TRACK-OPD: LEARNING FROM FUTURE CONSEQUENCES FOR MULTI-OBJECT TRACKING VIA ON-POLICY DISTILLATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Online multi-object tracking associates observations using the identity history created by earlier decisions. Learning from the future consequences of these decisions requires comparing alternative associations, yet each alternative can change which tracks survive and which objects their identities denote. We propose **Track-OPD**, an on-policy distillation method that makes these consequences comparable under fixed identity bindings. Alternative associations continue from the same tracker state under shared future observations and a frozen policy. Their outcomes are evaluated on the same annotated objects and frames, retaining missing and incorrect predictions in the comparison. The resulting future identity preferences are aligned with current association classes to construct a local teacher. This teacher redistributes probability between the ground-truth reference and verified poorer alternatives while preserving their total probability. On DanceTrack and SportsMOT, Track-OPD improves HOTA over continued training by 1.03–1.36 points across the evaluated tracker adaptations. At 4,000 updates, its target construction gains 0.543 HOTA and 1.184 AssA over unit ground-truth preferences on the same selected records. These results support learning current associations from the identity history they produce while retaining the original online inference structure.

1 INTRODUCTION

Multi-object tracking (MOT) maintains object identities by associating observations across video frames. Online trackers make these associations using motion and appearance histories (Bewley et al., 2016; Wojke et al., 2017; Bergmann et al., 2019; Zhang et al., 2022; Cao et al., 2023) or learned memories, queries, and identity representations (Cai et al., 2022; Meinhardt et al., 2022; Zeng et al., 2022; Gao & Wang, 2023; Gao et al., 2025; Wang et al., 2025). Each association updates this history and thus influences later decisions. An assignment made during overlap, for example, can enter the identity memory and affect tracking after the objects separate. Learning an association therefore requires considering both the present correspondence and the identity history it creates.

Ground-truth (GT) annotations specify the current correspondence, but competing assignments can produce different future identity histories. An incorrect assignment may be corrected at the next clear observation or persist after separation. Training on learner-visited histories addresses the mismatch between training and deployment states (Ross et al., 2011; Bengio et al., 2015; Agarwal et al., 2024), while rollout-based learning supervises actions through their consequences (Chang et al., 2015). For recurrent MOT, these consequences can refine current association probabilities using the tracker's subsequent behavior, with GT defining the reference identity.

Comparing these consequences is difficult because **the decision changes both the identity trajectory and the predictions used to judge it**. Continuing the tracker after each alternative creates a separate branch, which can lose a track, create a duplicate, or carry an identity onto a different object. Scoring only surviving trajectories can reward the disappearance of a difficult case, while rebinding identities afterward can hide the original switch. A comparable evaluation must therefore keep identity bindings and observable objects and frames fixed across branches. Evaluating the branches on this common evidence yields *future identity preferences*, relative scores that describe their identity outcomes within a finite continuation window (Figure 1).

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

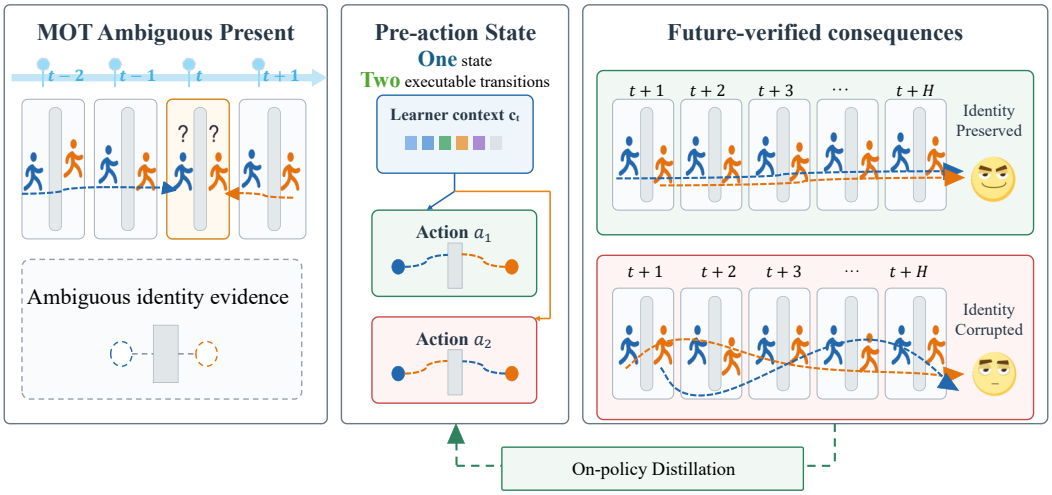


Figure 1. Learning current associations from future identity consequences. The schematic shows ambiguous identity evidence, two legal actions from independent copies of the same state, and their subsequent identity trajectories. Training compares these outcomes under common future observations and a frozen continuation policy. Online inference uses past and current observations.

We propose **Track-OPD** to turn these preferences into supervision for current associations. Track-OPD continues legal actions from copies of a visited state under common future observations and a frozen policy. It verifies their outcomes using fixed GT bindings and common observable object–frame pairs, including missing predictions. The resulting preferences are mapped to current association classes through *semantic alignment*. A *local teacher* then redistributes probability between the GT reference and verified poorer alternatives, preserving their total probability and all outside probabilities. Distillation trains the current association distribution toward this teacher alongside the tracker's original objective.

Track-OPD improves the official DanceTrack test results for MeMOTR, SambaMOTR, and MOTIP, gaining up to **1.36 HOTA** over continued training. Its MOTIP adaptation gains **1.03 HOTA** on SportsMOT, with AssA and IDF1 improving in all four comparisons. To examine the target construction, a controlled DanceTrack study fixes the 4,000-update budget and selected supervision records. On these records, Track-OPD exceeds a GT preference target, which assigns one to the reference and zero to its competitors, by **0.543 HOTA**, **1.184 AssA**, and **0.709 IDF1** in the three-run means.

Our contributions are as follows:

- We formulate **comparable future identity preferences** for recurrent tracking. Fixed identity bindings and common observable object–frame pairs preserve the basis of comparison as predicted trajectories change or disappear.
- We propose **Track-OPD** to learn current associations from these preferences. Semantic alignment preserves their identity correspondence, and a local teacher converts them into a bounded probability correction around the GT reference.
- We demonstrate experimentally that **Track-OPD improves association quality** on DanceTrack and SportsMOT across the evaluated tracker adaptations. Matched target comparisons and bidirectional transfer support learning from the identity history produced by a decision within the original online inference structure.

2 Related Work

Association as a sequential decision. Online trackers maintain identities through matching rules, queries, or recurrent memories (Zhang et al., 2022; Zeng et al., 2022; Gao & Wang, 2023; Gao et al., 2025; Segu et al., 2025). Their decisions change the histories on which later associations depend. Training on learner-visited histories addresses the resulting exposure bias (Ross et al., 2011; Maksai & Fua, 2019), while AggreVaTe and LOLS use rollout consequences to supervise actions (Ross &

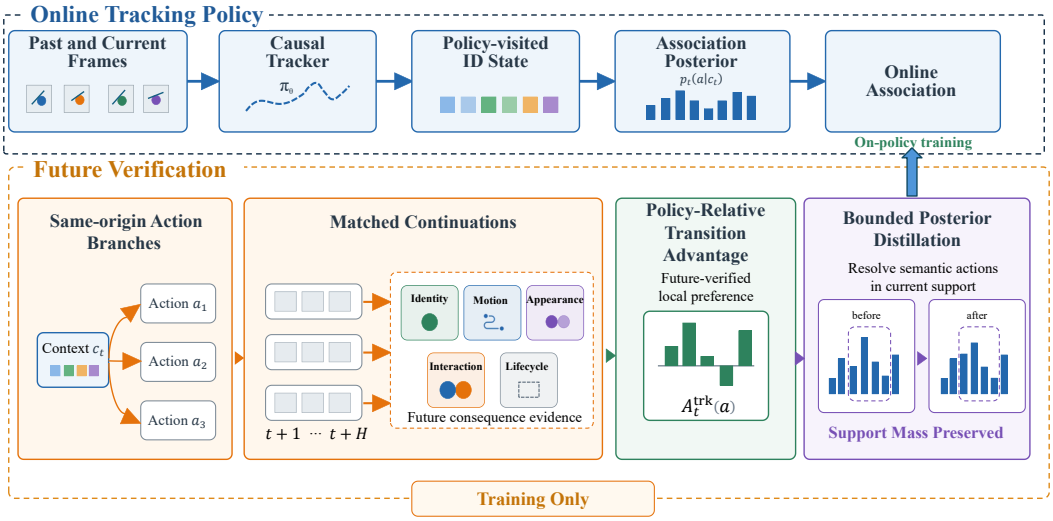


Figure 2. Track-OPD overview. Legal actions continue from context c_t under common future inputs and a frozen policy. GT verification combines five primitive scores and error penalties into four channels, yielding future identity preferences A_t^{trk}, labeled Policy-Relative Transition Advantage. Semantic alignment maps them to current association classes. Bounded Posterior Distillation constructs the local teacher and distills its correction into the tracker.

Bagnell, 2014; Chang et al., 2015). Track-OPD compares such consequences in recurrent tracking under fixed identity bindings and common observable object–frame pairs.

Privileged information for distillation. On-policy distillation trains on learner-generated histories (Agarwal et al., 2024), with privileged information providing additional supervision for those histories. Vision-OPD enriches perception targets with privileged information (Yuan et al., 2026), TOPD uses near-future teacher and student trajectories (Jiang & Ferraro, 2026), and AR-OPD transfers a full-view residual relative to a partial-view teacher (Zhang, 2026). Track-OPD uses future observations and identity annotations to verify alternative actions from the same tracker. The resulting preferences define a local teacher that redistributes probability within the semantically aligned comparison.

Trajectory-based supervision. Counterfactual and hindsight methods assign credit through alternative actions and observed outcomes (Foerster et al., 2018; Harutyunyan et al., 2019), and CRAFT uses sibling rollouts (Meng & Chen, 2026). In tracking, path consistency supervises association distributions across different observation paths (Lu et al., 2024). Track-OPD instead compares alternative initial associations under common future observations and a frozen continuation policy. Fixed identity bindings make these trajectories comparable, and semantic alignment attaches their preferences to the current decision. Appendix D.5 gives additional context.

3 Method

Track-OPD learns from future identity preferences that preserve their meaning across branch evaluation and training replay (Figure 2). Fixed identity bindings and common observable object–frame pairs establish this correspondence during evaluation. Semantic alignment preserves it when mapping preferences to current association classes. A local teacher refines probabilities around the GT reference on these classes, providing a training target while preserving the tracker's original inference computation.

3.1 Comparing Future Identity Consequences

Alternative actions must start from the same state for their consequences to be compared. At each epoch boundary, we freeze a tracker snapshot and use it to collect states from training streams with online observations. The pre-action context c_t contains the current observation, identity representations, history, and life-cycle state. At an ambiguous decision, the executed action and up to four

legal alternatives define the candidate set. Each candidate a uses the native transition on a copy of c_t to produce the post-decision state $s_{t^+}^{(a)}$. For association actions, this transition recomputes the remaining one-to-one allocation after the forced assignment. Candidate selection uses current and past information; annotations are consulted after branching (Appendix C.1).

From this state, each branch continues over a horizon H of 15 frames under common future inputs $\xi_{t+1:t+H}$. The frozen snapshot at epoch e, with parameters $\bar{\theta}_e$, supplies the collection policy μ and the branch trajectory

$$\tau_t^{\mu}(a) = \mathrm{Rollout}_{\mu}(s_{t^+}^{(a)}, \xi_{t+1:t+H}), \qquad \mu = \pi_{\bar{\theta}_e}. \tag{1}$$

The shared inputs include observations, detections, preprocessing, and stochastic choices. With the starting context and collection policy fixed, branches differ in their initial association and the state evolution it induces. Their comparison therefore describes consequences under this tracker and horizon, rather than optimal action values over all future policies.

Keeping the comparison fixed. As branch trajectories evolve, existing tracks retain the GT identity bindings established from their preceding histories. These bindings prevent a track that switches objects from being reassigned to the new object for scoring. A legal birth at a verified first appearance is bound once after branching, while a duplicate cannot obtain a new correct binding. The current detection and retained candidates determine which GT identities are evaluated. Their available annotations define a common observable object–frame mask across branches. Within this mask, a missing prediction receives zero correctness without shrinking the denominator; a missing annotation excludes the pair from every branch (Appendix A).

The common mask keeps an annotated object in the comparison even when one branch loses its track. Scoring surviving tracks alone would omit this error; the common mask instead assigns zero correctness. If the track follows another object, its fixed binding preserves the identity switch. Both failures remain visible on the same evaluation population.

On these common object–frame pairs, four verification channels score geometry, identity continuity, interactions, and life-cycle outcomes. A channel with unresolved evidence is excluded for every candidate. For each available channel m, we subtract the collection-policy mean from its score and divide by the uniform population standard deviation over candidates. This gives the channel preference $A_t^m(a)$, which contributes to the aggregate future identity preference

$$A_t^{\mathrm{trk}}(a) = \frac{\sum_m w_t^m \kappa_t^m A_t^m(a)}{\chi_t}, \qquad \chi_t = \sum_m w_t^m \kappa_t^m. \tag{2}$$

The weight w_t^m is uniform over available channels relevant to the event, while reliability κ_t^m reflects observability and score separation. The weighted sum of reliabilities defines aggregate confidence χ_t, shared by all candidates. Identical candidate scores give zero channel reliability; zero-confidence records are discarded. Retained records store preferences and confidence for training (Appendix A.3).

3.2 Aligning Preferences with Current Associations

Stored future identity preferences belong to candidate actions identified by their semantic identities and action types. Training must map these actions to the tracker's current association classes. For each supervised decision, indexed by token i, we replay the saved state and transformations and recompute its outputs. An identity-association action follows its current identity class (Gao et al., 2025); if that class moves from column 2 to 5, its preference follows column 5. The resulting aligned preference vector is V_i, with ambiguous or missing mappings excluded. Gradients pass through current association outputs, while saved history and future trajectories remain detached. This formulation uses MOTIP's logits; Appendix C.2 describes the query adaptations.

Among the aligned classes, the GT reference y_i must correspond to an evaluated legal action. We retain this reference with a set $\mathcal{N}_i$ of competitors whose verified preferences are lower, defining the comparison support

$$S_i = \{y_i\} \cup \mathcal{N}_i, \qquad V_i(y_i) > V_i(j) \quad \text{for all } j \in \mathcal{N}_i. \tag{3}$$

The support must contain at least one competitor and pass confidence, preference-gap, and entropy thresholds (Appendix B.1). Actions scoring above the GT reference remain outside it. A life-cycle action can enter the support only if it maps uniquely to a differentiable association output.

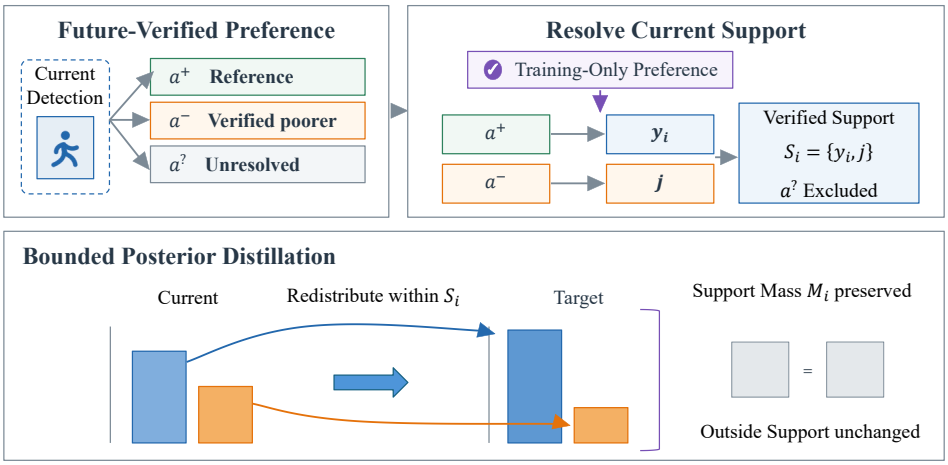


Figure 3. Local teacher construction for Bounded Posterior Distillation. The GT reference and verified poorer alternatives form comparison support S_i, labeled Verified Support. The teacher preserves its total probability M_i and all outside probabilities, with the change bounded by total variation (TV). Two of up to five supported actions are shown.

The retained support thus identifies the classes to be refined, with GT defining the reference and verification supplying the preference values. These values determine both the teacher's relative probability correction and its step size.

3.3 Constructing a Local Association Teacher

The preferences within S_i describe relative outcomes but do not specify how much probability the support should receive. The local teacher retains this total probability from the current association distribution and redistributes it within S_i. For the vector z_i of finite logits over legal classes, the current distribution and its detached copy are

$$p_i = \mathrm{softmax}(z_i), \qquad p_i^0 = \mathrm{stopgrad}[p_i]. \tag{4}$$

Here stopgrad stops gradient propagation. The detached distribution defines the support mass M_i and the conditional probabilities r_i within that support:

$$M_i = \sum_{k \in S_i} p_i^0(k), \qquad r_i(k) = p_i^0(k)/M_i. \tag{5}$$

We refine these conditional probabilities by exponentially weighting them with the aligned preferences:

$$\widehat{r}_i(k) = \frac{r_i(k)\exp(\eta_i V_i(k))}{\sum_{j \in S_i} r_i(j)\exp(\eta_i V_i(j))}. \tag{6}$$

The step η_i controls the strength of this correction. Its initial value combines record confidence, posterior maturity, and the clipped preference span. Maturity measures how strongly the collection policy favors the GT reference relative to a uniform distribution. We reduce the step when needed to satisfy a total-variation (TV) bound δ (Appendix B). Rescaling the refined conditional distribution by M_i gives the teacher, while retaining outside probabilities:

$$q_i|_{S_i} = M_i \widehat{r}_i, \qquad q_i|_{S_i^c} = p_i^0|_{S_i^c}. \tag{7}$$

The resulting teacher is normalized, preserves support mass, and satisfies

$$\sum_{k \in S_i} q_i(k) = M_i, \qquad \mathrm{TV}(q_i, p_i^0) = \frac{M_i}{2}\|\widehat{r}_i - r_i\|_1 \le \delta. \tag{8}$$

For a fixed step and fixed outside probabilities, this teacher (Figure 3) maximizes preference-weighted probability with a Kullback–Leibler (KL) penalty for departure from the detached current distribution (Appendix B.3). Its normalization and TV-search feasibility are established in Appendix B.4. These properties constrain the teacher distribution; updating shared model parameters can still change predictions elsewhere.

3.4 DISTILLING INTO THE ONLINE TRACKER

The tracker learns from this teacher by minimizing its KL divergence from the current association distribution:

$$\ell_{\mathrm{dist},i} = \mathrm{KL}(q_i \| p_i) = \sum_k q_i(k) \log \frac{q_i(k)}{p_i(k)}. \tag{9}$$

For classes k and j in the comparison support, the teacher log-odds equal the current log-odds plus the verified preference correction:

$$\log \frac{q_i(k)}{q_i(j)} = \mathrm{stopgrad}\big[z_i(k) - z_i(j) + \eta_i\big(V_i(k) - V_i(j)\big)\big]. \tag{10}$$

For a reference–competitor pair, the positive preference gap raises the reference's target odds. The amount of correction depends on the normalized preference gap and accepted step, so it need not reverse the current ranking. With the teacher detached, the distillation gradient at the current logits is

$$\nabla_{z_i} \ell_{\mathrm{dist},i} = p_i - q_i. \tag{11}$$

This gradient is zero outside the comparison support, and its entries within the support sum to zero. The distillation correction is thus confined to the verified alternatives at the current logits. The tracker's original training objective, termed the native objective $\mathcal{L}_{\mathrm{base}}$, remains responsible for the full association problem, including decisions with unresolved future evidence.

Alongside KL distillation, soft pairwise probability and bounded log-odds losses match the same teacher preferences. All three terms are normalized by the number of valid association tokens to form the auxiliary objective $\mathcal{L}_{\mathrm{aux}}$. At a shared differentiable association tensor u_i, approximate conflict removal and a norm cap constrain its gradient relative to the native gradient (Appendix B.5–B.6). MOTIP uses its association logits as this tensor. These local constraints provide no parameter-space or future-trajectory guarantee.

3.5 TRAINING PROCEDURE

Algorithm 1 Track-OPD post-training

Require: training streams $\mathcal{D}$, pretrained tracker π_θ, verification horizon H
for each post-training epoch e **do**
  Freeze $\bar{\theta}_e \leftarrow \theta$ and set $\mu \leftarrow \pi_{\bar{\theta}_e}$
  Visit $\mathcal{D}$ with μ; branch ambiguous legal decisions from copied states
  Continue each branch with common future inputs and frozen policy μ
  Verify branch outcomes on common evidence and retain resolved records $\mathcal{Q}_e$
  for each training batch B **do**
    Replay states; align preferences; compute $\mathcal{L}_{\mathrm{base}}$, scores z, and shared tensors u
    Form support S_i, detached teacher q_i, and $\mathcal{L}_{\mathrm{aux}}$ for eligible tokens
    Obtain constrained $\widetilde{g}_i^{\mathrm{a}}$ using Eq. B.22
    Compute g_θ by Eq. B.25 and apply one optimizer update
  end for
end for
Ensure: tracker with its original online inference computation

Algorithm 1 alternates collection with training. The frozen collection policy defines states and future identity preferences that remain fixed within each epoch. Training replays these states and recomputes the current association distribution, with saved history and future trajectories detached. Rebuilding the local teacher from this distribution applies the same verified preferences to the tracker's evolving probabilities. The native objective and constrained auxiliary gradient jointly determine each optimizer update.

At the next epoch, a new snapshot refreshes the states and preferences. Snapshot refresh updates the evidence, while teacher recomputation adapts corrections between refreshes. Future observations and annotations are used during collection; online inference retains the original tracker computation with past and current observations.

4 EXPERIMENTS

Table 1. Official test results without extra data. Red parentheses give gains over continued native training (Continue). Appendix D.2 specifies aggregation. Bold marks the best result per dataset.

Method	HOTA ↑	DetA ↑	AssA ↑	MOTA ↑	IDF1 ↑
DanceTrack					
MOTR (Zeng et al., 2022)	54.20	73.50	40.20	79.70	51.50
CO-MOT (Yan et al., 2025)	65.30	80.10	53.50	89.30	66.50
MeMOTR (Gao & Wang, 2023)	68.50	80.50	58.40	89.90	71.20
MeMOTR + Continue	68.71	80.16	58.82	89.79	71.28
SambaMOTR (Segu et al., 2025)	69.00	79.50	60.00	89.10	74.20
SambaMOTR + Continue	69.08	79.86	59.81	89.24	74.14
MOTIP (Gao et al., 2025)	69.60	80.40	60.40	90.60	74.70
MOTIP + Continue	69.98	80.33	61.15	90.55	74.62
MeMOTR + Track-OPD	69.86 (+1.15)	80.62 (+0.46)	61.07 (+2.25)	91.31 (+1.52)	72.78 (+1.50)
SambaMOTR + Track-OPD	70.38 (+1.30)	80.94 (+1.08)	61.28 (+1.47)	90.72 (+1.48)	75.37 (+1.23)
MOTIP + Track-OPD	**71.34** (+1.36)	**81.42** (+1.09)	**62.69** (+1.54)	**91.73** (+1.18)	**76.46** (+1.84)
SportsMOT					
FairMOT (Zhang et al., 2021b)	49.30	70.20	34.70	86.40	53.50
TrackFormer (Meinhardt et al., 2022)	63.30	66.00	61.10	74.10	72.40
MeMOTR (Gao & Wang, 2023)	68.80	82.00	57.80	90.20	69.90
MOTIP (Gao et al., 2025)	72.60	83.50	63.20	92.40	77.10
MOTIP + Continue	72.73	83.49	63.44	92.50	77.49
MOTIP + Track-OPD	**73.76** (+1.03)	**84.51** (+1.02)	**64.82** (+1.38)	**93.37** (+0.87)	**78.63** (+1.14)

4.1 EXPERIMENT SETUP

Implementation Details. We apply Track-OPD to pretrained MOTIP, MeMOTR, and SambaMOTR (Gao et al., 2025; Gao & Wang, 2023; Segu et al., 2025). All retain native training and share verification, alignment, teacher construction, auxiliary losses, and gradient constraints. MOTIP trains its identity modules with detection frozen; query trackers use existing representations (Appendix C.2). All collect 15-frame continuations with epoch refresh and retain native online inference.

Comparison Methods. The Continue baseline extends each tracker's native training procedure and objective. The controlled MOTIP study compares Continue, a GT preference control, and Track-OPD at the final 4k-update checkpoint, with shared initialization, detections, data, optimization, and evaluation. The GT preference control uses Track-OPD's selected records but assigns one to the reference and zero to competitors. All other target construction and training settings remain fixed (Appendix D.1).

Evaluation Metrics. Official results use the DanceTrack and SportsMOT test splits (Sun et al., 2022; Cui et al., 2023); controlled MOTIP studies use the DanceTrack training and validation splits. We report HOTA with detection accuracy (DetA) and association accuracy (AssA) (Luiten et al., 2021), identity F1 (IDF1) (Ristani et al., 2016), and available MOTA and identity-switch counts (IDSW). Compared methods share dataset protocols and non-target settings. Appendix D.2 specifies aggregation and dispersion; reported update counts exclude collection.

4.2 COMPARISON WITH EXISTING TRACKERS

Quantitative Results. Track-OPD improves all four tracker–dataset comparisons in Table 1. In the reported five-run medians, its HOTA gains over Continue range from 1.15 to 1.36 points on DanceTrack and reach 1.03 on SportsMOT. Each comparison also improves AssA and IDF1, with the AssA gain exceeding the DetA gain. These association improvements occur across all three tracker adaptations, and MOTIP with Track-OPD achieves the highest listed results on both datasets.

4.3 Ablation Study and Discussion

Effect of the Association Target. Table 2 compares Track-OPD with the GT preference control on the same supervised decisions at 4k updates. Track-OPD gains **0.543 HOTA**, **1.184 AssA**, and **0.709 IDF1** in the three-run means. Relative to Continue, the GT preference control improves HOTA and IDF1 but lowers AssA by 0.038, whereas Track-OPD improves all three. This supports using future identity preferences to construct the target on shared records. The values enter both exponential weighting and step selection, so the comparison measures their joint effect (Appendix D.1).

Table 2. DanceTrack validation at 4k updates. Values are three-run means; GT preference and Track-OPD share selected records.

Method	HOTA ↑	AssA ↑	IDF1 ↑
Continue	64.913	56.205	69.748
GT preference	65.473	56.167	70.686
Track-OPD	**66.016**	**57.351**	**71.395**

Role of Local Update Controls. Table 3 compares posterior maturity and local gradient control configurations. Full Track-OPD achieves the highest HOTA, AssA, and IDF1 and the fewest identity switches. Its HOTA is 66.016, compared with 64.340, 64.998, and 64.375 for the maturity, gradient-constraint, and norm-bound variants. These comparisons favor maturity weighting together with local gradient control. They complement the matched target study in Table 2 by supporting the full update configuration for learning from future identity preferences (Appendix D.3).

Table 3. Local update controls on DanceTrack validation. Appendix D.3 specifies the comparison scope.

Variant	HOTA ↑	AssA ↑	IDF1 ↑	IDSW ↓
w/o post. maturity	64.340	55.055	69.132	2932
w/o grad. constraint	64.998	56.225	70.253	2950
w/o grad. norm bound	64.375	55.244	69.146	2900
Track-OPD	**66.016**	**57.351**	**71.395**	**2740**

Choice of the Update Budget. Track-OPD exceeds 10k-update Continue on all three metrics after 3k updates (Table 4). At 4k, it gains 1.103 HOTA over 4k-update Continue, whereas extending Continue from 4k to 10k adds 0.021 HOTA. Track-OPD reaches its highest tested HOTA at 4k and declines to 64.998 at 10k. This trajectory supports a limited post-training budget in this setting. The update counts exclude collection, so the comparison measures optimizer budget rather than wall-clock speed. Appendix D.4 visualizes the same results.

Table 4. DanceTrack validation across optimizer update budgets, excluding collection.

Method	Steps (k)	HOTA ↑	AssA ↑	IDF1 ↑
None	0	64.070	54.689	69.106
Continue	4	64.913	56.205	69.748
Continue	10	64.934	56.184	69.727
Track-OPD	1	64.421	55.474	69.316
Track-OPD	3	65.353	56.754	70.336
Track-OPD	4	**66.016**	**57.351**	**71.395**
Track-OPD	10	64.998	56.333	70.325

How Does Association Transfer across Domains? The association improvement transfers to the other dataset without target-specific tuning (Table 5). Relative to Continue, Track-OPD gains 3.109 HOTA from DanceTrack to SportsMOT and 4.432 in the reverse direction. The corresponding AssA gains are 4.880 and 5.030, with IDF1 gains of 5.800 and 7.047. Improvements in both identity metrics and both transfer directions support the learned association behavior beyond the source domain under the fixed detector configuration.

Table 5. Transfer from source to target without target tuning. D and S denote DanceTrack and SportsMOT.

	D → S		S → D	
Metric	Continue	Track-OPD	Continue	Track-OPD
HOTA ↑	38.984	**42.093**	24.255	**28.687**
DetA ↑	41.172	**42.405**	37.246	**39.575**
AssA ↑	36.980	**41.860**	15.892	**20.922**
IDF1 ↑	38.031	**43.831**	25.795	**32.842**

Role of Motion Prediction. Track-OPD retains its advantage without motion prediction (Table 6). Disabling motion prediction reduces HOTA by 2.971 for MOTIP and 1.343 for Track-OPD, leaving gains of 2.762 HOTA and 4.704 IDF1 for Track-OPD. Its smaller drops in both metrics indicate reduced sensitivity to disabling motion prediction. This comparison uses paired checkpoints separate from the 4k study.

Table 6. Motion prediction enabled (On) or disabled (Off) over 25,508 validation frames. Parentheses give changes from On; checkpoints differ from the 4k study.

Metric	Pred.	MOTIP	Track-OPD
HOTA	On	64.007	**65.141**
	Off	61.036 (−2.971)	**63.798** (−1.343)
IDF1	On	68.734	**69.169**
	Off	63.301 (−5.433)	**68.005** (−1.164)

When Should Future Supervision Begin? Figure 4 compares Track-OPD introduction epochs during training from scratch. Starting at epoch 8 reaches 46.139 HOTA, exceeding ordinary training by 2.857 points. Starts at epochs 0, 2, and 4 give 28.904, 37.675, and 43.751. The improvement with later introduction supports applying Track-OPD after initial tracker training. This schedule determines when future supervision begins; posterior maturity in Section 3.3 determines the strength of each correction.

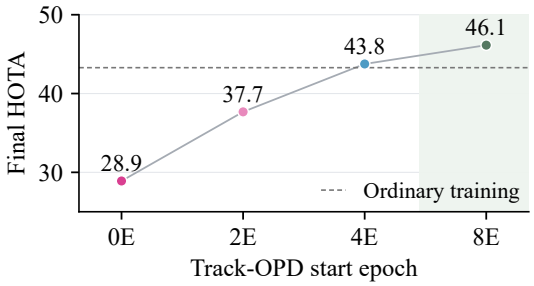


Figure 4. Track-OPD introduction epoch during training from scratch. The dashed line denotes ordinary training.

Qualitative Results. Figure 5 follows predicted identities through overlap and separation. Both methods initially assign predicted identity P6 to GT identity 14. After separation, MOTIP assigns P3 to this object, whereas Track-OPD retains P6. Across the shared interval, identity switches decrease from 88 to 69. The later persistence of P6 illustrates the type of future identity evidence used for learning.



Figure 5. Identity trajectories on DanceTrack sequence 0041. P and GT denote predicted and ground-truth identities. Track-OPD retains P6 for GT 14 through overlap and separation. The selected interval contains 88 identity switches for MOTIP and 69 for Track-OPD across all trajectories.

5 Conclusion

Track-OPD learns current associations from the identity histories they produce. It compares alternative associations under fixed identity bindings and common observable object–frame pairs, then aligns the resulting future identity preferences with current classes to construct a local teacher. Results across tracker adaptations, the matched target comparison, and domain transfer support this supervision for improving association while retaining the original online inference structure.

AI USE STATEMENT

Generative AI tools were used only for writing and language polishing.

REFERENCES

Rishabh Agarwal, Nino Vieillard, Yongchao Zhou, Piotr Stanczyk, Sabela Ramos Garea, Matthieu Geist, and Olivier Bachem. On-policy distillation of language models: Learning from self-generated mistakes. In *International Conference on Learning Representations*, volume 2024, pp. 21246–21263, 2024.

Samy Bengio, Oriol Vinyals, Navdeep Jaitly, and Noam Shazeer. Scheduled sampling for sequence prediction with recurrent neural networks. In *Advances in Neural Information Processing Systems*, volume 28, 2015.

Philipp Bergmann, Tim Meinhardt, and Laura Leal-Taixé. Tracking without bells and whistles. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 941–951, 2019.

Alex Bewley, Zongyuan Ge, Lionel Ott, Fabio Ramos, and Ben Upcroft. Simple online and realtime tracking. In *2016 IEEE International Conference on Image Processing (ICIP)*, pp. 3464–3468, 2016.

Jiarui Cai, Mingze Xu, Wei Li, Yuanjun Xiong, Wei Xia, Zhuowen Tu, and Stefano Soatto. MeMOT: Multi-object tracking with memory. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8090–8100, 2022.

Jinkun Cao, Jiangmiao Pang, Xinshuo Weng, Rawal Khirodkar, and Kris Kitani. Observation-centric SORT: Rethinking SORT for robust multi-object tracking. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9686–9696, 2023.

Kai-Wei Chang, Akshay Krishnamurthy, Alekh Agarwal, Hal Daumé III, and John Langford. Learning to search better than your teacher. In *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pp. 2058–2066. PMLR, 2015. URL `https://proceedings.mlr.press/v37/changb15.html`.

Yutao Cui, Chenkai Zeng, Xiaoyu Zhao, Yichun Yang, Gangshan Wu, and Limin Wang. SportsMOT: A large multi-object tracking dataset in multiple sports scenes. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 9921–9931, 2023.

Phuong Tuan Dat, Qi Li, and Xinchao Wang. dOPSD: On-policy self-distillation for diffusion language models, 2026. arXiv:2607.04428.

Shuxiao Ding, Eike Rehder, Lukas Schneider, Marius Cordts, and Juergen Gall. 3DMOTFormer: Graph transformer for online 3D multi-object tracking. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 9784–9794, 2023.

Jakob Foerster, Gregory Farquhar, Triantafyllos Afouras, Nantas Nardelli, and Shimon Whiteson. Counterfactual multi-agent policy gradients. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018. doi: 10.1609/aaai.v32i1.11794.

Ruopeng Gao and Limin Wang. MeMOTR: Long-term memory-augmented transformer for multi-object tracking. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 9901–9910, 2023.

Ruopeng Gao, Ji Qi, and Limin Wang. Multiple object tracking as ID prediction. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 27883–27893, 2025.

Anna Harutyunyan, Will Dabney, Thomas Mesnard, Mohammad Gheshlaghi Azar, Bilal Piot, Nicolas Heess, Hado van Hasselt, Gregory Wayne, Satinder Singh, Doina Precup, and Rémi Munos. Hindsight credit assignment. In *Advances in Neural Information Processing Systems*, volume 32, pp. 12467–12476, 2019.

Wenjin Hou, Shangpin Peng, Weinong Wang, Zheng Ruan, Yue Zhang, Zhenglin Zhou, Mingqi Gao, Yifei Chen, Kaiqi Wang, Hongming Yang, Chengquan Zhang, Zhuotao Tian, Han Hu, Yi Yang, Fei Wu, and Hehe Fan. Uni-OPD: Unifying on-policy distillation with a dual-perspective recipe, 2026. arXiv:2605.03677.

Yuxuan Jiang and Francis Ferraro. Bridging reasoning trajectories in on-policy distillation via near-future guidance, 2026. arXiv:2606.00305.

Jinho Kim and Kuk-Jin Yoon. Gated temporal fusion transformers for robust multi-object tracking. In *2026 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 4525–4534, 2026.

Zijia Lu, Bing Shuai, Yanbei Chen, Zhenlin Xu, and Davide Modolo. Self-supervised multi-object tracking with path consistency. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 19016–19026, 2024.

Jonathon Luiten, Aljosa Osep, Patrick Dendorfer, Philip Torr, Andreas Geiger, Laura Leal-Taixé, and Bastian Leibe. HOTA: A higher order metric for evaluating multi-object tracking. *International Journal of Computer Vision*, 129(2):548–578, 2021. doi: 10.1007/s11263-020-01375-2.

Andrii Maksai and Pascal Fua. Eliminating exposure bias and metric mismatch in multiple object tracking. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4639–4648, 2019.

Tim Meinhardt, Alexander Kirillov, Laura Leal-Taixé, and Christoph Feichtenhofer. TrackFormer: Multi-object tracking with transformers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8844–8854, 2022.

Zibin Meng and Kani Chen. CRAFT: Counterfactual credit assignment from free sibling rollouts for self-distilled agentic reinforcement learning, 2026. arXiv:2606.29476.

Thomas Mesnard, Theophane Weber, Fabio Viola, Shantanu Thakoor, Alaa Saade, Anna Harutyunyan, Will Dabney, Thomas S Stepleton, Nicolas Heess, Arthur Guez, Eric Moulines, Marcus Hutter, Lars Buesing, and Remi Munos. Counterfactual credit assignment in model-free reinforcement learning. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 7654–7664. PMLR, 2021.

Alexander Meulemans, Simon Schug, Seijin Kobayashi, Nathaniel Daw, and Gregory Wayne. Would I have gotten that reward? long-term credit assignment by counterfactual contribution analysis. In *Advances in Neural Information Processing Systems*, volume 36, pp. 68685–68735, 2023.

Ergys Ristani, Francesco Solera, Roger Zou, Rita Cucchiara, and Carlo Tomasi. Performance measures and a data set for multi-target, multi-camera tracking. In *Computer Vision – ECCV 2016 Workshops*, pp. 17–35. Springer, 2016. doi: 10.1007/978-3-319-48881-3_2.

Stéphane Ross and J. Andrew Bagnell. Reinforcement and imitation learning via interactive no-regret learning. *arXiv preprint arXiv:1406.5979*, 2014. URL `https://arxiv.org/abs/1406.5979`.

Stéphane Ross, Geoffrey Gordon, and Drew Bagnell. A reduction of imitation learning and structured prediction to no-regret online learning. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pp. 627–635, 2011.

Mattia Segu, Luigi Piccinelli, Siyuan Li, Yung-Hsu Yang, Luc Van Gool, and Bernt Schiele. Samba: Synchronized set-of-sequences modeling for multiple object tracking. In *International Conference on Learning Representations*, 2025.

Yuqing Shao, Yuchen Yang, Rui Yu, Weilong Li, Xu Guo, Huaicheng Yan, Wei Wang, and Xiao Sun. From detection to association: Learning discriminative object embeddings for multi-object tracking. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6878–6888, 2026.

Peize Sun, Jinkun Cao, Yi Jiang, Zehuan Yuan, Song Bai, Kris Kitani, and Ping Luo. DanceTrack: Multi-object tracking in uniform appearance and diverse motion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 20993–21002, 2022.

Peng Wang, Yongcai Wang, Hualong Cao, Wang Chen, and Deying Li. LA-MOTR: End-to-end multi-object tracking by learnable association. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 12438–12448, 2025.

Nicolai Wojke, Alex Bewley, and Dietrich Paulus. Simple online and realtime tracking with a deep association metric. In *2017 IEEE International Conference on Image Processing (ICIP)*, pp. 3645–3649, 2017.

Yu Xiang, Alexandre Alahi, and Silvio Savarese. Learning to track: Online multi-object tracking by decision making. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 4705–4713, 2015.

Yihong Xu, Aljosa Osep, Yutong Ban, Radu Horaud, Laura Leal-Taixé, and Xavier Alameda-Pineda. How to train your deep multi-object tracker. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6787–6796, 2020.

Feng Yan, Weixin Luo, Yujie Zhong, Yiyang Gan, and Lin Ma. CO-MOT: Boosting end-to-end transformer-based multi-object tracking via coopetition label assignment and shadow sets. In *The Thirteenth International Conference on Learning Representations*, 2025.

Xinlei Yu, Gen Li, Qingyi Si, Guibin Zhang, Yuqi Xu, Congcong Wang, Shuai Dong, Kaiwen Tuo, Xiangyu Zeng, Kaituo Feng, Qunzhong Wang, Yang Shi, Xiaobin Hu, Xiangyu Yue, Jiaqi Wang, and Shuicheng Yan. DOPD: Dual on-policy distillation, 2026. arXiv:2606.30626.

Qianhao Yuan, Jie Lou, Xing Yu, Hongyu Lin, Le Sun, Xianpei Han, and Yaojie Lu. Vision-OPD: Learning to see fine details for multimodal LLMs via on-policy self-distillation, 2026. arXiv:2605.18740.

Fangao Zeng, Bin Dong, Yuang Zhang, Tiancai Wang, Xiangyu Zhang, and Yichen Wei. MOTR: End-to-end multiple-object tracking with transformer. In *European Conference on Computer Vision*, pp. 659–675, 2022.

Wei Zhang, Lingxiao He, Peng Chen, Xingyu Liao, Wu Liu, Qi Li, and Zhenan Sun. Boosting end-to-end multi-object tracking and person search via knowledge distillation. In *Proceedings of the 29th ACM International Conference on Multimedia*, pp. 1192–1201, 2021a.

Wenhao Zhang. Beyond absolute imitation: Anchored residual guidance for privileged on-policy distillation, 2026. arXiv:2606.10385.

Yanfei Zhang, Xu Lin, and Chenglin Wu. StepOPSD: Step-aware online preference self-distillation for agent reinforcement learning, 2026. arXiv:2605.27140.

Yifu Zhang, Chunyu Wang, Xinggang Wang, Wenjun Zeng, and Wenyu Liu. FairMOT: On the fairness of detection and re-identification in multiple object tracking. *International Journal of Computer Vision*, 129(11):3069–3087, 2021b.

Yifu Zhang, Peize Sun, Yi Jiang, Dongdong Yu, Fucheng Weng, Zehuan Yuan, Ping Luo, Wenyu Liu, and Xinggang Wang. ByteTrack: Multi-object tracking by associating every detection box. In *European Conference on Computer Vision*, pp. 1–21, 2022.

A Verification of Future Consequences

A.1 Identity Binding and Common Evidence

Each predicted identity must retain the same GT binding throughout branch continuation. To establish this binding, detections are matched to GT boxes by one-to-one maximum intersection-over-union (IoU) assignment with a threshold of 0.5. An existing track is bound to the GT identity present in more than half of its valid matches over the preceding window of at most 15 frames. Uncertain bindings and identity conflicts are excluded. For a GT-confirmed first appearance, a legal birth is bound once to the current detection's GT identity after branching. A duplicate of an existing identity cannot establish a new correct binding. These bindings remain fixed during verification; GT information used to establish them does not enter online candidate selection or inference.

The GT identities associated with the current detection and retained candidates define the objects evaluated in every branch. Available annotations for these objects define the common observable object–frame mask. Within this mask, missing or incorrect predictions receive zero correctness; missing annotations exclude the pair from every branch. The same principle applies to appearance features and life-cycle events: only commonly available features and resolved events enter their comparisons. If the evidence needed for a verification channel remains unresolved, that channel is excluded for every candidate.

A.2 Primitive Scores and Penalties

The verification channels summarize geometry, identity continuity, interactions, and life-cycle outcomes. Each channel m combines a vector $\mathbf{C}_t(a)$ of five primitive scores with a vector $\mathbf{P}_t(a)$ of four error penalties:

$$G_t^m(a) = \langle \boldsymbol{\alpha}_m, \mathbf{C}_t(a) \rangle - \langle \boldsymbol{\beta}_m, \mathbf{P}_t(a) \rangle. \tag{A.1}$$

The coefficient vectors $\boldsymbol{\alpha}_m$ and $\boldsymbol{\beta}_m$ determine the contribution of each score and penalty. All five primitive scores lie between zero and one. Identity correctness is the fraction of common observable object–frame pairs assigned the fixed bound identity. Motion continuity evaluates the boxes predicted for that identity using overlap, center accuracy, and adjacent-frame displacement. These boxes retain their identity binding throughout the window and are not reassigned framewise to maximize GT overlap. For predicted and GT boxes b_s and $\widehat{b}_s$ at frame s, let u_s and $\widehat{u}_s$ be their centers and D_s the positive GT box diagonal. The normalized center and displacement errors are

$$e_s^{\mathrm{c}} = \frac{\|u_s - \widehat{u}_s\|_2}{D_s}, \qquad e_s^{\mathrm{v}} = \frac{\|(u_s - u_{s-1}) - (\widehat{u}_s - \widehat{u}_{s-1})\|_2}{D_s}. \tag{A.2}$$

These errors define motion continuity, averaged over common adjacent frame pairs with observable GT:

$$C_{\mathrm{motion}} = \mathrm{mean}_s \, \frac{\mathrm{IoU}(b_s, \widehat{b}_s) + \exp(-e_s^{\mathrm{c}}) + \exp(-e_s^{\mathrm{v}})}{3}. \tag{A.3}$$

Each motion component receives zero when its required prediction is missing. Overlap and center accuracy require the current predicted box, whereas displacement requires boxes at both adjacent frames. Missing predictions leave the common GT denominator unchanged. The motion score is unresolved when no adjacent GT frame pair is observable.

Appearance agreement compares the historical track representation f^{track} with the detection representation f^{det} assigned by the branch. Their nonnegative cosine similarities are averaged over the common available evidence under the track's fixed GT binding:

$$C_{\mathrm{app}} = \mathrm{mean} \max\left(0, \frac{\langle f^{\mathrm{track}}, f^{\mathrm{det}} \rangle}{\|f^{\mathrm{track}}\|_2 \|f^{\mathrm{det}}\|_2}\right). \tag{A.4}$$

The cosine similarity requires nonzero representations. A lost track receives zero when appearance evidence is observable; unavailable features exclude the affected evidence from every branch. Interaction correctness restricts identity correctness to frames where an evaluated GT box overlaps another GT box with IoU above 0.1. The interaction score is unavailable when the window contains no such frames.

Life-cycle correctness measures the fraction of common resolved opportunities assigned the correct track status. A birth is correct at an observable first appearance with no existing active or lost identity.

Once an identity exists, a visible object or one that reappears within the window requires that identity to remain recoverable. Creating a duplicate is a false-new event, discarding a needed identity is premature termination, and maintaining multiple active identities for the same object is an invalid keep. Failure to reappear within the window alone does not establish that termination is correct. Such unresolved opportunities are excluded from every branch. Invalid keeps reduce life-cycle correctness, while false-new and premature-termination events also incur the penalties defined below.

The four binary penalties indicate an identity switch, fragmentation, a false-new event, or premature termination. Fragmentation denotes an interruption followed by resumption of an observable GT trajectory. The short channel combines motion continuity and identity correctness with weights 0.80 and 0.20 and a switch penalty of 0.40. The long channel combines identity correctness and appearance agreement with weights 0.90 and 0.20 and switch and fragmentation penalties of 0.80 and 0.45. The interaction channel uses weights 0.75 and 0.20 for interaction and identity correctness, with a switch penalty of 0.70. The life-cycle channel uses weights 0.85 and 0.15 for life-cycle and identity correctness, with penalties of 0.80 for both false-new events and premature termination. All other coefficients are zero. The names short and long distinguish geometric and identity emphasis; both channels use the same 15-frame continuation.

A.3 Channel Statistics and Reliability

Within each channel, scores are compared over the common candidate set $\mathcal{B}_t$. The collection policy's action distribution is renormalized over this set to obtain weights $\bar{\pi}_t$. These weights define the policy-weighted mean score $\bar{G}_t^m$:

$$\bar{\pi}_t(a) = \frac{\pi_{\bar{\theta}_e}(a \mid c_t)}{\sum_{b \in \mathcal{B}_t} \pi_{\bar{\theta}_e}(b \mid c_t)}, \qquad \bar{G}_t^m = \sum_{a \in \mathcal{B}_t} \bar{\pi}_t(a) G_t^m(a). \tag{A.5}$$

Score variation is measured with uniform candidate weights. Let $\overline{G}_{t,\mathrm{u}}^m$ be the uniform mean over $\mathcal{B}_t$; the corresponding population variance is

$$(\sigma_t^m)^2 = \frac{1}{|\mathcal{B}_t|} \sum_{a \in \mathcal{B}_t} \left(G_t^m(a) - \overline{G}_{t,\mathrm{u}}^m\right)^2. \tag{A.6}$$

Subtracting the policy-weighted mean and scaling by this standard deviation gives the channel preference

$$A_t^m(a) = \frac{G_t^m(a) - \bar{G}_t^m}{\sigma_t^m + \epsilon_s}, \qquad a \in \mathcal{B}_t. \tag{A.7}$$

The stabilizer ϵ_s is 10^{-6}. A channel's reliability depends on the amount of observable evidence and the separation between candidate scores. With u_t^m denoting the common observable fraction of relevant frames or life-cycle opportunities, reliability is

$$\kappa_t^m = u_t^m \min\left(1, \frac{\max_{a \in \mathcal{B}_t} G_t^m(a) - \min_{a \in \mathcal{B}_t} G_t^m(a)}{\tau_G}\right). \tag{A.8}$$

The range scale τ_G is 0.1, so a constant-score channel has zero reliability. The weights w_t^m in Eq. 2 are uniform over available channels relevant to the event and sum to one; absent channels have zero weight. Their reliability-weighted sum gives aggregate confidence χ_t between zero and one. A record is discarded if fewer than two candidates remain or its aggregate confidence is zero.

B Target Construction and Mathematical Properties

B.1 Support Selection

The comparison support contains the GT reference and uniquely aligned competitors with lower future identity preferences. The reference must already belong to the evaluated legal candidates, and each competitor's preference must fall below it by more than the preference-gap threshold. Up to four competitors are retained in descending collection-policy probability. A record is eligible only if its confidence is at least 0.50 and its support contains at least one competitor. For such a support, a softmax over its preferences defines the normalized entropy used for selection:

$$s_i = \mathrm{softmax}(V_i|_{S_i}), \qquad \mathcal{H}_i = -\frac{\sum_{k \in S_i} s_i(k) \log s_i(k)}{\log |S_i|} \le h_{\max}. \tag{B.1}$$

The softmax temperature is one, the entropy ceiling $h_{\max}$ is 0.80, and the preference-gap threshold is 0.03. These conditions retain comparisons whose preference distribution is sufficiently resolved for teacher construction.

B.2 Step Selection and Numerical Evaluation

The correction step depends on how strongly the collection policy supports the GT reference. We quantify this posterior maturity using the collection distribution p_i^{col} over all K_i legal classes, relative to a uniform distribution:

$$m_i = \mathrm{clip}\left(\frac{p_i^{\mathrm{col}}(y_i) - 1/K_i}{1 - 1/K_i}, 0, 1\right), \qquad K_i > 1. \tag{B.2}$$

Here the collection probability is evaluated at the semantic reference corresponding to current label y_i, even if its earlier column index differs. Posterior maturity m_i and record confidence χ_i, inherited from χ_t, remain fixed within an epoch. Together with the preference span D_i^V, they determine the initial step:

$$D_i^V = \max_{k \in S_i} V_i(k) - \min_{k \in S_i} V_i(k), \qquad \eta_i^{(0)} = \chi_i m_i \,\mathrm{clip}(D_i^V, 0, 1). \tag{B.3}$$

The initial step is accepted when it satisfies the TV bound of 0.12 in Eq. 8. Otherwise, 20 bisection iterations between zero and the initial step select the feasible lower endpoint. To evaluate the exponential weighting stably, we center preferences using the detached conditional distribution:

$$\widetilde{V}_i(k) = V_i(k) - \sum_{j \in S_i} r_i(j) V_i(j), \qquad \widehat{r}_i = \mathrm{softmax}(\log r_i + \eta_i \widetilde{V}_i). \tag{B.4}$$

The conditional log probabilities are obtainable by applying log-softmax directly to the support logits, avoiding a logarithm of an underflowed probability. The resulting teacher is detached after construction. A numerically zero support mass, a single legal class, an unresolved reference, or a zero accepted step gives no auxiliary update.

B.3 A Constrained Teacher Distribution

The local teacher retains the current probabilities outside the comparison support. This constraint defines a feasible set $\mathcal{Q}_i$ within the probability simplex Δ_i over current legal association classes:

$$\mathcal{Q}_i = \{u \in \Delta_i : u|_{S_i^c} = p_i^0|_{S_i^c}\}. \tag{B.5}$$

Because every distribution in $\mathcal{Q}_i$ agrees outside the support, each assigns total probability M_i to it. For any fixed nonnegative step, the teacher in Eq. 7 solves

$$q_i = \arg\max_{u \in \mathcal{Q}_i} F_i(u), \qquad F_i(u) = \eta_i \sum_{k \in S_i} u(k) V_i(k) - \mathrm{KL}(u \| p_i^0). \tag{B.6}$$

The objective rewards probability assigned to higher-preference actions while penalizing departure from the detached current distribution. Finite legal logits make the negative KL term strictly concave on the feasible set. Substituting the teacher's exponential weighting gives the objective gap

$$F_i(q_i) - F_i(u) = \mathrm{KL}(u \| q_i) \geq 0, \qquad u \in \mathcal{Q}_i. \tag{B.7}$$

Expanding both KL terms yields this identity: the log-normalizer has coefficient M_i for every feasible distribution and cancels in the difference. The gap is zero precisely when u equals q_i, establishing the unique solution under the constraint. The TV search then selects a feasible step within this family of teachers; it does not optimize a general tracking objective under a TV constraint.

B.4 Normalization, Odds, and the TV Search

With finite legal logits and positive support mass, the refined conditional probabilities $\widehat{r}_i$ are positive and sum to one. Rescaling them by the support mass and retaining outside probabilities in Eq. 7 gives

$$\sum_k q_i(k) = M_i \sum_{k \in S_i} \widehat{r}_i(k) + \sum_{k \notin S_i} p_i^0(k) = M_i + (1 - M_i) = 1. \tag{B.8}$$

The unchanged outside probabilities also reduce TV to the conditional difference in Eq. 8. Within the support, dividing two teacher probabilities cancels both the support mass and the normalizer:

$$\frac{q_i(k)}{q_i(j)} = \frac{p_i^0(k)}{p_i^0(j)} \exp\big(\eta_i(V_i(k) - V_i(j))\big), \qquad k, j \in S_i. \tag{B.9}$$

The resulting ratio establishes Eq. 10 and depends on preference differences, so a shared centering offset leaves the teacher unchanged. For a positive step, the GT reference gains conditional probability because its preference is strictly higher than every competitor's. Competitors can exchange probability among themselves, while the reference's odds increase against each of them.

For fixed preferences and current conditional distribution, TV is continuous and nondecreasing as the nonnegative step increases. To show this, omit the token index and write $Z(\eta)$ for the exponential normalizer. Let U_η contain the actions whose refined probability exceeds their current probability:

$$Z(\eta) = \sum_k r(k) e^{\eta V(k)}, \qquad U_\eta = \{k : \eta V(k) > \log Z(\eta)\}. \tag{B.10}$$

Membership in U_η selects preferences above a common threshold. On intervals where this set is fixed, differentiating the sum of positive probability differences gives

$$\frac{\mathrm{d}}{\mathrm{d}\eta} \mathrm{TV}(\widehat{r}_\eta, r) = \mathrm{Cov}_{\widehat{r}_\eta}\big(\mathbf{1}_{U_\eta}, V\big) \geq 0. \tag{B.11}$$

The covariance is nonnegative because the indicator selects higher-preference actions. When membership changes, the entering or leaving action has zero probability difference, preserving continuity. Since the zero step is feasible, bisection with a feasible lower endpoint enforces the TV bound.

B.5 Auxiliary Losses and Normalization

The auxiliary objective combines KL distillation with two losses on the teacher's pairwise preferences. The soft pairwise probability loss $\mathcal{L}_{\mathrm{rank}}$ and bounded log-odds loss $\mathcal{L}_{\mathrm{value}}$ supplement $\mathcal{L}_{\mathrm{dist}}$ as

$$\mathcal{L}_{\mathrm{aux}} = \lambda_{\mathrm{aux}}\big(\mathcal{L}_{\mathrm{dist}} + \lambda_{\mathrm{rank}}\mathcal{L}_{\mathrm{rank}} + \lambda_{\mathrm{value}}\mathcal{L}_{\mathrm{value}}\big). \tag{B.12}$$

Both pairwise losses use the same set $\mathcal{P}_i$ of ordered class pairs within the comparison support. A pair is retained when its preference gap exceeds the threshold τ_V of 0.03:

$$\mathcal{P}_i = \{(k, j) \in S_i \times S_i : V_i(k) - V_i(j) > \tau_V\}. \tag{B.13}$$

For each retained pair, the current log-odds d_{ikj} and detached teacher log-odds d^*_{ikj} are

$$d_{ikj} = z_i(k) - z_i(j), \qquad d^*_{ikj} = \mathrm{stopgrad}\big[d_{ikj} + \eta_i\big(V_i(k) - V_i(j)\big)\big]. \tag{B.14}$$

Applying the sigmoid to the teacher log-odds gives its pairwise probability target:

$$s_{ikj} = \mathrm{sigmoid}(d^*_{ikj}). \tag{B.15}$$

The soft pairwise probability loss matches this target through binary cross-entropy:

$$\ell_{\mathrm{rank},i} = \frac{1}{|\mathcal{P}_i|} \sum_{(k,j)\in\mathcal{P}_i} \left[s_{ikj}\,\mathrm{softplus}(-d_{ikj}) + (1 - s_{ikj})\,\mathrm{softplus}(d_{ikj})\right]. \tag{B.16}$$

The bounded log-odds loss matches a hyperbolic-tangent transform of the same teacher log-odds:

$$\ell_{\mathrm{value},i} = \frac{1}{|\mathcal{P}_i|} \sum_{(k,j)\in\mathcal{P}_i} \mathrm{Huber}_1\big(\tanh d_{ikj} - \tanh d^*_{ikj}\big). \tag{B.17}$$

This loss uses the unit-threshold Huber function for a scalar residual v:

$$\mathrm{Huber}_1(v) = \tfrac{1}{2}\min(|v|, 1)^2 + \max(|v| - 1, 0). \tag{B.18}$$

An empty pair set contributes zero to either pairwise loss. The token losses are then aggregated over the eligible token set E and normalized by the total number N_B of valid association tokens in the native objective:

$$\mathcal{L}_r = \frac{1}{N_B} \sum_{i\in E} \ell_{r,i}, \qquad r \in \{\mathrm{dist}, \mathrm{rank}, \mathrm{value}\}. \tag{B.19}$$

A batch with no valid association tokens contributes no auxiliary loss. The weights λ_{aux}, λ_{rank}, and λ_{value} are 0.10, 0.50, and 0.25, respectively. Teacher quantities are detached in all three losses. At zero step, the teacher equals the current distribution, giving zero gradient at the current logits for every term. The soft pairwise probability loss can still have a nonzero entropy value, even though its gradient vanishes.

B.6 Local Gradient Constraint

All tracker adaptations apply the same local constraint at a differentiable association tensor u_i shared by the native and auxiliary objectives. MOTIP uses its association logits. For token i, the gradients g_i^{b} and g_i^{a} and their inner product c_i determine the projection coefficient γ_i:

$$g_i^{\mathrm{b}} = \nabla_{u_i}\mathcal{L}_{\mathrm{base}}, \qquad g_i^{\mathrm{a}} = \nabla_{u_i}\mathcal{L}_{\mathrm{aux}}, \qquad c_i = \langle g_i^{\mathrm{a}}, g_i^{\mathrm{b}}\rangle, \tag{B.20}$$

$$\gamma_i = \min\left(0, \frac{c_i}{\|g_i^{\mathrm{b}}\|_2^2 + \epsilon_g}\right). \tag{B.21}$$

This coefficient removes the opposing gradient component up to regularization. The resulting vector is then scaled to satisfy a norm cap:

$$g_i^{\perp} = g_i^{\mathrm{a}} - \gamma_i g_i^{\mathrm{b}}, \qquad \widetilde{g}_i^{\mathrm{a}} = \min\left(1, \frac{\rho\|g_i^{\mathrm{b}}\|_2}{\|g_i^{\perp}\|_2 + \epsilon_g}\right) g_i^{\perp}. \tag{B.22}$$

The norm ratio ρ is 0.25, and the stabilizer ϵ_g is 10^{-12}. When the native squared gradient norm is at most 10^{-12}, the auxiliary gradient for that token is set to zero. For the remaining tokens, Eq. B.22 gives

$$\|\widetilde{g}_i^{\mathrm{a}}\|_2 \le \rho\|g_i^{\mathrm{b}}\|_2. \tag{B.23}$$

When the two gradients oppose each other, regularization leaves the following inner product before norm scaling:

$$\langle g_i^{\perp}, g_i^{\mathrm{b}}\rangle = \frac{c_i\epsilon_g}{\|g_i^{\mathrm{b}}\|_2^2 + \epsilon_g}, \qquad c_i < 0. \tag{B.24}$$

This residual makes conflict removal approximate. The norm bound applies at the selected tensor u_i: propagation through shared parameters and summation across tokens can change gradient angles and norm ratios. For eligible tokens E, the parameter update retains the native gradient and propagates each constrained auxiliary vector through the corresponding Jacobian:

$$g_\theta = \nabla_\theta\mathcal{L}_{\mathrm{base}} + \sum_{i\in E}\left(\frac{\partial u_i}{\partial\theta}\right)^{\top}\mathrm{stopgrad}[\widetilde{g}_i^{\mathrm{a}}]. \tag{B.25}$$

The optimizer uses this combined vector for one update. An additional unconstrained auxiliary backward pass would change the specified update and is therefore excluded.

C Collection, Replay, and Experimental Configuration

C.1 Collection and State Replay

State replay restores the context in which a candidate action was evaluated. The pre-action context c_t and the state $s_{t^+}^{(a)}$ produced by a legal action a are

$$c_t = (x_t, \mathcal{T}_t, Q_t, M_t, L_t). \tag{C.1}$$

$$s_{t^+}^{(a)} = T(c_t, a), \qquad a \in \mathcal{A}(c_t). \tag{C.2}$$

Here x_t is the current observation, $\mathcal{T}_t$ the active identity set, Q_t the learned representations, M_t the motion and appearance history, and L_t the life-cycle state. The transition T applies an action from the legal set $\mathcal{A}(c_t)$ before the next observation becomes available during continuation.

The collection policy selects a decision when its second-largest legal action probability is at least half the largest, or when an association competes with a life-cycle action. For this decision, the candidate set retains the executed action and at most four alternatives whose probability is at least one tenth of the largest. Candidates with collisions, duplicate state effects, or identities or transitions unsupported by the native state machine are removed. A lost identity remains eligible when the tracker supports reconnection. Each retained candidate acts on a complete copy of the pre-action state, including hidden features, query order, identity maps, memories, motion state, life-cycle counters, and the identity allocator.

The collected record stores this pre-action state together with its sequence, frame, detection, query or track identity, action type, and collection snapshot. Replay restores the state with the same preprocessing and augmentation to preserve the context of the comparison. It keeps saved history detached while recomputing the trainable association outputs. A record contributes no auxiliary loss if its state cannot be reproduced or its action mapping is ambiguous.

C.2 Tracker-Specific Alignment

Replay must attach each action's stored preference to the corresponding current association class. A unique semantic mapping ϕ_i identifies the current column k for action a and assigns its preference as

$$k = \phi_i(a), \qquad V_i(k) = A_t^{\mathrm{trk}}(a). \tag{C.3}$$

MOTIP reconstructs this mapping from its saved identity tokens, history, batch-local vocabulary, and life-cycle state. Each semantic identity is assigned to the corresponding current row and legal label, including supported special labels for a new identity and background. MeMOTR reconstructs its identity correspondence from track memories, query representations, identity maps, and life-cycle state. SambaMOTR uses its synchronized hidden state, track representations, identity maps, and life-cycle state.

For the MeMOTR and SambaMOTR query adaptations, the stored track representation h_i defines the association to be trained. Its cosine similarity with each aligned detection or query representation e_j gives the candidate score

$$z_{ij}^{\mathrm{query}} = \frac{\langle h_i, e_j \rangle}{\|h_i\|_2 \|e_j\|_2}. \tag{C.4}$$

These scores require nonzero representations. A softmax over legal candidates converts them into an auxiliary association distribution. Its GT reference must correspond to a match or reconnection with both representations available. Other life-cycle events affect verification but receive no direct loss unless they have a differentiable correspondence. The query adaptations construct this distribution from existing representations and retain the original inference computation without adding a classifier.

All three trackers retain native objectives and continuation training while sharing verification, alignment, teacher construction, auxiliary losses, and the gradient rule in Appendix B.6. Both gradients are evaluated at the shared tensor u_i. For query trackers, the auxiliary loss is differentiated from the cosine scores back to this tensor, while the native loss follows its original graph. The native loss need not take the cosine scores as inputs. Saved history remains detached, and the constrained correction propagates through the recomputed association path. The norm bound holds at u_i; the teacher's probability and log-odds properties hold at its scores.

C.3 Optimization and Reproduction

The controlled MOTIP study starts from a pretrained tracker and updates its identity decoder, identity embeddings, and output layer, with the detection backbone and detection branch frozen. These parameters are optimized with AdamW using momentum coefficients 0.9 and 0.999, a numerical stabilizer of 10^{-8}, and weight decay of 10^{-4} except on biases and normalization parameters. The learning rate increases linearly from 10^{-6} to 10^{-5} over 100 updates and then remains constant. Each effective batch contains 16 clips of eight frames at stride one. Future verification uses a separate 15-frame window, and all compared groups retain the full native loss.

Data preparation, including images, detections, annotations, dataset conversion, and augmentation, follows the corresponding tracker. Collection saves the resulting states and transformations for replay. Tracker computations use mixed precision, while future verification and gradient projection use unscaled FP32. The experiments use eight A100-SXM4 GPUs with 80 GB each, Python 3.10, PyTorch 2.9.0 with CUDA 12.8, torchvision 0.24.0, NumPy 2.1.2, SciPy 1.15.3, and Accelerate 1.14.0.

Method coefficients were selected on DanceTrack validation before evaluating the remaining trackers and datasets. The explored values were 0.25, 0.30, and 0.50 for the soft pairwise probability loss weight, 0.03 and 0.05 for the preference-gap threshold, and 0.10 and 0.12 for the TV bound. HOTA, with AssA and IDF1 as secondary measures, selected the final values of 0.50, 0.03, and 0.12. The horizon, confidence and entropy thresholds, support size, and norm bound remained fixed. Because some trials also changed target construction, these ranges document development choices rather than a factorial sensitivity study.

Reproducing a configuration requires its starting checkpoint, data split, transformations, collection snapshots, saved states, training trace, final checkpoint, and evaluator output. Restoring the collection records preserves the verified supervision, whereas recollection can visit different states. The tracker-specific native training procedure is retained, with Track-OPD supplying the additional association

supervision described above. The planned release includes the tracker adaptations, configurations, environment specification, and evaluation scripts.

D Experimental Controls and Additional Context

D.1 Matched Preference Replacement

The GT preference control uses the same comparison support as Track-OPD and replaces its aligned future identity preferences with fixed values:

$$V_i^{\mathrm{GT}}(y_i) = 1, \qquad V_i^{\mathrm{GT}}(j) = 0 \quad \text{for } j \in \mathcal{N}_i. \tag{D.1}$$

The control shares each epoch's Track-OPD records, including saved states, valid positions, comparison support, confidence, and posterior maturity. It also retains the teacher construction, step rule, losses, gradient constraints, and final 4,000-update checkpoint rule. The replacement fixes every reference–competitor preference gap at one and makes all competitors equal in preference. These values enter both exponential weighting and step selection. Because both groups use the same future-selected evidence, their comparison measures the joint effect of these changes on common records; it does not isolate either as the sole source of the gain.

D.2 Aggregation and Diagnostic Protocols

Table 2 reports arithmetic means over three runs per method. The same DanceTrack Track-OPD runs supply the mean and standard deviation in Table 7, which also reports three runs on SportsMOT. These standard deviations describe variation in Track-OPD performance, not uncertainty in its differences from Continue or the GT preference control. The adapted official test results in Table 1 use five-run medians from separate runs. Within each comparison, methods share the dataset protocol, evaluation settings, and non-target training conditions.

Table 7. Track-OPD validation mean and standard deviation over three runs per dataset. The DanceTrack runs are the same as in Table 2.

Metric	DanceTrack	SportsMOT
HOTA ↑	66.016 ± 0.581	76.618 ± 0.171
AssA ↑	57.351 ± 0.986	68.747 ± 0.302
IDF1 ↑	71.395 ± 0.758	82.348 ± 0.184

The introduction-epoch study applies Track-OPD at epoch 0, 2, 4, or 8 during training from scratch. The corresponding HOTA values are 28.904, 37.675, 43.751, and 46.139, compared with 43.282 for ordinary training. The motion-prediction study uses its own paired checkpoints on the same 25,508 validation frames. Each study fixes its comparison protocol, with separate checkpoints across studies. Reported update budgets count optimizer steps and exclude collection, which continues up to five branches for 15 frames per selected decision. Wall-clock time and GPU-hour measurements are unavailable.

D.3 Additional Target and Update Comparisons

Table 8 compares alternative verification, weighting, and supervision configurations. Full Track-OPD achieves the highest HOTA, DetA, AssA, IDF1, and MOTA, supporting the combined target design for association quality. The variant without future verification has fewer identity switches, whereas the full method achieves higher AssA and IDF1.

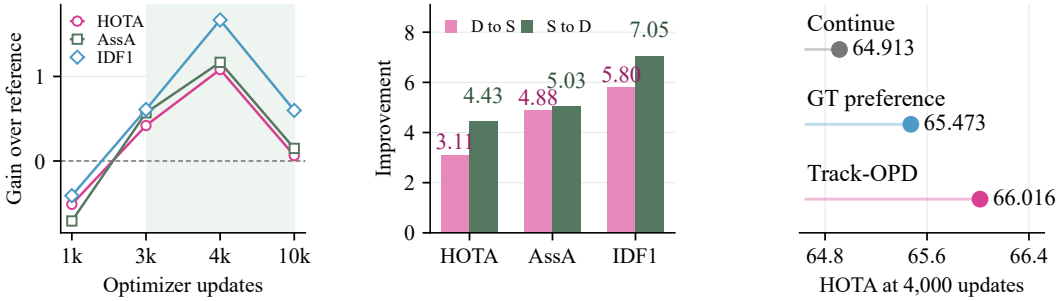
Tables 8 and 3 evaluate complete configurations under the shared dataset and comparison settings. Their variant labels identify the tested configurations, including Vanilla OPD and OPSD. The detailed historical removal operations and checkpoint specifications are not fully documented, so the metric gaps support the combined design rather than quantify independent component effects. The label "w/o contrast" leaves the changed operation unspecified. Subtracting a shared centering offset alone leaves the teacher unchanged. The matched 4k study compares preference values on identical records with a fixed update budget.

Table 8. Target configurations on DanceTrack validation. Appendix D.3 specifies the comparison scope.

Target design	HOTA ↑	DetA ↑	AssA ↑	IDF1 ↑	MOTA ↑	IDSW ↓
w/o future verifier	64.934	75.402	56.184	69.727	85.625	**2706**
w/o contrast	64.332	75.457	55.108	68.401	84.690	4615
w/o dynamic weight	64.162	75.476	54.799	68.672	84.774	4423
Hard-label OPD	64.713	75.485	55.732	69.724	85.538	2830
Hard-label only	64.619	75.580	55.493	69.196	85.572	2754
Vanilla OPD	64.354	75.554	55.077	68.721	85.561	2853
OPSD	64.050	75.713	54.443	68.449	85.561	2803
Track-OPD	**66.016**	**76.364**	**57.351**	**71.395**	**86.448**	2740

D.4 Additional Diagnostic Visualizations

Figure 6 visualizes Tables 4, 5, and 2: the intermediate-budget peak, the joint improvements in HOTA, AssA, and IDF1 under bidirectional transfer, and the gain over GT preferences on common records.



(a) Update trajectory **(b) Bidirectional transfer** **(c) Matched target control**

Figure 6. Additional validation diagnostics: (a) gains over 10k Continue across update budgets; (b) cross-domain gains over Continue; (c) the matched 4k comparison. D and S denote DanceTrack and SportsMOT.

D.5 Additional Related-Work Context

Tracking-by-detection combines motion, appearance, and matching rules (Bewley et al., 2016; Wojke et al., 2017; Bergmann et al., 2019; Cao et al., 2023), while learned representations improve detection and association (Zhang et al., 2021b; Shao et al., 2026; Kim & Yoon, 2026). Transformer trackers carry such representations across frames through queries or memories (Meinhardt et al., 2022; Cai et al., 2022; Yan et al., 2025; Wang et al., 2025). Training on model-generated histories addresses the training–inference mismatch in sequence prediction through scheduled sampling (Bengio et al., 2015) and in online 3D tracking through autoregressive training (Ding et al., 2023). Learned decision policies, structured losses, and distillation provide further ways to supervise tracking decisions (Xiang et al., 2015; Xu et al., 2020; Zhang et al., 2021a).

The source and form of this supervision also shape recent distillation methods, which vary supervision sources, perspectives, and self-distillation targets (Yu et al., 2026; Hou et al., 2026; Zhang et al., 2026; Dat et al., 2026). Outcome-conditioned credit methods use downstream effects to assess earlier actions (Mesnard et al., 2021; Meulemans et al., 2023). Track-OPD studies how such effects can supervise recurrent tracking when alternative actions change the identities and predictions being evaluated. Its fixed identity bindings and semantic alignment preserve the correspondence between future identity preferences and the current association decision.