)↑	FP (10^4)↓	FN (10^4)↓	IDs↓	AssA (%)↑	AssR (%)↑
FairMOT [61]	54.6	61.8	67.3	10.30	8.89	5,243	54.7	60.7
CSTrack [29]	54.0	66.6	68.6	2.54	14.4	3,196	54.0	57.6
TransMOT [8]	61.9	77.5	75.2	3.42	8.08	1,615	60.1	66.3
UTM [59]	62.5	74.3	79.8	3.00	8.15	1,228	-	-
GHOST [44]	61.2	73.7	75.2	-	-	1,264	-	-
ByteTrack [62]	61.3	77.8	75.2	2.62	8.76	1,223	59.6	66.2
OC-SORT [6]	62.1	75.5	75.9	1.80	10.8	913	62.0	67.5
StrongSORT [12]	61.5	72.2	75.9	-	-	1,066	63.2	-
AIPT [63]	62.1	76.9	75.4	2.10	9.80	1,134	61.1	-
Hybrid-SORT [58]	62.5	76.4	76.2	3.59	8.50	1,300	62.0	68.4
OA-SORT	62.6	76.5	76.6	3.59	8.46	1,274	62.1	68.7

Table 14. Results on MOT20-test with the private detections.

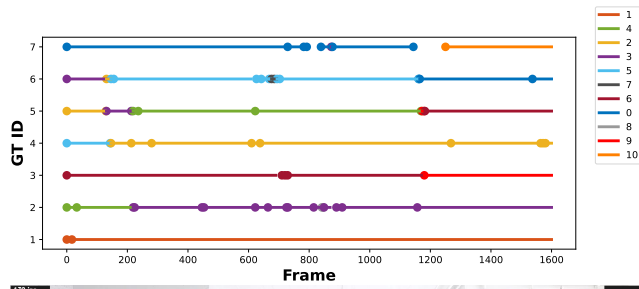


(a) The results through Hybrid-SORT.

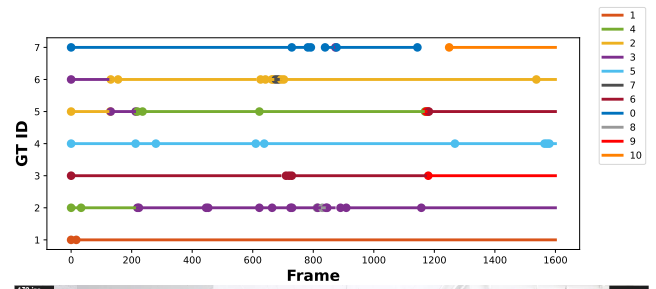


(b) The results through OA-SORT.

Figure 8. The results on DanceTrack0058. The different color indicates different trajectory number and the dots on the line represent the starting point of trajectory interruption or identity change. GT ID represents the actual ID. The number in the legend represents the trajectory number rather than GT ID. The legend also reflects the total number of trajectories in the sequence.



(a) The results through Hybrid-SORT.

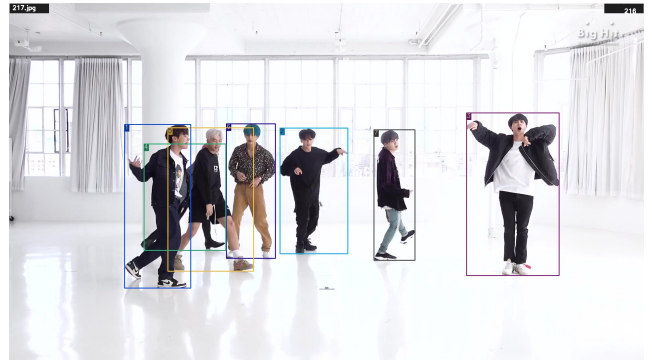


(b) The results through OA-SORT.

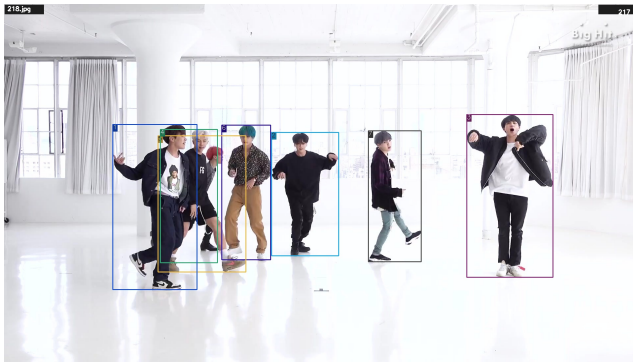
Figure 9. The results on DanceTrack0058. The different color indicates different trajectory number and the dots on the line represent the starting point of trajectory interruption or identity change. GT ID represents the actual ID. The number in the legend represents the trajectory number rather than GT ID. The legend also reflects the total number of trajectories in the sequence.



(a) Frame 215



(b) Frame 216



(c) Frame 217



(d) Frame 218

Figure 10. The failure case on DanceTrack0058. In Frame 217, due to the inaccurate detections of #4 and #5, the occlusion processing failed.