EMBRACING AMBIGUITY: HYPOTHESIS-GUIDED LINGUISTIC POLICY OPTIMIZATION FOR VIDEO ANOMALY REASONING

Anonymous authors
Paper under double-blind review

ABSTRACT

Video anomaly understanding requires models to identify anomalous events and explain why they are anomalous. Existing approaches improve recognition through informative observations or reusable linguistic guidance, yet a correct prediction may still rely on a criterion that misclassifies normal events. We introduce Anom-Ψ, a hypothesis-guided linguistic policy optimization framework that turns differences between competing interpretations into feedback for refining anomaly criteria while keeping model weights frozen. During training, the model interprets shared observations around temporal changes under normal and abnormal hypotheses, exposing differences in the evidence and contextual assumptions behind its judgments. A linguistic optimizer edits a hierarchical experience bank using reward-guided comparisons, keeping hypothesis-derived corrective references distinct from neutral-trajectory preferences. We derive a sufficient condition for these linguistic policy updates to improve expected reward under weak supervision. Across benchmarks, Anom-Ψ outperforms the compared tuning-free baselines, reaching $91.28 \pm 0.71\%$ Micro-AUC on MSAD with 20% of its training data. Ablations and repeated runs show that opposing-hypothesis comparisons complement neutral preferences, yielding more effective and consistent experience learning.

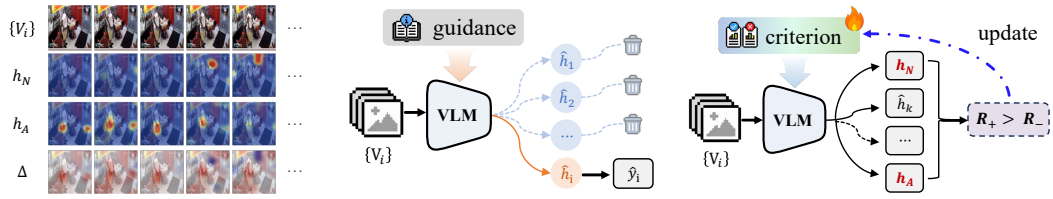
1 INTRODUCTION

Video anomaly reasoning requires identifying the evidence that makes an event anomalous in context. Different hypotheses can direct a model toward different aspects of the same scene. Consider the weapon-drawing event in Figure 1a. Under a normal hypothesis, the model emphasizes the room background, whereas an abnormal hypothesis brings the handgun at the person's waist into focus. The visual observations remain identical, but the hypotheses highlight different evidence. This contrast exposes decisive evidence overlooked under the normal hypothesis. Learning from what each hypothesis uses or overlooks can refine reusable conditions for distinguishing anomalous events from normal activity.

Progress in video anomaly analysis has strengthened temporal and semantic representations (Tian et al., 2021; Wu et al., 2024) and enabled more explicit event interpretations (Zhang et al., 2025a; Huang et al., 2025a). Recent approaches further optimize evidence acquisition (Mo et al., 2026) or adapt reusable linguistic guidance (Ye et al., 2025; Leng et al., 2026). These advances help resolve ambiguity into a final judgment (Figure 1b). Resolving competing hypotheses, however, does not by itself exploit their differences for learning. Reward-ranked reasoning trajectories may still share an interpretive premise, leaving the evidence and assumptions behind it unchallenged. A correct prediction can also reinforce a cue without establishing the conditions under which it is diagnostic. We therefore ask *how competing hypotheses can turn ambiguity into feedback for learning contextual anomaly criteria*.

We propose Anom-Ψ, a hypothesis-guided linguistic policy optimization framework that learns from differences between competing interpretations (Figure 1c). Change-aware sampling prioritizes observations around candidate behavioral transitions while retaining salient anomaly evidence, providing temporal context for examining when a cue becomes diagnostic. Under explicit normal and

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.



(a) Ambiguity in VAU (b) Existing Methods (c) Anom-Ψ (Ours)

Figure 1: Learning the contextual scope of anomaly criteria. (a) Ambiguity: shared behavioral cues admit normal and abnormal interpretations. (b) Existing: outcome-level feedback can leave the conditions for applying a criterion unresolved. (c) Ours: competing hypotheses on shared observations guide experience-bank edits that refine when a cue supports an anomaly judgment.

abnormal hypotheses, the learner generates reasoning trajectories from identical observations and a shared experience bank. The hypotheses change the interpretive premise, making different uses of the same evidence available for comparison even when neutral sampling repeatedly reaches the same decision. Video-level supervision and temporal reward feedback identify useful contrasts between these trajectories, exposing evidence, conditions, and exceptions missing from a judgment.

Hypothesis-derived corrective references remain distinct from reward-ranked neutral trajectories, which provide complementary within-policy preferences. Building on feedback-driven experience adaptation (Ye et al., 2025; Leng et al., 2026), a linguistic optimizer integrates both signals into edits of a hierarchical experience bank. Principles, Patterns, and Clues retain the resulting conditions and evidence at different levels of abstraction. The learned bank guides subsequent neutral reasoning with frozen model parameters, carrying the benefit of hypothesis comparisons into inference. Our theoretical analysis connects linguistic edits to local policy changes and derives a sufficient condition for expected-reward improvement under weakly supervised feedback.

Across benchmarks, Anom-Ψ improves detection with limited adaptation data and retains its benefits in cross-dataset transfer (Tables 1 and 12). Under a matched generation budget, the full hypothesis-guided update improves XD-Violence AP by 3.51 percentage points over its hypothesis-free control (Table 4(c)). This gain supports using explicit contrasts between interpretations as feedback for experience learning. The learned entries and their use in case studies further illustrate how such feedback can be retained as contextual conditions that distinguish visually similar events (Appendix D).

Our main contributions are threefold. (1) We identify differences between competing hypotheses as informative feedback for learning the evidence and contextual conditions supporting anomaly judgments. (2) We propose Anom-Ψ, which combines change-aware observation selection with opposing hypotheses, keeps hypothesis-derived corrective references distinct from neutral-trajectory preferences, and integrates both into a hierarchical experience bank without updating model parameters. (3) We connect linguistic policy adaptation, group-relative comparisons, and hypothesis-driven decision coverage, and derive a sufficient condition for expected-reward improvement.

2 Related Work

Tuning-based Video Anomaly Understanding. Weakly supervised detection learns temporal representations and anomaly scores from video-level labels, increasingly incorporating vision–language semantics and explanations (Tian et al., 2021; Wu et al., 2024; Huang et al., 2025a). Holmes-VAU (Zhang et al., 2025a) combines anomaly-oriented temporal sampling with hierarchical instruction data for queries at multiple temporal granularities. Vad-R1 (Huang et al., 2025b) combines perception-to-cognition supervision with self-verification-based reinforcement learning, while Anom-π (Mo et al., 2026) optimizes interleaved reasoning, evidence acquisition, and stopping through trajectory preferences under video-level supervision. These developments extend learned anomaly representations to reasoning procedures and observation policies.

Tuning-free Video Anomaly Understanding. Structured inference combines visual descriptions with temporal aggregation, event-level prompting, or hierarchical reasoning (Zanella et al., 2024; Shao et al., 2025; Li et al., 2025). AnomalyRuler (Yang et al., 2024) derives reusable detection rules from normal references. Feedback-driven language optimization learns textual guidance with frozen

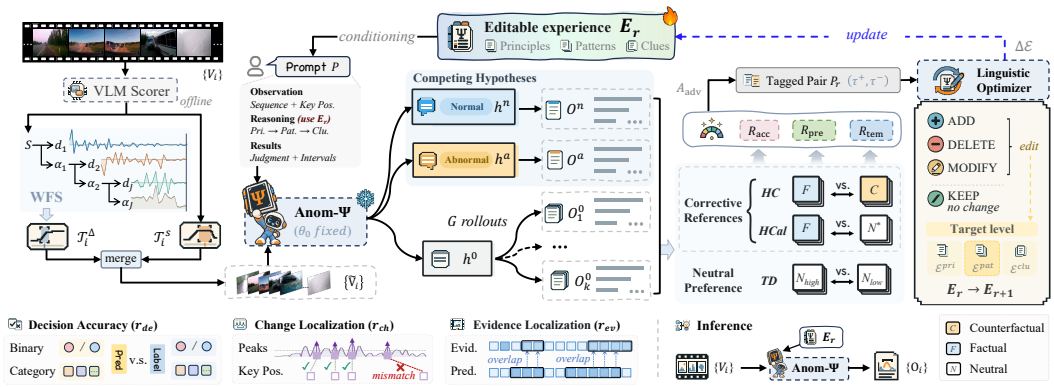


Figure 2: Overview of Anom-Ψ. Shared frames support normal, abnormal, and neutral reasoning; reward-guided comparisons drive hierarchical experience updates. The learned bank guides inference with frozen model weights.

model weights (Pryzant et al., 2023; Cai et al., 2025). VERA (Ye et al., 2025) refines guiding questions through verbal feedback, while LRPO (Leng et al., 2026) distills reward-ranked trajectories into general experience and adds scenario-specific experience for contextual adaptation. Anom-Ψ turns ambiguity into learning feedback by contrasting normal and abnormal hypotheses over shared evidence to refine contextual anomaly criteria.

3 Method

Anom-Ψ learns contextual anomaly criteria by editing a textual experience bank while keeping model weights frozen. Figure 2 traces this process from shared observations through competing interpretations and reward-guided comparisons to experience edits. These comparisons expose differences in the evidence and conditions supporting each judgment, which the optimizer turns into reusable guidance for subsequent videos.

3.1 Problem Formulation

Given a video $V = \{I_t\}_{t=1}^{T}$, video anomaly reasoning aims to predict frame-level anomaly scores $s_t \in [0,1]$, localize anomalous intervals, and explain the judgments. We consider a weakly supervised training set $\mathcal{D} = \{(V_v, Y_v, C_v)\}_{v=1}^{N}$, where $Y_v \in \{0,1\}$ indicates the presence of anomalies and C_v denotes the anomaly-category set. Frame-level annotations are unavailable during training. For sampled observations and a task prompt $x = (\widetilde{V}, P)$, a frozen VLM with parameters θ_0 generates a reasoning trajectory $o \sim \pi_{\mathcal{E}}(\cdot \mid x)$ conditioned on a hierarchical experience bank $\mathcal{E}$. Following LRPO (Leng et al., 2026), we use the bank's full, hierarchy-ordered text as an editable policy representation. With a fixed distribution ν over $z = (x, Y, C)$ and reward evaluator $R(z, o)$, our objective is to improve the expected reward $J(\mathcal{E}) = \mathbb{E}_{z\sim\nu}\mathbb{E}_{o\sim\pi_{\mathcal{E}}(\cdot|x)}[R(z,o)]$ by refining this context. Training observations and the reward evaluator remain fixed throughout experience optimization, so updates act on the learner's interpretation of the same evidence.

3.2 Shared Change-Aware Observations

A salient cue can be ambiguous without the event context that makes it diagnostic. We therefore combine observations around candidate temporal changes with high-score anomaly evidence, as shown in Figure 2(a).

Change-aware video selection. Following the score-guided temporal sampling approach of Holmes-VAU (Zhang et al., 2025a), we obtain fixed offline VLM anomaly scores $u_j \in [0,1]$ for training videos. Wavelet responses to the transformed scores identify candidate transitions. Ranking videos by change richness within each category concentrates the learning budget on sequences with temporal variation while maintaining category coverage.

Hybrid frame selection. For each retained video $V_i = (I_{i,t})_{t=1}^{T_i}$, let $\mathcal{J}_i \subseteq \{1, \ldots, T_i\}$ be its candidate frame indices and $\mathbf{u}_i = (u_{i,t})_{t \in \mathcal{J}_i}$ the aligned VLM scores. We select change-aware and high-score index sets with respective budgets K_Δ and K_s,

$$\begin{aligned} \mathcal{T}_i^{\Delta} &= \mathrm{WFS}_{K_\Delta}(V_i, \mathbf{u}_i; \mathcal{J}_i), \\ \mathcal{T}_i^{s} &= \underset{t \in \mathcal{J}_i \setminus \mathcal{T}_i^{\Delta}}{\mathrm{TopK}_{K_s}} u_{i,t}, \end{aligned} \tag{1}$$

where wavelet-based frame selection (WFS) captures temporal context and TopK returns the highest-scoring remaining indices. Their chronological union retains both sources of evidence without duplicate frames,

$$\widetilde{V}_i = (I_{i,t})_{t \in \mathrm{Sort}\left(\mathcal{T}_i^{\Delta} \cup \mathcal{T}_i^{s}\right)}. \tag{2}$$

All branches receive these same frames in the same temporal order under the current bank. Their comparisons therefore concern interpretations of shared evidence. Sampling scores, allocation rules, and default budgets are specified in Appendix A.1 and Table 5.

3.3 Competing Hypotheses and Structured Trajectories

Neutral sampling can repeatedly return the same decision, leaving its underlying criterion unchallenged. We introduce opposing hypotheses to expose alternative interpretations (Appendix B.3).

At step r, let $\pi_r = \pi_{\mathcal{E}_r}$. We supplement G neutral trajectories $o_{r,i}^0$ with one abnormal-premise trajectory o_r^a and one normal-premise trajectory o_r^n, as shown in Figure 2(b). The abnormal branch must predict at least one anomaly interval and, when ground-truth categories are available, is restricted to that category set. The normal branch must predict no anomaly interval. When both outputs are valid, these constraints supply opposing decisions even if neutral sampling favors only one. Each trajectory records observations, key positions, reasoning, and category-tagged anomaly intervals, making the evidence behind these decisions available for comparison.

The video label assigns factual and counterfactual roles. We set $o_r^F = o_r^a$ and $o_r^C = o_r^n$ when $Y = 1$, and reverse the assignment when $Y = 0$. Thus F (factual) and C (counterfactual) denote label-consistent and label-inconsistent premises, while N denotes a neutral judgment. A factual role alone does not establish a preferred trajectory. Rewards evaluate the resulting decision and localization to identify useful corrective comparisons.

3.4 From Weak Rewards to Corrective Comparisons

Video-level correctness alone does not establish whether a judgment identifies the relevant temporal changes and supporting evidence. We therefore evaluate each valid trajectory with accuracy, change-localization, and evidence-localization rewards. Accuracy checks the anomaly decision and category prediction. Change localization matches key positions to wavelet peaks, while evidence localization measures overlap with positions ranked highly by the raw VLM scores. The active components $\mathcal{A}(o)$ form the normalized reward

$$R(z, o) = \frac{\sum_{m \in \mathcal{A}(o)} w_m R_m(z, o)}{\sum_{m \in \mathcal{A}(o)} w_m}. \tag{3}$$

This combines video-level outcome supervision with temporal proxies from the fixed VLM scores, providing feedback about evidence use without frame-level training annotations. Reward definitions, weights, and activation conditions appear in Appendix A.2.

Three complementary relations. Corrective references and neutral judgments provide different information about the criterion being learned. Figure 2(c) separates them into three relations. Hypothesis-contrast (HC) takes the factual trajectory selected by the binary ground-truth label as chosen and the opposing hypothesis trajectory as rejected. Hypothesis-calibrated (HCal) keeps the factual trajectory as chosen and selects the highest-reward neutral trajectory strictly below it as rejected, providing the hardest lower-reward alternative for calibrating independent judgments. Trajectory-discriminative (TD) compares only valid neutral trajectories, choosing the highest-reward output and rejecting the lowest-reward one to retain within-policy preferences.

We retain at most one pair per relation, requiring two valid outputs and a strictly positive chosen-minus-rejected reward gap. HC and HCal can supply corrective feedback even when neutral rewards agree and TD yields no preference. Relation tags preserve the distinction between these references and neutral preferences when the pairs and rewards are passed to the editor. Detailed selection rules appear in Appendix A.3.

3.5 Hierarchical Linguistic Experience Updates

A comparison becomes useful beyond its source video when the distinguishing evidence and conditions are retained as reusable experience. Following feedback-based textual optimization (Ye et al., 2025; Pryzant et al., 2023; Leng et al., 2026), the linguistic optimizer examines the selected relation-tagged pairs $\mathcal{P}_r$, their rewards, the current bank, and video-level supervision (Y, C). It identifies which entries support or mislead the judgment and proposes additions, revisions, or deletions, each citing a supporting relation. Validated edits $\Delta\mathcal{E}_r$ produce the next bank through $\mathcal{E}_{r+1} = \mathrm{Apply}(\mathcal{E}_r, \Delta\mathcal{E}_r)$.

The hierarchy $\mathcal{E} = (\mathcal{E}^{\mathrm{pri}}, \mathcal{E}^{\mathrm{pat}}, \mathcal{E}^{\mathrm{clu}})$ organizes these revisions by their role in a judgment. Principles specify when a cue is diagnostic, Patterns describe the relevant event progression, and Clues identify supporting visual evidence. Figure 2(d) illustrates a criterion revision from "contact" to "contact with escalation." Such an edit adds an applicability condition to a broad cue, making the comparison useful for interpreting subsequent interactions. Algorithm 1 and Appendix A.4 give the learning procedure and edit constraints.

Frozen-bank inference. At test time, the frozen VLM uses the fixed final bank with a neutral prompt, without auxiliary models or competing hypotheses. Inference details appear in Appendix A.5.

3.6 Conditions for Expected-Reward Improvement

Experience editing changes the context-conditioned output distribution while preserving model weights. Appendix B.2 characterizes this response locally. Building on group-relative feedback (Shao et al., 2024), Appendix B.4 expresses the gradient of a neutral-group reward surrogate as a sum of reward-weighted pairwise log-odds gradients. This connects trajectory comparisons to a policy-update direction. The identity concerns the full neutral group, while our editor receives sparse preferences and separately conditioned corrective references.

To characterize when a candidate edit improves expected reward, fix ν and R and write $p = \pi_{\mathcal{E}'}$ and $A_r(z, o) = R(z, o) - \mathbb{E}_{\pi_r}[R(z, o) \mid z]$. Let $g_r(z, o)$ be an analytical advantage proxy defined on the output support and consistent with the selected preference directions. Finite comparisons leave its values on other outputs unspecified, so the discrepancy from A_r captures feedback error.

For $\beta > 0$, the standard KL-regularized policy construction (Rafailov et al., 2023) gives the reference

$$q_r^\star(o \mid z) = \frac{\pi_r(o \mid x)\exp(g_r(z, o)/\beta)}{Z_r(z)}, \quad Z_r(z) = \mathbb{E}_{\pi_r}[\exp(g_r(z, o)/\beta) \mid z]. \tag{4}$$

This reference shifts pairwise log odds by the corresponding proxy gap divided by β, providing a target for analyzing the response to an edit. The proxy and reference provide an analytical framework for characterizing the policy response to an edit.

Proposition 3.1 (Sufficient condition for a candidate policy) *Assume $p(\cdot \mid x) \ll \pi_r(\cdot \mid x)$ and $0 < Z_r(z) < \infty$ almost surely, with finite expected $|\log Z_r|$, KL terms (including $D_{\mathrm{KL}}(p\|\pi_r)$), and absolute reward and proxy-error expectations. Let d_{old} and d_{new} be the expected KL divergences from π_r and p to $q_r^\star$, and let e_{old} and e_{new} be the expected errors $|g_r - A_r|$ under those two policies, respectively. Then*

$$J(\mathcal{E}') - J(\mathcal{E}_r) \geq \beta(d_{\mathrm{old}} - d_{\mathrm{new}}) - e_{\mathrm{old}} - e_{\mathrm{new}}. \tag{5}$$

Expected reward therefore improves when $\beta(d_{\mathrm{old}} - d_{\mathrm{new}}) > e_{\mathrm{old}} + e_{\mathrm{new}}$. The bound separates feedback quality, measured by the proxy errors, from the edit-induced policy response, measured by movement toward the reference. Appendix B.6 proves the bound, and Appendices B.7 and B.8 analyze sparse feedback and finite sampling.

Table 1: Frame-level detection and open-set transfer. Bold/underline mark the best/second-best scores within each group. Expl. and Lrn. denote explanation output and learned guidance. ⋆ marks closed-source models, † XD-learned priors, and ◇ adaptive segmentation. Dashes indicate unreported or inapplicable entries.

Method (venue)	**Supervision**	**Expl.**	**Lrn.**	**Stride**	**Multi-Scenario**		**Open-Set**
					XD AP (%)	**UCF AUC (%)**	**UB AUC (%)**
Task-Specific Anomaly Detectors							
AED-MAE (Ristea et al., 2024) (CVPR'24)	Semi	✗	–	–	–	–	<u>58.50</u>
STPAG (Rai et al., 2024) (CVPRW'24)	Semi	✗	–	–	–	–	57.98
RTFM (Tian et al., 2021) (ICCV'21)	Weak	✗	–	1	77.81	84.30	**64.94**
PEL4VAD (Pu et al., 2024) (TIP'24)	Weak	✗	–	1	<u>85.59</u>	86.76	–
VadCLIP (Wu et al., 2024) (AAAI'24)	Weak	✗	–	1	84.51	<u>88.02</u>	–
Ex-VAD (Huang et al., 2025a) (ICML'25)	Weak	✓	–	–	**86.52**	**88.29**	–
Tuning-Free Anomaly Understanding							
ZS CLIP (Radford et al., 2021) (ICML'21)	TF	✗	✗	–	17.83	53.16	46.20
LLaVA-1.5 (Liu et al., 2024) (CVPR'24)	TF	✓	✗	–	50.26	72.84	53.71
LAVAD (Zanella et al., 2024) (CVPR'24)	TF	✓	✗	1	62.01	80.28	64.23
EventVAD (Shao et al., 2025) (MM'25)	TF	✓	✗	◇	64.04	82.03	–
URF (Lin et al., 2025) (NeurIPS'25)	TF	✓	✗	1	68.07	84.28	69.02
VADTree (Li et al., 2025) (NeurIPS'25)	TF	✓	✗	◇	68.85	84.74	–
PANDA⋆ (Yang et al., 2025b) (NeurIPS'25)	TF	✓	✗	8	70.16	84.89	75.78
AnomalyRuler⋆ (Yang et al., 2024) (ECCV'24)	TF	✓	✓	1	–	–	71.90
VERA (Ye et al., 2025) (CVPR'25)	VL	✓	✓	1	70.11	86.55	71.65†
LRPO (Leng et al., 2026) (ICML'26)	VL	✓	✓	4	<u>74.09</u>	<u>86.59</u>	76.24†
Anom-Ψ (Ours)	VL	✓	✓	8	73.40	84.79	**77.09**†
				4	**78.43**	**86.73**	<u>77.01</u>†

Table 2: MSAD detection results. Subscript a denotes evaluation on abnormal videos only.

Method (venue)	**Supervision**	**AUC (%)**	**AUC$_a$ (%)**	**AP (%)**	**AP$_a$ (%)**
RTFM (Tian et al., 2021) (ICCV'21)	Weak	86.65	–	–	–
MGFN (Chen et al., 2023b) (AAAI'23)	Weak	84.96	–	–	–
TEVAD (Chen et al., 2023a) (CVPRW'23)	Weak	86.82	–	–	–
UR-DMU (Zhou et al., 2023) (AAAI'23)	Weak	85.78	67.95	67.35	75.30
GS-MoE (D'Amicantonio et al., 2025) (ICCV'25)	Weak	87.72	69.54	68.26	76.68
π-VAD (Majhi et al., 2025) (CVPR'25)	Weak	88.68	71.25	71.26	77.86
VAD-DPO (Zhang et al., 2025b) (NeurIPS'25)	Pref	85.40	–	–	–
VADTree (Li et al., 2025) (NeurIPS'25)	TF	89.32	67.85	71.41	75.49
LaGoVAD (Liu et al., 2026) (ICLR'26)	Weak	90.41	–	–	–
Anom-Ψ (Ours, 20%)	VL	**91.28** ± 0.71	**78.59** ± 1.15	**82.55** ± 1.20	**85.21** ± 1.20

4 EXPERIMENTS

Datasets and evaluation. We evaluate frame-level detection on XD-Violence (Wu et al., 2020) (AP), UCF-Crime (Sultani et al., 2018) (AUC), and MSAD (Zhu et al., 2024) (Micro-AUC and AP), and cross-dataset transfer on UBnormal (Acsintoae et al., 2022) (AUC). On MSAD, we follow the experimental settings of VADTree (Li et al., 2025). All metrics are percentages. Unless varied, we use Qwen3.5-9B with frozen weights. XD-Violence results are from single runs unless stated otherwise, while our MSAD main results report means and sample SDs across three runs. XD benchmark and data-efficiency results use evaluation stride 4, whereas controlled comparisons use stride 8 unless specified otherwise. Published baselines retain their source protocols. Implementation details and complete evaluation settings appear in Appendices A and C.1.

Detection performance and generalization. With model weights frozen, Anom-Ψ leads the listed tuning-free methods on XD-Violence and UCF-Crime, reaching 78.43% AP and 86.73% AUC, respectively. The XD-Violence gain over LRPO is 4.34 percentage points (Table 1). On MSAD, experience learned from 20% of the training split yields 91.28 ± 0.71% Micro-AUC, while AP$_a$ reaches 85.21% versus 77.86% for π-VAD (Table 2). The abnormal-video advantage reflects better discrimination between event intervals and surrounding normal activity, consistent with learning when a visual cue becomes diagnostic. Transferring only XD-learned Principle-level experience to

Table 3: Core component ablations on XD-Violence. VS and FS denote video and frame selection. Parentheses show changes in percentage points from row Ref.

Idx	How to Learn			What to Learn From		Metrics		
	Optimization	Hypothesis	Preference	VS	FS	Ref.	AP (%) ↑	AUC (%) ↑
0						–	57.14	85.57
1	✓					0	64.78 (+7.64)	86.95 (+1.38)
2	✓	✓				1	66.25 (+1.47)	89.08 (+2.13)
3	✓	✓	✓			2	68.96 (+2.71)	88.86 (−0.22)
4	✓	✓	✓	✓		3	69.75 (+0.79)	87.40 (−1.46)
5	✓	✓	✓		✓	3	70.39 (+1.43)	89.07 (+0.21)
6	✓			✓	✓	–	66.80	88.49
7	✓		✓	✓	✓	6	69.89 (+3.09)	88.16 (−0.33)
8	✓	✓	✓	✓	✓	7	**73.40** (+3.51)	**90.30** (+2.14)

Table 4: Preference learning ablations on XD-Violence. N denotes neutral responses, D additional distinct hypothesis-free samples, and H_{A}/H_{N} abnormal/normal hypotheses.

(a) Preference composition

TD	HC	HCal	AP (%)	AUC (%)
✓			69.36	87.37
✓	✓		71.65	89.89
✓		✓	72.17	89.55
	✓	✓	70.28	88.04
✓	✓	✓	**73.40**	**90.30**

(b) Pair selection

Strategy	AP (%)	AUC (%)
Random	67.13	87.23
Largest-gap	72.92	**90.50**
Best-anchor	69.37	89.71
Stratified random	66.83	86.46
Relation-aware	**73.40**	90.30

(c) Budget controls

Mix	AP (%)	AUC (%)
$4N$	68.55	88.83
$4N + 2D$	69.89	88.16
$4N + 2H_{\mathrm{A}}$	66.27	84.25
$4N + 2H_{\mathrm{N}}$	67.31	88.65
$4N + H_{\mathrm{A}} + H_{\mathrm{N}}$	**73.40**	**90.30**

UBnormal further yields 77.01% AUC without target training or bank updates, showing that these criteria remain useful beyond the source scenes.

Component contributions. Explicit comparisons improve how the model learns from the same observations. With category-balanced random videos and uniform frames, adding hypothesis guidance and preference updates improves AP over direct learning by 4.18 percentage points (Table 3). With both observation selectors enabled, the complete method reaches 73.40% AP, gaining 16.26 percentage points over the no-experience baseline. The matched-observation comparison shows that the learning procedure contributes beyond observation selection.

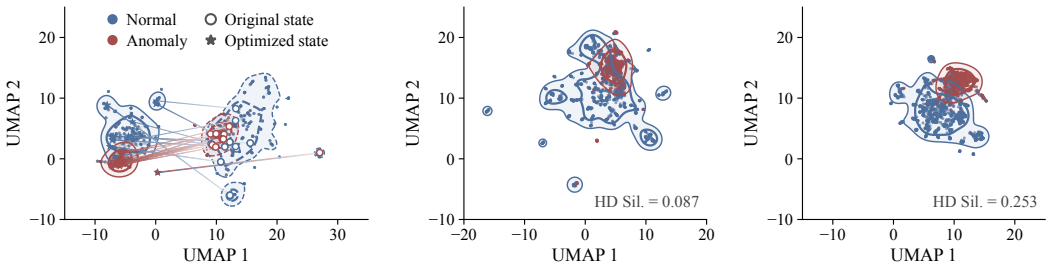
How to learn. Useful feedback should both challenge an anomaly criterion and reflect how the learner currently applies it. Adding HC and HCal to TD improves AP by 4.04 percentage points, while removing TD from the complete method lowers it by 3.12 percentage points (Table 4(a)). Removing either HC or HCal also reduces AP. These results support complementary roles for corrective references that challenge current decisions and neutral preferences that reveal differences among the learner's independent judgments.

At a fixed generation budget, the complete opposing-hypothesis update improves AP by 3.51 percentage points over the $4N + 2D$ control (panel (c)). Repeating either the normal or abnormal premise also performs worse than opposing premises. With generation and pairing varied jointly, these controls favor supplying both interpretations over additional neutral or same-premise samples.

Pair selection should focus the editing budget on informative competitors. Relation-aware selection exceeds Stratified random by 6.57 percentage points in AP even though both retain the same relation types (panel (b)). The gain supports targeted comparisons within each relation, using the hardest lower-reward neutral alternative for HCal and the reward extremes for TD.

The combined feedback also makes learning more consistent across runs. Anom-Ψ improves mean AP over TD-only while reducing the sample SD from 4.70 to 0.90 percentage points across independent learning runs (Figure 5(c), Appendix C.2).

What to learn from. Salient moments can identify an event while leaving the context of its development undersampled. With the Hybrid frame-selection strategy fixed, wavelet-prioritized video selec-



(a) Paired representations (b) Base (c) Learned Experience

Figure 3: UMAP of MSAD representations. (a) Eight pairs per class in a shared embedding (circles: Base; stars: Learned Experience). (b,c) Separate fits with matched samples, preprocessing, and parameters.

tion improves AP over category-balanced random selection by 3.01 percentage points (Table 8(a)). With the frame budget and high-score component fixed, replacing uniform contextual frames with WFS frames adds 5.00 percentage points (panel (b)). Video selection prioritizes informative training cases, while WFS complements high-score frames with evidence of event progression.

Experience structure. The experience hierarchy organizes general criteria, event patterns, and local evidence for reuse. Within the proposed learning scheme, it improves AP over a flat bank by 2.61 percentage points, whereas it does not improve AP under direct learning (Table 10). This contrast suggests that the hierarchy is most useful when updates already provide informative distinctions to organize. Per-level removal and auxiliary-reward results appear in Appendix C.2.

Refining contextual criteria. Useful updates should suppress false alarms around an event while preserving the event itself. With the starting bank, training video, observations, and editing budgets matched, hypothesis-guided updates improve AP by 0.42 percentage points on average, compared with 0.14 for neutral-only updates (Figure 5(b)). They also reduce adjacent-normal FPR by 0.37 versus 0.17 percentage points while improving anomaly recall. These paired gains indicate better temporal selectivity, with fewer boundary false alarms and greater coverage of annotated events (Appendix C.3).

Experience also affects the representations from which the frozen model produces its judgments. On matched MSAD inputs, the high-dimensional cosine silhouette score rises from 0.0866 under Base to 0.2533 with learned experience (Figure 3). Under label-opposite hints on XD-Violence, the prediction flip rate is 13.31% for Anom-Ψ, compared with 22.69% for the zero-shot baseline (Appendix C.6). Greater class separation and lower hint sensitivity provide complementary evidence for adaptation through textual context. The error transitions remain class-dependent, with less false-positive induction but more false-negative induction.

Figure 4 illustrates how learned conditions help distinguish categories that share a salient weapon cue. Anom-Ψ interprets the firearm in relation to physical restraint and the absence of firing, grounding its judgment in how the weapon is used within the interaction. Full responses and visualization details appear in Appendix D.3.

Data efficiency. Reusable experience should benefit videos beyond those used to learn it. With only 3.5% of the XD-Violence training split, Anom-Ψ trails full-data learning by 1.81 percentage points in AP and 0.28 percentage points in transferred UBnormal AUC (Figure 5(a), Table 12). These small gaps support learning a bank that serves both detection and transfer from a compact adaptation set.

Experience learning also depends on the diversity of its feedback. On MSAD, the category-balanced 20% subset exceeds full-data learning by 2.36 percentage points in both AUC and AP (Table 12). The full split retains its predominance of normal videos without rebalancing. This result suggests that balanced category exposure can be more useful for experience learning than additional normal-dominated data (Appendix C.4).

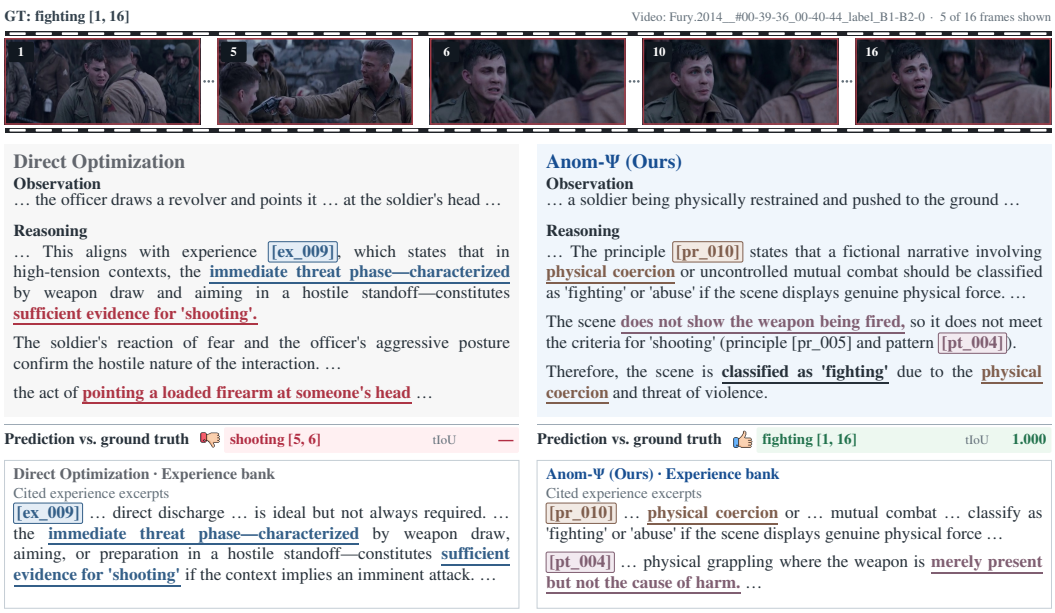
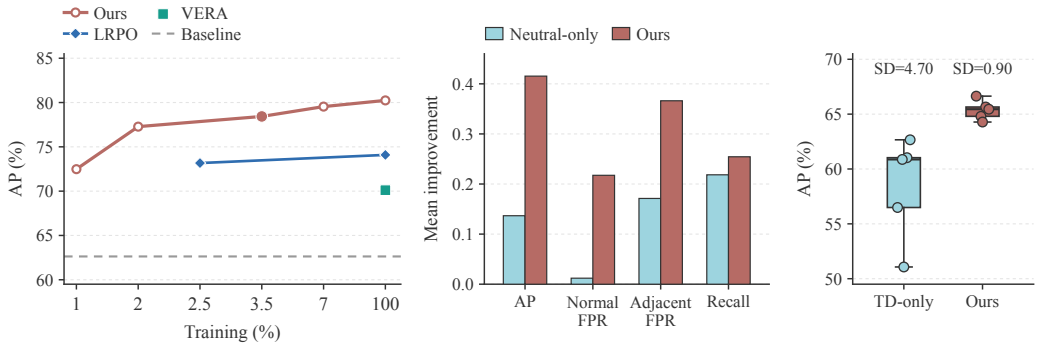


Figure 4: Distinguishing a firearm threat from shooting in *Fury* (2014). Direct Optimization predicts shooting from aiming alone, whereas Anom-Ψ recognizes physical coercion and correctly predicts fighting.



(a) Data efficiency (b) One-step update gains (c) Learning stability

Figure 5: Experience learning on XD-Violence. (a) AP by training budget, with equally spaced ticks. The filled circle marks the default setting and the dashed line the zero-shot baseline. (b) Mean one-step improvements in percentage points. (c) AP distributions over five independent learning runs per method.

5 Conclusion

We presented Anom-Ψ, which learns contextual anomaly criteria by turning contrasts between normal and abnormal hypotheses into corrective feedback. Under video-level supervision, comparisons over shared, change-aware observations complement neutral preferences and guide edits to a hierarchical experience bank, adapting the model's judgments without changing its weights. Our analysis derives a sufficient condition for these linguistic updates to improve expected reward. Experiments show that the learned experience improves detection with limited adaptation data and remains useful in cross-dataset transfer. Ablations and repeated runs further support the complementary value of hypothesis contrasts for more effective and consistent experience learning. Improving the reliability of experience generation and editing for smaller backbones remains a direction for future work.

AI USE STATEMENT

Generative AI tools assisted language polishing, literature retrieval, and the discovery of related work. The authors are responsible for verifying AI-assisted text, technical claims, and bibliographic information against the underlying research records and original sources, and for the final content of the paper.

ETHICS STATEMENT

This work studies video anomaly reasoning on established benchmarks. Anomaly judgments depend on context, and false positives can mischaracterize benign behavior. Biases in pretrained models and limited dataset coverage may also affect anomaly scores, generated explanations, and learned experience. Potential misuse includes intrusive surveillance and unsupported judgments about individuals. Any deployment should therefore include validation in the intended setting, human review of flagged events, and safeguards for the privacy of people appearing in videos. Generated explanations should be checked against the visual evidence before informing decisions that affect individuals.

REPRODUCIBILITY STATEMENT

Section 3 describes the learning and inference procedures, and Section 4 specifies the datasets, metrics, training budgets, and evaluation protocols. Appendix B provides the assumptions and proofs underlying the analysis. Appendix A, including Algorithm 1 and Table 5, details observation selection, trajectory generation, reward computation, preference construction, experience editing, and frame scoring. Additional component comparisons and budget studies appear in Appendix C; experiment-specific settings and departures from the default configuration are stated with the corresponding results.

REFERENCES

Andra Acsintoae, Andrei Florescu, Mariana-Iuliana Georgescu, Tudor Mare, Paul Sumedrea, Radu Tudor Ionescu, Fahad Shahbaz Khan, and Mubarak Shah. UBnormal: New benchmark for supervised open-set video anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 20143–20153, 2022.

Yuzheng Cai, Siqi Cai, Yuchen Shi, Zihan Xu, Lichao Chen, Yulei Qin, Xiaoyu Tan, Gang Li, Zongyi Li, Haojia Lin, Yong Mao, Ke Li, and Xing Sun. Training-free group relative policy optimization. *arXiv preprint arXiv:2510.08191*, 2025. doi: 10.48550/arXiv.2510.08191. URL https://arxiv.org/abs/2510.08191.

Weiling Chen, Keng Teck Ma, Zi Jian Yew, Minhoe Hur, and David Aik-Aun Khoo. TEVAD: Improved video anomaly detection with captions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 5549–5559, 2023a.

Yingxian Chen, Zhengzhe Liu, Baoheng Zhang, Wilton Fok, Xiaojuan Qi, and Yik-Chung Wu. MGFN: Magnitude-contrastive glance-and-focus network for weakly-supervised video anomaly detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 387–395, 2023b. doi: 10.1609/aaai.v37i1.25112.

Damai Dai, Yutao Sun, Li Dong, Yaru Hao, Shuming Ma, Zhifang Sui, and Furu Wei. Why can GPT learn in-context? language models secretly perform gradient descent as meta-optimizers. In *Findings of the Association for Computational Linguistics: ACL 2023*, pp. 4005–4019. Association for Computational Linguistics, 2023. doi: 10.18653/v1/2023.findings-acl.247.

Giacomo D'Amicantonio, Snehashis Majhi, Quan Kong, Lorenzo Garattoni, Gianpiero Francesca, Francois Bremond, and Egor Bondarev. Mixture of experts guided by gaussian splatters matters: A new approach to weakly-supervised video anomaly detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 10275–10285, 2025.

Xi Ding, Lei Wang, Piotr Koniusz, and Yongsheng Gao. Learning time in static classifiers. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 20816–20825, 2026. doi: 10.1609/aaai.v40i25.39221.

Chao Huang, Yushu Shi, Jie Wen, Wei Wang, Yong Xu, and Xiaochun Cao. Ex-VAD: Explainable fine-grained video anomaly detection based on visual-language models. In *Forty-second International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 25750–25761, 2025a.

Chao Huang, Benfeng Wang, Wei Wang, Jie Wen, Chengliang Liu, Li Shen, and Xiaochun Cao. Vad-R1: Towards video anomaly reasoning via perception-to-cognition chain-of-thought. In *Advances in Neural Information Processing Systems*, volume 38, pp. 118486–118518, 2025b. doi: 10.52202/085713-3952.

Jiaxu Leng, Jiankang Zheng, Mengjingcheng Mo, Zhanjie Wu, Haosheng Chen, Ji Gan, and Xinbo Gao. Linguistic relative policy optimization for video anomaly reasoning. In *International Conference on Machine Learning*, 2026. URL https://arxiv.org/abs/2607.00654.

Wenlong Li, Yifei Xu, Yuan Rao, Zhenhua Wang, and Shuiguang Deng. VADTree: Explainable training-free video anomaly detection via hierarchical granularity-aware tree. In *Advances in Neural Information Processing Systems*, volume 38, pp. 148372–148404, 2025. doi: 10.52202/085713-4959.

Dongheng Lin, Mengxue Qu, Kunyang Han, Jianbo Jiao, Xiaojie Jin, and Yunchao Wei. A unified reasoning framework for holistic zero-shot video anomaly analysis. In *Advances in Neural Information Processing Systems*, volume 38, pp. 29713–29744, 2025. doi: 10.52202/085713-0996.

Haotian Liu, Chunyuan Li, Yuheng Li, and Yong Jae Lee. Improved baselines with visual instruction tuning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 26296–26306, 2024.

Zihao Liu, Xiaoyu Wu, Jianqin Wu, Xuxu Wang, and Linlin Yang. Language-guided open-world video anomaly detection under weak supervision. In C. Vondrick, B. Hariharan, C. Raffel, L. Pinto, D. Yang, and A. Faust (eds.), *International Conference on Learning Representations*, volume 2026, pp. 153197–153224, 2026.

Snehashis Majhi, Giacomo D'Amicantonio, Antitza Dantcheva, Quan Kong, Lorenzo Garattoni, Gianpiero Francesca, Egor Bondarev, and Francois Bremond. Just dance with π! a poly-modal inductor for weakly-supervised video anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 24265–24274, 2025.

Leland McInnes, John Healy, and James Melville. UMAP: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*, 2018. URL https://arxiv.org/abs/1802.03426.

Mengjingcheng Mo, Jiaxu Leng, and Xinbo Gao. Learning to watch: Active video anomaly understanding via interleaved policy optimization. In *International Conference on Machine Learning*, 2026. URL https://arxiv.org/abs/2607.00622.

Kevin R. Moon, David van Dijk, Zheng Wang, Scott Gigante, Daniel B. Burkhardt, William S. Chen, Kristina Yim, Antonia van den Elzen, Matthew J. Hirn, Ronald R. Coifman, Natalia B. Ivanova, Guy Wolf, and Smita Krishnaswamy. Visualizing structure and transitions in high-dimensional biological data. *Nature Biotechnology*, 37:1482–1492, 2019. doi: 10.1038/s41587-019-0336-3. URL https://www.nature.com/articles/s41587-019-0336-3.

Wenbo Pan, Zhichao Liu, Xianlong Wang, Haining Yu, and Xiaohua Jia. Towards long-horizon interpretability: Efficient and faithful multi-token attribution for reasoning LLMs. *arXiv preprint arXiv:2602.01914*, 2026. URL https://arxiv.org/abs/2602.01914.

Reid Pryzant, Dan Iter, Jerry Li, Yin Lee, Chenguang Zhu, and Michael Zeng. Automatic prompt optimization with "gradient descent" and beam search. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pp. 7957–7968. Association for Computational Linguistics, 2023. doi: 10.18653/v1/2023.emnlp-main.494.

Yujiang Pu, Xiaoyu Wu, Lulu Yang, and Shengjin Wang. Learning prompt-enhanced context features for weakly-supervised video anomaly detection. *IEEE Transactions on Image Processing*, 33:4923–4936, 2024. doi: 10.1109/TIP.2024.3451935.

Qwen Team. Qwen3.5: Towards native multimodal agents, February 2026. URL https://qwen.ai/blog?id=qwen3.5.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In *International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 8748–8763, 2021.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D. Manning, Stefano Ermon, and Chelsea Finn. Direct preference optimization: Your language model is secretly a reward model. In *Advances in Neural Information Processing Systems*, volume 36, pp. 53728–53741, 2023. doi: 10.52202/075280-2338.

Ayush K. Rai, Tarun Krishna, Feiyan Hu, Alexandru Drimbarean, Kevin McGuinness, Alan F. Smeaton, and Noel E. O'Connor. Video anomaly detection via spatio-temporal pseudo-anomaly generation: A unified approach. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 3887–3899, 2024.

Nicolae-Catalin Ristea, Florinel-Alin Croitoru, Radu Tudor Ionescu, Marius Popescu, Fahad Shahbaz Khan, and Mubarak Shah. Self-distilled masked auto-encoders are efficient video anomaly detectors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 15984–15995, 2024.

Yihua Shao, Haojin He, Sijie Li, Siyu Chen, Xinwei Long, Fanhu Zeng, Yuxuan Fan, Muyang Zhang, Ziyang Yan, Ao Ma, Xiaochen Wang, Hao Tang, Yan Wang, and Shuyan Li. EventVAD: Training-free event-aware video anomaly detection. In *Proceedings of the 33rd ACM International Conference on Multimedia*, pp. 2586–2595, 2025. doi: 10.1145/3746027.3754500.

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, Y. K. Li, Y. Wu, and Daya Guo. DeepSeekMath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*, 2024. doi: 10.48550/arXiv.2402.03300. URL https://arxiv.org/abs/2402.03300.

Waqas Sultani, Chen Chen, and Mubarak Shah. Real-world anomaly detection in surveillance videos. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6479–6488, 2018. URL https://openaccess.thecvf.com/content_cvpr_2018/html/Sultani_Real-World_Anomaly_Detection_CVPR_2018_paper.html.

V Team, Wenyi Hong, Wenmeng Yu, Xiaotao Gu, Guo Wang, Guobing Gan, Haomiao Tang, Jiale Cheng, Ji Qi, Junhui Ji, Lihang Pan, Shuaiqi Duan, Weihan Wang, Yan Wang, Yean Cheng, Zehai He, Zhe Su, Zhen Yang, Ziyang Pan, Aohan Zeng, Baoxu Wang, Bin Chen, Boyan Shi, Changyu Pang, Chenhui Zhang, Da Yin, Fan Yang, Guoqing Chen, Haochen Li, Jiale Zhu, Jiali Chen, Jiaxing Xu, Jiazheng Xu, Jing Chen, Jinghao Lin, Jinhao Chen, Jinjiang Wang, Junjie Chen, Leqi Lei, Letian Gong, Leyi Pan, Mingdao Liu, Mingde Xu, Mingzhi Zhang, Qinkai Zheng, Ruiliang Lyu, Shangqin Tu, Sheng Yang, Shengbiao Meng, Shi Zhong, Shiyu Huang, Shuyuan Zhao, Siyan Xue, Tianshu Zhang, Tianwei Luo, Tianxiang Hao, Tianyu Tong, Wei Jia, Wenkai Li, Xiao Liu, Xiaohan Zhang, Xin Lyu, Xinyu Zhang, Xinyue Fan, Xuancheng Huang, Yadong Xue, Yanfeng Wang, Yanling Wang, Yanzi Wang, Yifan An, Yifan Du, Yiheng Huang, Yilin Niu, Yiming Shi, Yu Wang, Yuan Wang, Yuanchang Yue, Yuchen Li, Yusen Liu, Yutao Zhang, Yuting Wang, Yuxuan Zhang, Zhao Xue, Zhengxiao Du, Zhenyu Hou, Zihan Wang, Peng Zhang, Debing Liu, Bin Xu, Juanzi Li, Minlie Huang, Yuxiao Dong, and Jie Tang. GLM-4.5V and GLM-4.1V-Thinking: Towards Versatile Multimodal Reasoning with Scalable Reinforcement Learning. *arXiv preprint arXiv:2507.01006*, 2025. URL https://arxiv.org/abs/2507.01006.

Yu Tian, Guansong Pang, Yuanhong Chen, Rajvinder Singh, Johan W. Verjans, and Gustavo Carneiro. Weakly-supervised video anomaly detection with robust temporal feature magnitude learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 4975–4986, 2021.

Weiyun Wang, Zhangwei Gao, Lixin Gu, Hengjun Pu, Long Cui, Xingguang Wei, Zhaoyang Liu, Linglin Jing, Shenglong Ye, Jie Shao, Zhaokai Wang, Zhe Chen, Hongjie Zhang, Ganlin Yang, Haomin Wang, Qi Wei, Jinhui Yin, Wenhao Li, Erfei Cui, Guanzhou Chen, Zichen Ding, Changyao Tian, Zhenyu Wu, Jingjing Xie, Zehao Li, Bowen Yang, Yuchen Duan, Xuehui Wang, Zhi Hou, Haoran Hao, Tianyi Zhang, Songze Li, Xiangyu Zhao, Haodong Duan, Nianchen Deng, Bin Fu, Yinan He, Yi Wang, Conghui He, Botian Shi, Junjun He, Yingtong Xiong, Han Lv, Lijun Wu, Wenqi Shao, Kaipeng Zhang, Huipeng Deng, Biqing Qi, Jiaye Ge, Qipeng Guo, Wenwei Zhang, Songyang Zhang, Maosong Cao, Junyao Lin, Kexian Tang, Jianfei Gao, Haian Huang, Yuzhe Gu, Chengqi Lyu, Huanze Tang, Rui Wang, Haijun Lv, Wanli Ouyang, Limin Wang, Min Dou, Xizhou Zhu, Tong Lu, Dahua Lin, Jifeng Dai, Weijie Su, Bowen Zhou, Kai Chen, Yu Qiao, Wenhai Wang, and Gen Luo. InternVL3.5: Advancing Open-Source Multimodal Models in Versatility, Reasoning, and Efficiency. *arXiv preprint arXiv:2508.18265*, 2025. URL https://arxiv.org/abs/2508.18265.

Peng Wu, Jing Liu, Yujia Shi, Yujia Sun, Fangtao Shao, Zhaoyang Wu, and Zhiwei Yang. Not only look, but also listen: Learning multimodal violence detection under weak supervision. In *European Conference on Computer Vision*, 2020. URL https://roc-ng.github.io/XD-Violence/.

Peng Wu, Xuerong Zhou, Guansong Pang, Lingru Zhou, Qingsen Yan, Peng Wang, and Yanning Zhang. VadCLIP: Adapting vision-language models for weakly supervised video anomaly detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 6074–6082, 2024. doi: 10.1609/aaai.v38i6.28423.

Biao Yang, Bin Wen, Boyang Ding, Changyi Liu, Chenglong Chu, Chengru Song, Chongling Rao, Chuan Yi, Da Li, Dunju Zang, Fan Yang, Guorui Zhou, Guowang Zhang, Han Shen, Hao Peng, Haojie Ding, Hao Wang, Haonan Fan, Hengrui Ju, Jiaming Huang, Jiangxia Cao, Jiankang Chen, Jingyun Hua, Kaibing Chen, Kaiyu Jiang, Kaiyu Tang, Kun Gai, Muhao Wei, Qiang Wang, Ruitao Wang, Sen Na, Shengnan Zhang, Siyang Mao, Sui Huang, Tianke Zhang, Tingting Gao, Wei Chen, Wei Yuan, Xiangyu Wu, Xiao Hu, Xingyu Lu, Yi-Fan Zhang, Yiping Yang, Yulong Chen, Zeyi Lu, Zhenhua Wu, Zhixin Ling, Zhuoran Yang, Ziming Li, Di Xu, Haixuan Gao, Hang Li, Jing Wang, Lejian Ren, Qigen Hu, Qianqian Wang, Shiyao Wang, Xinchen Luo, Yan Li, Yuhang Hu, and Zixing Zhang. Kwai Keye-VL 1.5 Technical Report. *arXiv preprint arXiv:2509.01563*, 2025a. URL https://arxiv.org/abs/2509.01563.

Yuchen Yang, Kwonjoon Lee, Behzad Dariush, Yinzhi Cao, and Shao-Yuan Lo. Follow the rules: Reasoning for video anomaly detection with large language models. In *European Conference on Computer Vision*, pp. 304–322, 2024. doi: 10.1007/978-3-031-73004-7_18.

Zhiwei Yang, Chen Gao, and Mike Zheng Shou. PANDA: Towards generalist video anomaly detection via agentic AI engineer. In *Advances in Neural Information Processing Systems*, volume 38, pp. 83182–83211, 2025b. doi: 10.52202/085713-2788.

Muchao Ye, Weiyang Liu, and Pan He. VERA: Explainable video anomaly detection via verbalized learning of vision-language models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8679–8688, 2025.

Tianyu Yu, Zefan Wang, Chongyi Wang, Fuwei Huang, Wenshuo Ma, Zhihui He, Tianchi Cai, Weize Chen, Yuxiang Huang, Yuanqian Zhao, Bokai Xu, Junbo Cui, Yingjing Xu, Liqing Ruan, Luoyuan Zhang, Hanyu Liu, Jingkun Tang, Hongyuan Liu, Qining Guo, Wenhao Hu, Bingxiang He, Jie Zhou, Jie Cai, Ji Qi, Zonghao Guo, Chi Chen, Guoyang Zeng, Yuxuan Li, Ganqu Cui, Ning Ding, Xu Han, Yuan Yao, Zhiyuan Liu, and Maosong Sun. MiniCPM-V 4.5: Cooking Efficient MLLMs via Architecture, Data, and Training Recipe. *arXiv preprint arXiv:2509.18154*, 2025. URL https://arxiv.org/abs/2509.18154.

Luca Zanella, Willi Menapace, Massimiliano Mancini, Yiming Wang, and Elisa Ricci. Harnessing large language models for training-free video anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 18527–18536, 2024.

Huaxin Zhang, Xiaohao Xu, Xiang Wang, Jialong Zuo, Xiaonan Huang, Changxin Gao, Shanjun Zhang, Li Yu, and Nong Sang. Holmes-VAU: Towards long-term video anomaly understanding

at any granularity. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 13843–13853, 2025a.

Menghao Zhang, Huazheng Wang, Pengfei Ren, Kangheng Lin, Qi Qi, Haifeng Sun, Zirui Zhuang, Lei Zhang, Jianxin Liao, and Jingyu Wang. Do LVLMs truly understand video anomalies? revealing hallucination via co-occurrence patterns. In *Advances in Neural Information Processing Systems*, volume 38, pp. 106738–106765, 2025b. doi: 10.52202/085713-3560.

Hang Zhou, Junqing Yu, and Wei Yang. Dual memory units with uncertainty regulation for weakly supervised video anomaly detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 3769–3777, 2023. doi: 10.1609/aaai.v37i3.25489.

Liyun Zhu, Lei Wang, Arjun Raj, Tom Gedeon, and Chen Chen. Advancing video anomaly detection: A concise review and a new dataset. In *Advances in Neural Information Processing Systems*, volume 37, pp. 89943–89977, 2024. doi: 10.52202/079017-2856.

Table of Contents

A. Implementation Details **15**
A.1. Observation selection ... 15
A.2. Trajectory sampling and rewards ... 16
A.3. Admissible preference pairs ... 18
A.4. Linguistic edit application ... 18
A.5. Frame scoring and inference ... 18
A.6. Core prompt excerpts ... 19
B. Policy Analysis and Proofs **20**
B.1. Fixed-round conventions ... 20
B.2. Context-induced policy changes ... 20
B.3. Hypothesis constraints and decision coverage ... 22
B.4. Group-relative feedback and pairwise directions ... 22
B.5. Relative improvement and finite policy changes ... 23
B.6. Proof of Proposition 3.1 ... 23
B.7. From selected comparisons to the full policy change ... 25
B.8. Finite-group bias and adaptive editing ... 25
B.9. Likelihood diagnostics beyond the editing inputs ... 26
C. Additional Experimental Analysis **26**
C.1. Evaluation Protocols and Detailed Results ... 26
C.2. Component Ablations and Learning Stability ... 27
C.3. Local-update Evaluation and Diagnostics ... 28
C.4. Data and Budget Sensitivity ... 29
C.5. Computational Cost and Test-time Selection ... 30
C.6. Sensitivity to Prompt Interference ... 31
D. Visualizations and Case Studies **32**
D.1. Experience Evolution and Final Bank ... 32
D.2. Representation and Visual Attribution ... 32
D.3. Prediction Curves and Response Comparisons ... 34

The appendices provide implementation and prompt specifications (Appendix A) and analyze how experience edits change the learner's policy (Appendix B). Additional experiments examine component contributions, learning stability, and computational cost (Appendix C). Experience evolution, representation and attribution analyses, and response comparisons complement these results by illustrating how learned criteria affect video interpretation (Appendix D).

A Implementation Details

Learning compares judgments over fixed visual observations and turns supported differences into revisions of a shared experience bank. Algorithm 1 specifies this process with frozen model parameters; Table 5 gives the default settings.

A.1 Observation selection

Experience learning benefits from observations that reveal changes in event state and the context surrounding them. Following the score-guided temporal sampling approach of Holmes-VAU (Zhang et al., 2025a), we use fixed offline VLM anomaly scores $u_j \in [0, 1]$ to select training observations. On the stride-16 candidate timeline, we emphasize mid-range scores through $q_j = \exp[-\frac{1}{2}((u_j - 0.5)/0.2)^2]$, and apply a Daubechies-4 wavelet transform to identify candidate temporal transitions. Videos are ranked by their length-normalized peak prominence, $\mathrm{Rich}(V) = T^{-1} \sum_{b \in \mathcal{B}(V)} \mathrm{prom}(b)$, where $\mathcal{B}(V)$ contains the detected peaks and T is the candidate-timeline length. Selecting the twenty highest-ranked videos per category yields 140 training videos on XD-Violence, concentrating the learning budget on change-rich sequences while maintaining category coverage.

The two index sets in Equation (1) retain complementary information about event context and anomaly evidence. On the candidate grid $\mathcal{J}_i$, wavelet peaks partition each video into segments.

Algorithm 1 Hypothesis-guided linguistic policy optimization

Require: Training set $\mathcal{D}$ and precomputed VLM scores; frozen learner θ_0 and linguistic editor Edit; initial hierarchical bank $\mathcal{E}_0$, rollout count $G = 4$, and step budget R
Ensure: Final experience bank $\mathcal{E}_R$; model parameters remain frozen

Fix training observations $\{z_r = (x_r, Y_r, C_r)\}_{r=0}^{R-1}$ by offline video selection and hybrid frame sampling (Appendix A.1)
for $r = 0, \ldots, R-1$ **do**
  $\pi_r \leftarrow \pi_{\mathcal{E}_r}$, $\mathcal{P}_r \leftarrow \varnothing$, $\mathcal{E}_{r+1} \leftarrow \mathcal{E}_r$
  Shared-observation rollouts and rewards
  Sample G neutral trajectories $o_{r,i}^0 \sim \pi_r(\cdot \mid x_r)$
  Generate $o_r^a \sim \pi_r^a(\cdot \mid x_r)$ and $o_r^n \sim \pi_r^n(\cdot \mid x_r)$ under abnormal and normal premises
  $(o_r^F, o_r^C) \leftarrow (o_r^a, o_r^n)$ if $Y_r = 1$, else (o_r^n, o_r^a)
  Validate outputs; for each valid o, compute $\rho(o) \leftarrow R(z_r, o)$ using active rewards in Equation (3)
  $\mathcal{V}_r \leftarrow \{i : o_{r,i}^0 \text{ is valid}\}$; use rollout indices to break reward ties
  Within-policy preference (TD)
  if $|\mathcal{V}_r| \geq 2$ **then**
    $i^+ \leftarrow \arg\max_{i \in \mathcal{V}_r} \rho(o_{r,i}^0)$, $i^- \leftarrow \arg\min_{i \in \mathcal{V}_r} \rho(o_{r,i}^0)$
    if $\rho(o_{r,i^+}^0) > \rho(o_{r,i^-}^0)$ **then**
      $\mathcal{P}_r \leftarrow \mathcal{P}_r \cup \{(\mathrm{TD}, o_{r,i^+}^0, o_{r,i^-}^0)\}$
    end if
  end if
  Hypothesis-derived comparisons (HC and HCal)
  if o_r^F is valid **then**
    if o_r^C is valid and $\rho(o_r^F) > \rho(o_r^C)$ **then**
      $\mathcal{P}_r \leftarrow \mathcal{P}_r \cup \{(\mathrm{HC}, o_r^F, o_r^C)\}$
    end if
    $\mathcal{H}_r \leftarrow \{i \in \mathcal{V}_r : \rho(o_{r,i}^0) < \rho(o_r^F)\}$
    if $\mathcal{H}_r \neq \varnothing$ **then**
      $j^\star \leftarrow \arg\max_{i \in \mathcal{H}_r} \rho(o_{r,i}^0)$
      $\mathcal{P}_r \leftarrow \mathcal{P}_r \cup \{(\mathrm{HCal}, o_r^F, o_{r,j^\star}^0)\}$
    end if
  end if
  Linguistic experience update
  if $\mathcal{P}_r \neq \varnothing$ **then**
    $\Delta\mathcal{E}_r \leftarrow \mathrm{Edit}(Y_r, C_r, \mathcal{E}_r, \mathcal{P}_r)$ using relation-tagged trajectories and rewards
    At most three supported edits to Principles, Patterns, or Clues
    if $\Delta\mathcal{E}_r$ is a valid, nonempty batch within bank capacity **then**
      $\mathcal{E}_{r+1} \leftarrow \mathrm{Apply}(\mathcal{E}_r, \Delta\mathcal{E}_r)$
    end if
  end if
end for
return $\mathcal{E}_R$

WFS allocates its budget K_Δ according to segment duration and importance, then uses CLIP-based diversity selection to reduce redundant observations. The high-score set $\mathcal{T}_i^s$ selects K_s additional indices from $\mathcal{J}_i \setminus \mathcal{T}_i^\Delta$. With sufficient candidates, these disjoint sets yield $K_\Delta + K_s$ distinct frames. Table 5 specifies the default budgets and fallback rules.

Preference comparisons are more informative when competing interpretations are grounded in the same observations. We therefore form $\widetilde{V}_i$ by the chronological union in Equation (2) and fix it before experience optimization; all neutral and hypothesis-conditioned trajectories for a training video receive the same visual input. Their comparisons consequently reflect differences in judgment over shared evidence rather than differences in the available frames.

A.2 Trajectory sampling and rewards

Trajectory sampling. Neutral rollouts reveal the learner's current judgments, while opposing hypotheses expose interpretations that these rollouts may leave unexplored. At step r, we generate $G = 4$ neutral trajectories together with one abnormal-premise and one normal-premise trajectory. The video label assigns the two hypothesis-conditioned trajectories their factual and counterfactual

Table 5: Default sampling, experience-editing, and inference settings.

Component	Default setting
Training VLM scorer	Qwen3.5-9B takes video-frame sequences and directly returns anomaly scores $u_j \in [0, 1]$, fixed for training observation selection and auxiliary rewards.
Video selection	Stride-16 candidate grid; Daubechies-4 wavelets on q_j. For detail signal d, peak height is $\mathrm{mean}(\lvert d\rvert) + 0.5\,\mathrm{std}(\lvert d\rvert)$, prominence is $0.05(\max\lvert d\rvert - \min\lvert d\rvert)$, and minimum distance is $\max(32, \lfloor 0.02T \rfloor)$. XD-Violence retains the top 20 videos per category by $\mathrm{Rich}(V)$.
Frame selection	Default budgets are $K_\Delta = 12$ WFS frames and $K_s = 4$ additional unique top-score frames (Hybrid-12+4), merged chronologically. WFS duration/mean/max/variance weights are $(0.4, 0.2, 0.3, 0.1)$; filtering strictness is 1.2, allocation temperature is 1.0, and CLIP-based MMR trade-off is 0.5. Without reliable boundaries, WFS uses half-uniform, half-top-importance selection; remaining slots are filled in VLM-score order.
Bank context	Every learner request receives the full current bank, without query-dependent retrieval. All training branches share the same bank and visual observations.
Generation	Four neutral trajectories and one per hypothesis. Learner temperature 1.0; editor temperature 0; top-$p = 1.0$; model thinking disabled.
Output format	JSON with a sequence summary, at most four distinct key positions, a Principle–Pattern–Clue reasoning trace, and category-tagged inclusive intervals. Positions are 1-based sampled-frame indices; valid outputs must satisfy position, category, and premise constraints.
Rewards	Accuracy/change/evidence weights $0.5/0.25/0.25$, renormalized over active components. Change: matching radius 1 and $F_{0.5}$. Evidence: top 25% of sampled positions by raw VLM score, requiring a 90th–10th percentile spread of at least 0.05.
Pair selection	At most one pair per HC, HCal, and TD relation. Both trajectories must be valid, with a chosen-minus-rejected reward gap strictly greater than $\eta = 0$.
Editor input	Video-level labels and categories; the bank and its capacity; relation-tagged trajectory pairs with observations, reasoning, predictions, and component and total rewards.
Experience edits	Two initial entries per level; capacity $K_{\max} = 10$ per level. At most three add, update, or delete operations per step, each citing a supporting relation. Requested entry length: 10–85 English words. Default XD-Violence budget: 140 video-processing steps.
Inference	Neutral prompt, temperature 0, 16-frame windows, reference stride 16, tail-window retention, and mean aggregation. Dataset-specific strides are reported in Appendix C.1.

roles, as defined in Section 3.3. This combination provides both within-policy variation and explicit alternatives for constructing corrective comparisons.

Accuracy reward. Preference feedback should first distinguish correct anomaly decisions and category predictions. Let $\widehat{C}(o)$ denote the predicted category set and $\mathcal{I}(o)$ the predicted intervals. We define

$$R_{\mathrm{acc}}(z, o) = \begin{cases} \mathbf{1}[\mathcal{I}(o) = \varnothing], & Y = 0, \\ \frac{1}{2}\mathbf{1}[\mathcal{I}(o) \neq \varnothing] + \frac{1}{2}F_1\left(\widehat{C}(o), C\right), & Y = 1. \end{cases} \tag{6}$$

For normal videos, this reward favors an empty anomaly prediction; for anomalous videos, it balances anomaly detection and category agreement. It anchors trajectory ranking in the available video-level supervision.

Change-localization reward. Video-level correctness alone does not indicate whether a trajectory identifies the changes that characterize an event. We therefore compare its predicted key positions $K(o)$ with reliable wavelet peaks mapped to their nearest sampled-frame positions, denoted by Q. Maximum one-to-one matching allows a tolerance of one sampled position. This component is inactive when no reliable peak is available or WFS uses its fallback. Otherwise, for $m > 0$ matches, precision and recall are $P = m/|K(o)|$ and $R = m/|Q|$, and the reward is $R_{\mathrm{ch}}(z, o) = F_{0.5}(P, R) = 1.25PR/(0.25P + R)$. With no matches, including an empty predicted key-position set, we set $R_{\mathrm{ch}}(z, o) = 0$. The precision-weighted score favors selective identification of transitions, while one-to-one matching prevents repeated credit around the same peak.

Evidence-localization reward. Predicted anomaly intervals should also cover observations that support the assigned categories. For an anomalous sample, we take the top 25% of sampled positions ranked by raw VLM scores as the evidence set T_{ev}. Let $P_{\mathrm{ev}}(o)$ contain positions covered by predicted intervals whose categories belong to the ground-truth set C. Their intersection over union defines $R_{\mathrm{ev}}(z, o) = |P_{\mathrm{ev}}(o) \cap T_{\mathrm{ev}}|/|P_{\mathrm{ev}}(o) \cup T_{\mathrm{ev}}|$, rewarding evidence coverage while penalizing unsupported interval expansion. This component is active only for anomalous samples with

available target categories and valid, sufficiently varying VLM scores; the score-spread threshold is specified in Table 5.

Combined feedback. Together, the three components assess outcome correctness, transition selection, and anomaly evidence. We combine them using Equation (3), omitting inactive components and renormalizing the remaining weights. Default weights are given in Table 5.

A.3 Admissible preference pairs

A corrective comparison should identify both a better judgment and the alternative it improves upon. We retain the HC, HCal, and TD relation tags to distinguish these sources of feedback, selecting at most one pair per relation. Every pair requires two valid trajectories and a chosen-minus-rejected reward gap greater than $\eta = 0$, so the factual role alone does not establish a preference. Let $\mathcal{V}_r$ index the valid neutral trajectories, and write $R_i^0 = R(z, o_{r,i}^0)$ and $R_F = R(z, o_r^F)$.

Opposing hypotheses reveal how the same evidence supports different decisions. HC therefore compares the factual and counterfactual trajectories as $\mathcal{P}_r^{\mathrm{HC}} = (o_r^F, o_r^C)$. To connect the factual reference to the learner's own judgments, HCal considers $\mathcal{H}_r = \{i \in \mathcal{V}_r : R_F - R_i^0 > \eta\}$ and selects $j^\star = \arg\max_{i \in \mathcal{H}_r} R_i^0$ when this set is nonempty, forming $\mathcal{P}_r^{\mathrm{HCal}} = (o_r^F, o_{r,j^\star}^0)$. This selects the strongest neutral judgment that still falls below the factual reference in reward, exposing a remaining gap in the learner's current interpretation.

Neutral trajectories also reveal useful differences within the current policy. TD selects $i^+ = \arg\max_{i \in \mathcal{V}_r} R_i^0$ and $i^- = \arg\min_{i \in \mathcal{V}_r} R_i^0$, forming $\mathcal{P}_r^{\mathrm{TD}} = (o_{r,i^+}^0, o_{r,i^-}^0)$ when at least two valid outputs have different rewards. This supplies a comparison between independently sampled judgments without an imposed premise. Rollout indices resolve selection ties; tied rewards never form a preference. An unavailable relation contributes no pair, and editing is skipped when none is available.

A.4 Linguistic edit application

For a trajectory comparison to help on subsequent videos, its distinguishing conditions and evidence must be expressed as reusable experience. The editor receives the selected pairs $\mathcal{P}_r$, their component and total rewards, the video-level supervision (Y, C), and the current bank. It diagnoses which entries support the preferred judgment, miss a necessary condition, or encourage the rejected interpretation. At temperature 0, it proposes $\Delta\mathcal{E}_r = \mathrm{Edit}(Y, C, \mathcal{E}_r, \mathcal{P}_r)$ with at most three add, update, or delete operations, each citing a supporting relation. Principles express applicability conditions, Patterns describe event progression, and Clues specify visual evidence.

Because the full bank conditions every judgment, bounded capacity keeps the accumulated experience compact. The default bank starts with two entries per level and permits $K_{\max} = 10$ per level, or thirty in total. For the $K_{\max} = 1$ capacity ablation (Figure 6(b)), each level instead starts with one designated initial experience, giving three initial entries in total. The editor is instructed to propose edits only when a reusable conclusion supports them and may return no operations. A valid batch is applied as $\mathcal{E}_{r+1} = \mathrm{Apply}(\mathcal{E}_r, \Delta\mathcal{E}_r)$; an empty or invalid batch leaves the bank unchanged. The default XD-Violence budget counts 140 processed videos, so a learning step need not produce an accepted edit.

A.5 Frame scoring and inference

The learned bank supplies the contextual guidance needed for test-time judgments. Inference therefore uses only the frozen learner and the final bank, with a neutral prompt and temperature 0; the offline scorer, CLIP-based frame selection, and other auxiliary models are confined to training. The reference configuration uses 16-frame windows and retains the final window to cover the video tail. The bank remains fixed throughout evaluation.

Predicted intervals use inclusive, 1-based positions within the sixteen sampled frames. Each inference record stores their original-video frame indices as `frame_ids`; evaluation maps interval positions through these recorded indices. On XD-Violence, adjacent sampled frames are sixteen original frames apart. An interval $[a, b]$ therefore covers original frames $[f_a, \min(f_b + 16, L))$,

where f_i is the recorded frame index at sampled position i and L is the video's ground-truth annotation length. For example, with frame indices $[0, 16, 32, 48, \ldots]$, the predicted interval $[2, 4]$ expands to $[16, 64)$, covering original frames 16–63 before any tail clipping. This mapping uses the recorded frame indices rather than linear rescaling by video length. Both inference and evaluation strides on XD-Violence are measured in sampled-frame positions: stride s corresponds to a window-start displacement of $16s$ original frames, so stride 4 corresponds to 64 original frames.

For a window B_k mapped to the original-video timeline, let $u_{k,t}$ denote its contribution at original frame t. A parseable anomalous prediction assigns 1 inside its mapped intervals and 0.5 elsewhere in the window; a parseable normal prediction assigns 0 throughout, and an unparseable output assigns 0.5 throughout. Where windows overlap, their contributions are averaged as $s_t = |\mathcal{K}_t|^{-1} \sum_{k \in \mathcal{K}_t} u_{k,t}$, with $\mathcal{K}_t = \{k : t \in B_k\}$. On XD-Violence, adaptive temporally weighted averaging is then applied to the aggregated frame scores separately within each video, and metrics are computed from the resulting scores. No additional smoothing is applied.

A.6 Core prompt excerpts

The learner and editor need a shared representation to connect a judgment difference to an experience revision. The learner prompt therefore requests observations, hierarchical reasoning, and anomaly intervals; the editor prompt asks which existing experience supports or misleads that judgment and whether the comparison supports a reusable revision. Requiring experience citations and supporting relations makes the connection between evidence, judgment, and edit explicit.

The flame marks the editable bank $\mathcal{E}_r$, whose text supplies the policy context $c_{\mathcal{E}_r}$ in Equation (7). Gray placeholders denote other runtime inputs, and bracketed ellipses mark omitted text. Input names, mathematical annotations, headings, and emphasis are adapted for presentation.

Learner prompt excerpts

Shared system. You are a video anomaly detection assistant. [...] First review the temporal sequence across all input frames. In observation, use sequence_summary to briefly describe the full event progression, then use key_positions to select at most four frames containing **visible changes, state transitions, or strong evidence**. [...] Then, in reasoning, form one coherent chain: Principle boundary → Pattern evolution → Clue evidence → conclusion. Prefer directly applicable experience, and use similar experience only through a credible analogy; cite every used experience by its real ID and explain its role. Skip a level when none applies, and reason from visible frames when no level applies. Do not classify a clip as normal merely because it appears cinematic or staged. [...]

Shared inputs. Frame index range: ⟨*Sampled frame range*⟩.

Experience bank $\mathcal{E}_r$ 🔥

Neutral (training and inference). [...] Determine whether these frames contain an anomaly and return observation, reasoning, and results in that order. [...] **Return an empty list when no anomaly is present.** [...]

Normal premise (training). [...] **The video is known to be normal.** Review the complete sequence and return observation with a full progression summary and key positions, then hierarchical reasoning and **an exactly empty results list**.

Abnormal premise (training). [...] **The ground-truth anomaly categories include** ⟨*Anomaly category set*⟩. Review the complete sequence and return observation with a full progression summary and key positions, then hierarchical reasoning and **at least one results item**.

Optimizer prompt excerpts

System. You are an experience-bank optimizer for video anomaly judgment. [...] Given video-level ground truth and preference pairs of Learner trajectories, diagnose where the current experience bank supports or departs from better judgments, then maintain these three levels conservatively. [...]

Inputs. Current capacity: ⟨*Bank capacity status*⟩. Video-level ground truth: ⟨*Video label Y and categories C*⟩. Valid preference pairs: ⟨*Selected trajectory pairs $\mathcal{P}_r$*⟩.

Experience bank $\mathcal{E}_r$ 🔥

Compare trajectories. [...] Compare the observation, reasoning, and results of chosen and rejected against the ground truth and reward components. Identify why the key judgment in chosen is more sound and the

specific full-sequence review, change-awareness, key-evidence, boundary, event-process, or local-evidence error in rejected.

Diagnose experience. Audit current experiences relevant to that difference. For each relevant level, state whether an existing experience supports chosen, **misses a necessary condition, is too narrow or too broad**, or would steer the Learner toward rejected; cite the relevant experience ID when applicable.

Decide edits. Decide whether to retain, **minimally update, add, delete**, or make no change, and explain how that decision would move future Learner judgments toward chosen and away from rejected. When no reliable reusable conclusion exists, explicitly explain why no operation is warranted.

[...] Extract only differences that can be stated as a general judgment rule, event pattern, or visual clue. Do not write experience that depends on the current video name, frame indices, a single character's appearance, or incidental scene details. Any pair may support experience at any level or no experience at all; do not force coverage of all three levels.

[...] When there is no reliable new knowledge or basis for modification, **operations may be empty**; do not operate merely to fill a count, cover levels, or fill capacity. Three is only a maximum. [...]

B Policy Analysis and Proofs

B.1 Fixed-round conventions

To characterize one experience update, fix the distribution ν of $z = (x, Y, C)$, the reward function R, and the bank-context construction. Write $J(\mathcal{E}) = \mathbb{E}_{z\sim\nu, o\sim\pi_{\mathcal{E}}(\cdot|x)}[R(z, o)]$ for expected reward under these fixed conditions. The learner conditions on x; training labels enter the reward and the analytical reference. Let $p = \pi_{\mathcal{E}'}$ and $\pi_r = \pi_{\mathcal{E}_r}$. Required integrals are finite, and $p \ll \pi_r$ holds ν-almost surely. Likelihood ratios must use the actual decoding distributions, including any truncation.

Any bank dependence in the reward evaluator or observation sampler is frozen at $\mathcal{E}_r$ throughout the comparison, so changes in J isolate the effect of the edited policy. Allowing these components to change would require corresponding reward or distribution-shift terms. The implemented editor receives reward-ranked trajectory pairs; the continuous representations, standardized statistics, and reference policies below are analytical tools for characterizing its response.

The hierarchy-ordered bank forms textual context $c_{\mathcal{E}}$. The complete-output policy of the frozen model, including termination, is

$$\pi_{\mathcal{E}}(o \mid x) = \prod_{\ell=1}^{L+1} p_{\theta_0}\big(o_\ell \mid x, c_{\mathcal{E}}, o_{<\ell}\big), \qquad o_{L+1} = \mathrm{EOS}, \tag{7}$$

For $p = \pi_{\mathcal{E}'}$, write $\rho(x, o) = p(o \mid x)/\pi_r(o \mid x)$. Define the population baseline as $b_r(z) = \mathbb{E}_{\pi_r}[R(z, o) \mid z]$ and the advantage as $A_r(z, o) = R(z, o) - b_r(z)$.

For a finite group of M valid neutral trajectories, use the population-divisor convention for the sample variance and a positive stabilizer ε:

$$\begin{aligned} &R_i^0 = R(z, o_i^0), \qquad \mu_r = M^{-1}\sum_i R_i^0, \quad \sigma_r = \sqrt{M^{-1}\sum_i (R_i^0 - \mu_r)^2}, \\ &a_i^0 = R_i^0 - \mu_r, \quad \widehat{A}_i^{\mathrm{grp}} = \frac{a_i^0}{\sigma_r + \varepsilon}. \end{aligned} \tag{8}$$

These statistics express the reward differences within a neutral group and their common scale.

B.2 Context-induced policy changes

An experience edit can change output preferences while the backbone remains frozen. To describe this response locally, fix a finite collection $\{(x_k, o_k)\}_{k=1}^K$ and write $\ell_k(\theta, h) = \log \pi_{\theta,h}(o_k \mid x_k)$. Assume that the old and edited contexts admit a common fixed-dimensional continuous representation h and that these log-probabilities are twice continuously differentiable near (θ_0, h_r), with bounded Hessians and positive output probabilities. The derivatives apply to this continuous relaxation; contexts with different lengths or masks must also admit the common smooth representation.

Local response to context editing. With θ_0 fixed and $h_{r+1} = h_r + \Delta h$, Taylor's theorem gives, for each pair,

$$\begin{aligned}\ell_k(\theta_0, h_{r+1}) - \ell_k(\theta_0, h_r) = \nabla_h \ell_k(\theta_0, h_r)^\top \Delta h \\ + \frac{1}{2}\Delta h^\top H_k(\xi_k)\Delta h,\end{aligned} \tag{9}$$

where H_k is the context Hessian and ξ_k lies on the segment between the two representations. Write $\boldsymbol{\ell}(\theta, h) = (\ell_k(\theta, h))_{k=1}^K$ and $J_h = \nabla_h \boldsymbol{\ell}(\theta_0, h)|_{h_r}$. Stacking these expansions yields

$$\boldsymbol{\ell}(\theta_0, h_{r+1}) - \boldsymbol{\ell}(\theta_0, h_r) = J_h \Delta h + O(\|\Delta h\|^2). \tag{10}$$

If $\|H_k\|_2 \le B$ throughout this segment, the Euclidean norm of the stacked remainder is at most $\sqrt{K}B\|\Delta h\|^2/2$. For two outputs of the same input, subtracting their expansions yields the corresponding first-order change in preference log odds. Thus context edits can perturb output preferences without changing model weights. The linguistic optimizer proposes discrete text operations without computing J_h.

Conditional correspondence to parameter adaptation. For comparison, perturbing parameters at the old context gives

$$\boldsymbol{\ell}(\theta_0 + \Delta\theta, h_r) - \boldsymbol{\ell}(\theta_0, h_r) = J_\theta \Delta\theta + O(\|\Delta\theta\|^2). \tag{11}$$

Here J_h and J_θ are the context and parameter Jacobians at (θ_0, h_r). A parameter perturbation reproduces the context-induced first-order change exactly when $J_\theta \Delta\theta = J_h \Delta h$ is solvable, equivalently when $J_h \Delta h \in \mathrm{col}(J_\theta)$. More generally, the Moore–Penrose pseudoinverse yields the least-squares solution and residual

$$\begin{aligned}\Delta\theta^\star = J_\theta^\dagger J_h \Delta h, \\ \min_{\Delta\theta} \|J_h\Delta h - J_\theta \Delta\theta\|_2 = \|(I - J_\theta J_\theta^\dagger) J_h \Delta h\|_2.\end{aligned} \tag{12}$$

Here $J_\theta J_\theta^\dagger$ is the orthogonal projector onto $\mathrm{col}(J_\theta)$. For the chosen outputs and sufficiently small perturbations, the difference between the actual log-probability changes additionally contains terms of order $\|\Delta h\|^2 + \|\Delta\theta^\star\|^2$. The column-space condition thus identifies the local policy responses that context editing and parameter adaptation can both produce.

Supplementary ICL interpretation. Following Dai et al. (2023), fix one attention head and one query $q \in \mathbb{R}^{d_k}$. Let $K_\mathcal{E}, V_\mathcal{E}$ collect the experience-token keys and values, and K_X, V_X those of the remaining visible tokens. Write $K = [K_\mathcal{E}, K_X]$ and $V = [V_\mathcal{E}, V_X]$, with token columns $k_j \in \mathbb{R}^{d_k}$ and $v_j \in \mathbb{R}^{d_v}$. Only positions visible to this query under the attention mask are included. The usual head output is $\mathcal{F}_\mathcal{E}(q) = V\,\mathrm{softmax}(K^\top q/\sqrt{d_k})$. Replacing softmax attention by the unnormalized linear map, and absorbing the fixed scale into the values, gives

$$\begin{aligned}\widetilde{\mathcal{F}}_\mathcal{E}(q) &= [V_\mathcal{E}, V_X][K_\mathcal{E}, K_X]^\top q \\ &= V_X K_X^\top q + V_\mathcal{E} K_\mathcal{E}^\top q \\ &= \big(W_X + \Delta W_\mathcal{E}\big)q, \qquad W_X = V_X K_X^\top, \quad \Delta W_\mathcal{E} = \sum_j v_j k_j^\top.\end{aligned} \tag{13}$$

Holding q, K_X, V_X fixed, an edit from $\mathcal{E}_r$ to $\mathcal{E}_{r+1}$ changes the relaxed head output by

$$\widetilde{\mathcal{F}}_{\mathcal{E}_{r+1}}(q) - \widetilde{\mathcal{F}}_{\mathcal{E}_r}(q) = \big(\Delta W_{\mathcal{E}_{r+1}} - \Delta W_{\mathcal{E}_r}\big)q. \tag{14}$$

The bank-token term $\Delta W_\mathcal{E}$ acts as an effective context operator with frozen model weights. This decomposition holds for the stated linear relaxation with fixed q, K_X, V_X; in a full transformer, earlier layers can change these representations and softmax couples attention weights through normalization.

Correspondence to a linear-layer update. For a linear map $y_i = W k_i$ and a differentiable loss $L = \sum_i \ell_i(y_i)$, one gradient-descent step evaluated at W_0 gives

$$\begin{aligned}g_i &= \nabla_{y_i} \ell_i(y_i)|_{y_i = W_0 k_i}, \\ \Delta W_{\mathrm{GD}} &= -\eta \sum_i g_i k_i^\top, \qquad (W_0 + \Delta W_{\mathrm{GD}})q = W_0 q - \eta \sum_i g_i (k_i^\top q).\end{aligned} \tag{15}$$

The context term $\sum_j v_j(k_j^\top q)$ has the same outer-product form, with exact equality of update terms if the keys match and $v_j = -\eta g_j$. For a bank of linguistic rules, reward-guided comparisons provide the feedback from which the editor seeks useful context changes. The correspondence explains how those changes can alter computation while keeping the stored weights fixed.

B.3 Hypothesis constraints and decision coverage

Opposing premises address a limitation of neutral sampling: a confident policy may repeatedly produce only one decision. Condition on fixed observations, the old bank, and the validity of each of G independently generated neutral outputs. Let $u_r = \Pr_{\pi_r}(\widehat{Y} = 1 \mid x, \text{valid})$, where $\widehat{Y}$ is the binary anomaly decision. The two events missing one decision class are "all abnormal," with probability u_r^G, and "all normal," with probability $(1-u_r)^G$. They are disjoint for $G \geq 1$, so their complement has probability $1 - u_r^G - (1-u_r)^G$, which approaches zero as u_r approaches zero or one. If only M independent valid outputs remain, the same argument uses M in place of G.

A valid normal branch has an empty result list, and a valid abnormal branch has a nonempty list. Their joint validity therefore ensures both decisions are available for comparison regardless of u_r, while explanation quality and reward separation determine their usefulness. Video labels assign the factual and counterfactual roles, and categories constrain the abnormal branch during training. Both branches interpret the same visual evidence; counterfactual here denotes an opposing interpretive premise, not a causal intervention.

B.4 Group-relative feedback and pairwise directions

Reward differences specify which judgments an edit should favor. To connect these comparisons to a local policy direction, fix $M \geq 2$ valid neutral trajectories with rewards R_i^0, mean μ_r, and scale $s_r = \sigma_r + \varepsilon > 0$ as defined in Equation (8). Let π_h denote the relaxed context-conditioned policy, with $\pi_{h_r} = \pi_r$, and assume positive probabilities and differentiability near h_r. Define the sequence-level reward surrogate

$$\mathcal{L}_{\text{grp}}(h) = \frac{1}{M}\sum_{i=1}^{M} \frac{\pi_h(o_i^0 \mid x)}{\pi_r(o_i^0 \mid x)} \frac{R_i^0 - \mu_r}{s_r}. \tag{16}$$

Holding the sampled outputs, rewards, mean, and scale fixed, and writing $v_i = \nabla_h \log \pi_h(o_i^0 \mid x)\big|_{h_r}$, differentiation at the current context gives

$$\nabla_h \mathcal{L}_{\text{grp}}|_{h_r} = \frac{1}{M s_r}\sum_i (R_i^0 - \mu_r) v_i. \tag{17}$$

The vector-valued pair sum expands as

$$\begin{aligned}\sum_{i<j}(R_i^0 - R_j^0)(v_i - v_j) &= M\sum_i R_i^0 v_i - \Big(\sum_i R_i^0\Big)\Big(\sum_i v_i\Big) \\ &= M\sum_i (R_i^0 - \mu_r) v_i.\end{aligned} \tag{18}$$

Dividing by $M^2 s_r$ gives the pairwise identity

$$\nabla_h \mathcal{L}_{\text{grp}}|_{h_r} = \frac{1}{M^2 s_r}\sum_{i<j}(R_i^0 - R_j^0)\,\nabla_h \log \frac{\pi_h(o_i^0 \mid x)}{\pi_h(o_j^0 \mid x)}\bigg|_{h_r}. \tag{19}$$

Each term favors increasing the log odds of a higher-reward trajectory relative to a lower-reward one. When all rewards agree, the gradient is zero and TD supplies no preference. In a twice differentiable local relaxation, the connection to an edit direction is

$$\begin{aligned}\mathcal{L}_{\text{grp}}(h_r + \Delta h) - \mathcal{L}_{\text{grp}}(h_r) = \frac{1}{M^2 s_r}\sum_{i<j}(R_i^0 - R_j^0)(v_i - v_j)^\top \Delta h \\ + O(\|\Delta h\|^2).\end{aligned} \tag{20}$$

A direction aligned with the weighted preference gradients gives a positive first-order surrogate change. This identity provides an analytical basis for using reward-ranked judgment comparisons as editing feedback.

Equation (16) isolates an unclipped complete-trajectory reward surrogate. The full GRPO objective (Shao et al., 2024) uses token-level ratios, clipping, sequence-length averaging, and a reference-policy KL term. The pairwise identity applies to the full neutral group, from which TD retains one comparison. Retaining a sparse subset does not ensure the same direction. HC/HCal supply additional, differently conditioned references rather than iid neutral samples. The editor receives relation-tagged comparisons and rewards, without computing standardized advantages or numerical policy gradients. The following analyses relate finite edits and sparse comparisons to changes in expected reward.

B.5 Relative improvement and finite policy changes

Discrete text edits can produce finite policy changes. Likelihood ratios characterize their reward effect directly, extending the local directional interpretation above.

Proposition B.1 (Relative improvement and pairwise comparisons) *If $p(\cdot \mid x) \ll \pi_r(\cdot \mid x)$ almost surely and the required expectations are finite, then*

$$J(\mathcal{E}') - J(\mathcal{E}_r) = \mathbb{E}_{\nu,\pi_r}[(\rho - 1)A_r]. \tag{21}$$

For any fixed neutral group of size $M \geq 2$, writing $\rho_i = \rho(x, o_i^0)$ gives the algebraic identity

$$\frac{1}{M}\sum_i a_i^0(\rho_i - 1) = \frac{1}{M^2}\sum_{i<j}(R_i^0 - R_j^0)(\rho_i - \rho_j). \tag{22}$$

Proof. For fixed z, absolute continuity gives $\mathbb{E}_{\pi_r}[\rho \mid z] = 1$ and

$$\mathbb{E}_p[R \mid z] - \mathbb{E}_{\pi_r}[R \mid z] = \mathbb{E}_{\pi_r}[(\rho - 1)R \mid z]. \tag{23}$$

Because $b_r(z)$ does not depend on o, $\mathbb{E}_{\pi_r}[(\rho-1)b_r(z) \mid z] = 0$. Subtracting this term and averaging over $z \sim \nu$ yields Equation (21). The argument applies to any output-independent conditional baseline; it makes no assumption about differentiability of $\mathcal{E}$.

For a fixed finite group, abbreviate $R_i = R_i^0$ and $\mu = M^{-1}\sum_i R_i$. The pair sum expands as

$$\begin{aligned}\sum_{i<j}(R_i - R_j)(\rho_i - \rho_j) &= M\sum_i R_i\rho_i - \Big(\sum_i R_i\Big)\Big(\sum_i \rho_i\Big) \\ &= M\sum_i (R_i - \mu)\rho_i \\ &= M\sum_i (R_i - \mu)(\rho_i - 1).\end{aligned} \tag{24}$$

The last step uses $\sum_i(R_i-\mu) = 0$. Division by M^2 proves Equation (22). Dividing by a shared positive $\sigma_r + \varepsilon$ gives the standardized version and reweights the contributions of different groups. The identity expresses finite-group gain through reward differences and the probability shifts induced by the edit.

B.6 Proof of Proposition 3.1

Reference construction and preference directions. The improvement condition separates feedback quality from the policy response to an edit. Fix the feedback at step r and choose an analytical extension g_r on supported outputs, with $g_r(z, o^+) > g_r(z, o^-)$ for each retained preference. Finite rankings determine these directions but leave gap magnitudes and values on unobserved outputs unspecified. The ideal choice is $g_r = A_r$, for which the proxy errors vanish.

For two supported outputs at the same z, taking logarithms in Equation (4) and subtracting cancels the log-normalizer:

$$\log\frac{q_r^\star(o^+ \mid z)}{q_r^\star(o^- \mid z)} - \log\frac{\pi_r(o^+ \mid x)}{\pi_r(o^- \mid x)} = \frac{g_r(z, o^+) - g_r(z, o^-)}{\beta}. \tag{25}$$

Thus the reference increases the log odds of the preferred output. Hypothesis-generated trajectories are evaluated under the common neutral policy π_r, provided they belong to its support. The distances below measure movement toward this reference, while the proxy errors measure its agreement with the population advantage.

KL decomposition and reward bound. The distances and proxy errors in the proposition are

$$\begin{aligned} d_{\text{old}} &= \mathbb{E}_\nu D_{\text{KL}}(\pi_r \| q_r^\star), \quad d_{\text{new}} = \mathbb{E}_\nu D_{\text{KL}}(p \| q_r^\star), \\ e_{\text{old}} &= \mathbb{E}_{\nu,\pi_r} |g_r - A_r|, \quad e_{\text{new}} = \mathbb{E}_{\nu,p} |g_r - A_r|. \end{aligned} \tag{26}$$

For a support-defined proxy g_r and $\beta > 0$, consider conditional distributions $q(\cdot \mid z) \ll \pi_r(\cdot \mid x)$ and the regularized functional

$$\mathcal{F}_r(q) = \mathbb{E}_{\nu,q}[g_r] - \beta \mathbb{E}_\nu D_{\text{KL}}(q \| \pi_r). \tag{27}$$

Assume finite expected $|\log Z_r|$, the KL terms used below, and the relevant reward and proxy expectations. Taking the logarithm of Equation (4) gives

$$\log \frac{q_r^\star(o \mid z)}{\pi_r(o \mid x)} = \frac{g_r(z,o)}{\beta} - \log Z_r(z). \tag{28}$$

For any q with finite terms, use $\log(q/q_r^\star) = \log(q/\pi_r) - \log(q_r^\star/\pi_r)$ and take expectations:

$$\mathbb{E}_\nu D_{\text{KL}}(q \| q_r^\star) = \mathbb{E}_\nu D_{\text{KL}}(q \| \pi_r) - \frac{1}{\beta} \mathbb{E}_{\nu,q}[g_r] + \mathbb{E}_\nu[\log Z_r]. \tag{29}$$

Rearrangement yields

$$\mathcal{F}_r(q) = \beta \mathbb{E}_\nu[\log Z_r] - \beta \mathbb{E}_\nu D_{\text{KL}}(q \| q_r^\star). \tag{30}$$

When $q_r^\star$ has finite expected proxy and KL terms, it maximizes this functional over the conditional distributions under consideration. The realizable bank policies may form a smaller family. The candidate bound follows by evaluating Equation (29) at p and π_r, whether or not a bank realizes $q_r^\star$.

Specifically, evaluating Equation (29) at these two policies gives

$$\begin{aligned} d_{\text{new}} &= \mathbb{E}_\nu D_{\text{KL}}(p \| \pi_r) - \frac{1}{\beta} \mathbb{E}_{\nu,p}[g_r] + \mathbb{E}_\nu[\log Z_r], \\ d_{\text{old}} &= -\frac{1}{\beta} \mathbb{E}_{\nu,\pi_r}[g_r] + \mathbb{E}_\nu[\log Z_r]. \end{aligned} \tag{31}$$

Subtracting eliminates the common log-normalizer and gives

$$\mathbb{E}_{\nu,p}[g_r] - \mathbb{E}_{\nu,\pi_r}[g_r] = \beta \left(d_{\text{old}} - d_{\text{new}} + \mathbb{E}_\nu D_{\text{KL}}(p \| \pi_r) \right). \tag{32}$$

To pass from proxy gain to true reward gain, define the signed error $\epsilon_r(z,o) = g_r(z,o) - A_r(z,o)$. The baseline in A_r is output-independent, so

$$\begin{aligned} J(\mathcal{E}') - J(\mathcal{E}_r) = \mathbb{E}_{\nu,p}[g_r] - \mathbb{E}_{\nu,\pi_r}[g_r] \\ - \big(\mathbb{E}_{\nu,p}[\epsilon_r] - \mathbb{E}_{\nu,\pi_r}[\epsilon_r] \big). \end{aligned} \tag{33}$$

The last parenthesis is at most $\mathbb{E}_{\nu,p}|\epsilon_r| + \mathbb{E}_{\nu,\pi_r}|\epsilon_r| = e_{\text{new}} + e_{\text{old}}$. Combining this inequality with Equation (32) gives

$$\begin{aligned} J(\mathcal{E}') - J(\mathcal{E}_r) \geq \beta \left(d_{\text{old}} - d_{\text{new}} + \mathbb{E}_\nu D_{\text{KL}}(p \| \pi_r) \right) \\ - e_{\text{old}} - e_{\text{new}}. \end{aligned} \tag{34}$$

Dropping the nonnegative KL term proves Equation (5). Under the stronger uniform assumption $|g_r - A_r| \leq \epsilon_A$, both error terms are at most ϵ_A, recovering the $2\epsilon_A$ form.

Movement toward the preference reference has a direct objective interpretation:

$$\mathcal{F}_r(p) - \mathcal{F}_r(\pi_r) = \beta(d_{\text{old}} - d_{\text{new}}). \tag{35}$$

Distance reduction equals a regularized proxy-objective gain. When this gain exceeds $e_{\text{old}} + e_{\text{new}}$, expected reward improves. The condition makes two roles explicit: comparison construction targets reliable feedback, and experience editing must induce a sufficient policy response to that feedback.

B.7 From selected comparisons to the full policy change

An editor sees a small set of comparisons, while a bank edit can change probabilities across the output distribution. To isolate the contribution of retained feedback, let $\mathcal{I} = \{(i,j) : i < j\}$ be all neutral pairs, $\mathcal{S} \subseteq \mathcal{I}$ the selected pairs, and $D_{ij} = (R_i^0 - R_j^0)(\rho_i - \rho_j)$. The group statistic $\widehat{L}_M = M^{-1} \sum_i a_i^0(\rho_i - 1)$ decomposes as

$$\widehat{L}_M = \underbrace{\frac{1}{M^2} \sum_{(i,j)\in\mathcal{S}} D_{ij}}_{L_\mathcal{S}\text{: selected comparisons}} + \underbrace{\frac{1}{M^2} \sum_{(i,j)\in\mathcal{I}\setminus\mathcal{S}} D_{ij}}_{B_\mathcal{S}\text{: unselected comparisons}} . \tag{36}$$

The decomposition holds for adaptively selected pairs as well. Define the observed likelihood-ratio spread $\omega_p = \max_i \rho_i - \min_i \rho_i$. Since $|\rho_i - \rho_j| \le \omega_p$, the triangle inequality gives

$$\begin{aligned} |B_\mathcal{S}| &\le \frac{1}{M^2} \sum_{(i,j)\in\mathcal{I}\setminus\mathcal{S}} |R_i^0 - R_j^0|\,|\rho_i - \rho_j| \\ &\le \frac{\omega_p}{M^2} \sum_{(i,j)\in\mathcal{I}\setminus\mathcal{S}} |R_i^0 - R_j^0|. \end{aligned} \tag{37}$$

Consequently,

$$\widehat{L}_M \ge L_\mathcal{S} - \frac{\omega_p}{M^2} \sum_{(i,j)\in\mathcal{I}\setminus\mathcal{S}} |R_i^0 - R_j^0|. \tag{38}$$

Selected-pair gains exceeding the residual bound certify a positive finite-group change. Each contribution depends on both reward separation and edit-induced probability changes, requiring likelihoods under both banks. This explains why feedback selection must be considered together with the learner's response to an edit.

For example, consider three outputs at a fixed input with rewards $(R_1, R_2, R_3) = (1, 1/2, 0)$ and policies $\pi_r = (1/3, 1/3, 1/3)$ and $p = (2/5, 1/10, 1/2)$. The odds of the preferred o_1 over o_2 rise from 1 to 4, yet expected reward falls from $1/2$ to $9/20$ as more probability moves to the zero-reward output. For a group containing each output once and $\mathcal{S} = \{(1,2)\}$, the selected-pair gain $L_\mathcal{S} = 1/20$ is outweighed by $B_\mathcal{S} = -1/10$, giving $\widehat{L}_M = -1/20$.

B.8 Finite-group bias and adaptive editing

Finite rollout groups also affect estimation of reward gain. Fix z and a candidate p independent of M iid samples from π_r, with $M \ge 2$. If validity filtering changes the sampling law, use the corresponding conditional policies and likelihood ratios; the resulting gain then concerns conditional expected reward. Define

$$\widehat{L}_M = \frac{1}{M} \sum_i (R_i - \bar{R})(\rho_i - 1), \qquad \Delta_z = \mathbb{E}_p[R \mid z] - \mathbb{E}_{\pi_r}[R \mid z]. \tag{39}$$

The fixed-candidate assumptions give $\mathbb{E}[\rho_i - 1 \mid z] = 0$ and $\mathbb{E}[R_i(\rho_i - 1) \mid z] = \Delta_z$. By independence of different samples,

$$\mathbb{E}[\bar{R}(\rho_i - 1) \mid z] = \frac{1}{M}\mathbb{E}[R_i(\rho_i - 1) \mid z] = \frac{\Delta_z}{M}. \tag{40}$$

Consequently,

$$\mathbb{E}[\widehat{L}_M \mid z] = \left(1 - \frac{1}{M}\right) \Delta_z. \tag{41}$$

Multiplying by $M/(M-1)$, equivalently replacing the baseline for sample i by the leave-one-out mean, removes this centering bias for a fixed candidate. Choosing $\mathcal{E}'$ from the same group, normalizing by a random standard deviation, and selecting relations introduce additional dependence beyond this correction. Population-improvement assessment therefore requires control of estimation error and edit-selection dependence. Independent evaluation separates measurement from edit selection; reusing selection samples requires an adaptive-analysis argument.

B.9 LIKELIHOOD DIAGNOSTICS BEYOND THE EDITING INPUTS

A pairwise likelihood change provides an observable measure of the policy response to an edit. For fixed observations and a preference pair (o^+, o^-) with positive probabilities under both banks, define

$$\begin{aligned}\Delta_{\text{pair}} &= \log \frac{\pi_{\mathcal{E}'}(o^+ \mid x)}{\pi_{\mathcal{E}'}(o^- \mid x)} - \log \frac{\pi_r(o^+ \mid x)}{\pi_r(o^- \mid x)} \\ &= \log \rho(x, o^+) - \log \rho(x, o^-).\end{aligned} \tag{42}$$

Positive Δ_{pair} indicates that the edit increases the preferred output's relative likelihood. Its computation uses complete-sequence log-probabilities, including termination, with consistent tokenization and conditioning. This requires token probabilities rather than frame-level anomaly scores or token-averaged likelihoods.

Applied to edit-input pairs, this quantity measures responsiveness to supplied feedback. Applied to neutral pairs whose observations and reward comparisons are fixed independently of the candidate edit and excluded from edit selection, it assesses whether the response extends beyond that feedback. These measurements describe the evaluated outputs; the KL distances and proxy errors in Proposition 3.1 concern the full output support.

C ADDITIONAL EXPERIMENTAL ANALYSIS

C.1 EVALUATION PROTOCOLS AND DETAILED RESULTS

Evaluation settings. Unless varied, controlled XD-Violence experiments use Qwen3.5-9B at r448 resolution, 140 training videos, and 16 frames per video, with the remaining defaults in Appendix A. XD-Violence results are from single runs unless stated otherwise, with a shared full-method reference bank. Our MSAD main benchmark and scene-wise results report means and sample SDs across three runs. XD-Violence uses average precision (AP) and ROC-AUC on the same frame-level predictions. On MSAD, AP and AP_a are computed as trapezoidal PR-AUC.

Unless specified otherwise, our XD-Violence analyses use evaluation stride 8. Main benchmark and partial/full-data comparisons on XD-Violence, UCF-Crime, and UBnormal use stride 4, as does the data-efficiency analysis. Table 1 also reports stride-8 results. Learning-stability comparisons use the same evaluation settings for both methods at stride 16.

Published baselines retain their original evaluation protocols. We learn separate experience banks for XD-Violence and UCF-Crime and transfer only XD-learned Principles to UBnormal, without target training or bank updates.

Model families and capacity. Linguistic adaptation draws on the backbone's ability to revise and apply textual criteria. Backbone comparisons cover Qwen3.5 (Qwen Team, 2026), InternVL3.5 (Wang et al., 2025), GLM (Team et al., 2025), Keye (Yang et al., 2025a), and MiniCPM (Yu et al., 2025). Each model serves as both learner and optimizer and learns its own bank under the same video and frame settings.

Six of eight models improve on both AP and AUC (Table 6). GLM-4.6V-Flash gains 26.02 percentage points in AP, while Qwen3.5-27B reaches the highest final AP of 74.19%. Within Qwen3.5, the 9B and 27B models improve but the 2B and 4B models decline. The gains across families support adapting different frozen backbones through reusable textual criteria, while the variation across model scales shows that the benefit depends on the model used to generate, revise, and apply that experience.

Scene-wise performance on MSAD. Scene-wise evaluation examines whether the overall MSAD gain extends across environments. With experience learned from 20% of the training split, Anom-Ψ achieves the highest AP in all twelve listed scenes and the highest AUC in ten against the reported baselines (Ding et al., 2026) (Table 7). These gains span indoor, outdoor, and transport scenes. Warehouse remains challenging: its AUC is 76.63%, compared with 84.7% for EGO, despite having the highest AP among the listed methods. The Baseline column denotes a single zero-shot run under its own prompt and output protocol.

Table 6: Backbone comparisons on XD-Violence (AP/AUC, %). Each model learns its own experience bank. Shaded rows add Anom-Ψ; Δ measures the change from the paired baseline (pp).

(a) Model scale

Model	**AP (%)**	**ΔAP**	AUC (%)	ΔAUC
Qwen3.5-2B	44.01	–	77.84	–
+ *Anom*-Ψ	40.46	−3.55	74.51	−3.33
Qwen3.5-4B	57.40	–	84.89	–
+ *Anom*-Ψ	47.61	−9.79	82.09	−2.80
Qwen3.5-9B	57.14	–	85.57	–
+ *Anom*-Ψ	73.40	+16.26	90.30	+4.73
Qwen3.5-27B	62.78	–	88.36	–
+ *Anom*-Ψ	74.19	+11.41	92.52	+4.16

(b) Model families

Model	**AP (%)**	**ΔAP**	AUC (%)	ΔAUC
InternVL3.5-8B	60.77	–	87.58	–
+ *Anom*-Ψ	71.17	+10.40	90.09	+2.51
GLM-4.6V-Flash	42.98	–	73.11	–
+ *Anom*-Ψ	69.00	+26.02	88.41	+15.30
Keye-VL-1.5-8B	53.63	–	84.14	–
+ *Anom*-Ψ	60.52	+6.89	85.52	+1.38
MiniCPM-V-4.5	53.73	–	82.09	–
+ *Anom*-Ψ	62.94	+9.21	88.03	+5.94

Table 7: Scene-wise results on MSAD (%). MGFN/MGFN* use SwinT/I3D backbones. Bold/underline mark the best/second-best scores. Our Overall scores use the full test set.

Scene	RTFM		MGFN		MGFN*		UR-DMU		EGO		Baseline		**Anom-Ψ (Ours)**	
	AUC	AP	AUC	AP	AUC	AP	AUC	AP	AUC	AP	AUC	AP	AUC	AP
Frontdoor	81.8	79.3	59.5	51.7	82.5	80.8	84.8	82.8	85.2	81.6	70.50	69.78	**85.67** ± 0.45	**84.53** ± 2.11
Mall	88.1	76.6	18.5	20.1	73.8	71.3	91.0	83.8	82.3	73.4	64.21	59.55	**96.10** ± 3.72	**91.69** ± 9.45
Office	76.6	72.8	64.1	52.3	71.5	58.2	77.8	67.3	80.0	71.7	72.54	70.80	**86.54** ± 3.35	**83.00** ± 3.95
Parkinglot	80.7	45.8	67.9	19.0	68.9	14.8	91.4	53.9	96.8	75.2	78.58	44.91	**99.39** ± 0.07	**94.89** ± 1.00
Pedestr. st.	94.0	48.5	75.9	9.7	94.8	36.2	81.9	11.5	97.5	52.0	89.32	46.13	**99.46** ± 0.17	**89.14** ± 7.65
Restaurant	88.3	79.1	67.9	44.0	95.1	91.3	93.1	87.4	94.3	73.9	81.48	73.25	**97.90** ± 0.37	**95.07** ± 1.48
Road	84.3	57.9	70.6	26.3	76.5	35.8	83.0	64.4	**89.8**	64.6	73.59	45.34	89.53 ± 0.14	**76.83** ± 0.93
Shop	85.3	75.6	62.7	43.0	85.6	78.4	81.3	64.5	83.4	72.2	63.77	53.80	**87.08** ± 3.53	**84.84** ± 0.78
Sidewalk	88.3	68.8	69.0	25.9	78.5	57.2	86.5	64.1	87.1	45.0	68.68	37.10	**96.59** ± 0.40	**84.42** ± 2.16
St. highview	72.0	28.5	75.3	23.3	77.9	29.3	85.0	37.7	28.2	10.1	86.38	62.12	**98.20** ± 0.04	**96.33** ± 0.29
Train	51.4	3.3	65.4	5.2	40.3	2.1	59.0	3.1	80.8	7.8	67.93	22.50	**98.85** ± 0.41	**66.71** ± 8.61
Warehouse	82.7	57.0	70.1	30.1	58.3	24.2	81.2	59.1	**84.7**	46.6	65.61	40.18	76.63 ± 2.35	**71.01** ± 5.73
Overall	86.6	68.4	69.3	33.6	81.2	59.3	85.0	68.3	87.3	64.4	72.61	50.69	**91.28** ± 0.71	**82.55** ± 1.20

C.2 Component Ablations and Learning Stability

Learning mechanisms and trajectory controls. The controls examine whether explicit comparisons improve experience learning (Table 3). All learning variants use six trajectories. Row 6 uses a flat bank, while the other learning rows use hierarchical banks. Without Preference, the editor learns from individual trajectories and video-level labels. Disabled video/frame selectors use category-balanced random videos and uniform frames.

The trajectory controls in Table 4(c) compare complete generation-and-pairing configurations. The $4N + 2D$ control supplies at most one TD pair, while the opposing-hypothesis configuration allows up to three relations. Under the same six-trajectory generation budget, the opposing-hypothesis configuration achieves the highest AP and AUC.

Pair-selection controls examine whether covering the comparison relations is sufficient to obtain useful feedback. Random and Largest-gap draw from the same valid, positive-gap candidate pool; Best-anchor fixes the factual trajectory as chosen, and Stratified random samples one pair per relation. All strategies allow at most three pairs per video (Table 4(b)). The default relation-specific selection raises AP from 66.83% for Stratified random to 73.40%, supporting the importance of which judgments are compared within each relation.

Stability across learning runs. Repeated learning runs assess whether the gain persists across different learned banks. Across five independent runs per method on XD-Violence, Anom-Ψ achieves $65.36 \pm 0.90\%$ AP versus $58.42 \pm 4.70\%$ for TD-only, improving both mean performance and consistency (mean ± sample SD, Figure 5(c)). Boxes span the interquartile range with medians marked, whiskers cover the observed range, and dots show individual runs.

Table 8: Video and frame selection on XD-Violence. (a) Fixed Hybrid frame-selection strategy. (b) Fixed wavelet-prioritized videos and 16 frames per video; Uniform+Top replaces the WFS component with uniform sampling. Shading marks Anom-Ψ.

(a) Video selection

Strategy	AP (%)	AUC (%)
Global random	68.73	89.27
Balanced random	70.39	89.07
Wavelet-prioritized	**73.40**	**90.30**

(b) Frame selection and hybrid composition

Strategy	AP (%)	AUC (%)	Strategy	AP (%)	AUC (%)
Random	66.81	87.15	Top-score	67.75	87.94
Uniform	69.75	87.40	Uniform+Top	68.40	86.99
WFS	70.28	89.60	Hybrid	**73.40**	**90.30**

Table 9: Training-score source comparison on XD-Violence. Scores use evaluation stride 16 and fixed smoothing. Training videos are fixed; each source recomputes training frames and score-based rewards.

Training-score source	AP (%)	AUC (%)
HyperVD	**64.80**	86.35
MACIL_SD	62.69	85.47
UR-DMU	61.53	84.92
Qwen3.5-9B	64.27	**87.23**

Video and frame selection. A balanced subset can still contain videos with little temporal change. With 140 videos and a fixed Hybrid frame-selection strategy, wavelet-prioritized selection improves AP over global-random and category-balanced random selection by 4.67 and 3.01 percentage points, respectively (Table 8(a)). The latter comparison supports prioritizing change-rich videos within each category in addition to balancing category counts.

High-score frames can omit the context needed to interpret an event. With sixteen frames and the same four-frame high-score component, replacing uniform contextual frames with WFS frames raises AP from 68.40% to 73.40% (panel (b)). The hybrid also outperforms either WFS or top-score selection alone, supporting their complementary roles in selecting observations for experience learning.

The score source determines the temporal signals available for frame selection and auxiliary rewards. We compare four sources with the training videos fixed, recomputing frames and score-based rewards for each source (Table 9). Qwen3.5-9B scores achieve the highest AUC (87.23%), with AP only 0.53 percentage points below the best result. These results support using frozen VLM scores for training-time frame selection and auxiliary feedback.

Experience levels and auxiliary rewards. The hierarchy retains general criteria, event progression, and local evidence at distinct levels. To examine their contributions, each removal keeps the remaining levels at capacity ten without reallocating the deleted capacity (Table 10). Removing any level lowers AP from 73.40% to 71.64%–72.07%, supporting the value of retaining all three levels; removing Pattern yields slightly higher AUC. The direct-learning controls use six neutral trajectories, individual-trajectory updates, category-balanced random videos, and uniform frames.

Auxiliary rewards provide temporal feedback beyond video-level accuracy. Each improves AP individually, and their combination raises AP from 70.30% to 73.40%, while AUC changes from 91.04% to 90.30% (Table 11). The joint reward thus favors the precision–recall objective used for XD-Violence evaluation.

C.3 Local-update Evaluation and Diagnostics

To assess whether hypothesis-derived conflicts yield more useful experience revisions, we compare single-step updates from the same old bank, training video, and observations (Figure 5(b)). Neutral-only uses six neutral trajectories and reward-ranked neutral comparisons; hypothesis-guided learning uses four neutral trajectories and one under each opposing premise, with the default TD/HC/HCal comparisons. Both share the backbone, rewards, bank capacity, and editor interface, with at most one editor call and three edit operations per step.

Table 10: Experience organization under direct learning and Anom-Ψ on XD-Violence. Gray parentheses give changes in percentage points from the flat bank within each learning scheme. P, M, and C denote Principle, Pattern, and Clue.

Learning scheme	Bank	P	M	C	AP (%)	AUC (%)
Direct learning	Flat bank	–	–	–	65.36	86.52
	Hierarchical bank	✓	✓	✓	64.78 (−0.58)	86.95 (+0.43)
Anom-Ψ	Flat bank	–	–	–	70.79	89.39
	w/o Principle	✗	✓	✓	71.82 (+1.03)	88.70 (−0.69)
	w/o Pattern	✓	✗	✓	71.64 (+0.85)	**90.72** (+1.33)
	w/o Clue	✓	✓	✗	72.07 (+1.28)	89.66 (+0.27)
	Hierarchical bank	✓	✓	✓	**73.40** (+2.61)	90.30 (+0.91)

Table 11: Contributions of change and evidence rewards on XD-Violence.

Variant	Accuracy	Change	Evidence	AP (%)	AUC (%)
Accuracy only	✓	✗	✗	70.30	**91.04**
+ Change Reward	✓	✓	✗	71.79	90.58
+ Evidence Reward	✓	✗	✓	71.53	89.45
Anom-Ψ (Ours)	✓	✓	✓	**73.40**	90.30

Evaluation uses a fixed subset of 35 videos from the official XD-Violence test split: five normal videos and five from each of the six anomaly categories. These videos do not overlap with the 140 training videos in the recorded update history. All 140 update steps, including ineffective updates, enter the paired analysis. The subset, step order, threshold, and boundary neighborhood were fixed before this diagnostic.

AP is computed over this subset. For FPR and recall, smoothed frame scores at or above 0.5 are classified as anomalous. Normal FPR measures false positives in normal videos; Adjacent FPR uses normal frames within two seconds before or after an annotated anomaly interval, excluding anomalous frames and clipping neighborhoods at video boundaries. Anomaly recall uses annotated anomalous frames. Adjacent FPR and anomaly recall are macro-averaged. Panel (b) reports mean updated-minus-old differences for AP and recall and old-minus-updated differences for FPR.

Hypothesis-guided updates yield a larger mean AP gain than neutral-only updates (0.42 versus 0.14 percentage points) and a larger reduction in Adjacent FPR (0.37 versus 0.17 percentage points), while also improving anomaly recall. Background rejection and event coverage therefore improve together in this paired comparison.

C.4 Data and Budget Sensitivity

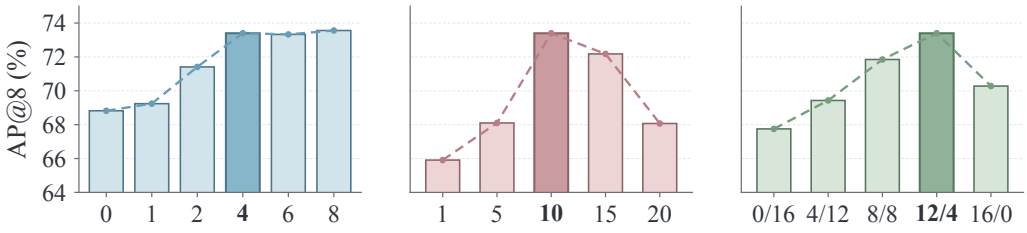
Adaptation data. We compare partial- and full-data learning to assess how much adaptation is needed for detection and transfer (Table 12). Using 3.5% of the XD-Violence training split trails full-data learning by only 1.81 percentage points in AP and 0.28 percentage points in transferred UBnormal AUC. These small gaps support learning useful criteria from a limited source-domain adaptation set. Figure 5(a) also includes published VERA (Ye et al., 2025) and LRPO (Leng et al., 2026) results as references across training budgets.

Training-set composition determines how often the learner encounters each event category. The 20% MSAD subset balances specific event categories, treating normal videos as an additional category, while full-data learning retains all 360 normal and 120 abnormal videos without rebalancing. The balanced subset improves both AUC and AP by 2.36 percentage points over full-data learning. This contrast suggests that category balance matters alongside data volume.

Rollout, capacity, and frame budgets. We vary the learning budgets to assess the value of additional sampling and experience capacity (Figure 6). Doubling neutral rollouts from $G = 4$ to $G = 8$ improves AP by only 0.16 percentage points (panel (a)), favoring the smaller rollout budget. AP peaks at a per-level capacity of $K_{\max} = 10$ and declines at larger capacities (panel (b)). Under the

Table 12: Partial-data versus full-data experience learning for Anom-Ψ. UBnormal transfers XD-learned experience; its training ratios refer to XD-Violence. Subscript a denotes evaluation on abnormal videos only.

Dataset	Training Ratio	AUC (%)	AUC$_a$ (%)	AP (%)	AP$_a$ (%)
XD-Violence	3.5%	92.20	82.56	78.43	80.76
	100%	**93.43**	**84.46**	**80.24**	**81.86**
UCF-Crime	9%	86.73	73.60	37.23	39.18
	100%	**88.08**	**75.97**	**43.20**	**46.09**
UBnormal	3.5% (XD)	77.01	**69.27**	82.40	**87.73**
	100% (XD)	**77.29**	69.09	**82.43**	87.42
MSAD	20%	**91.28** $\pm$ 0.71	**78.59** $\pm$ 1.15	**82.55** $\pm$ 1.20	**85.21** $\pm$ 1.20
	100%	88.92 $\pm$ 0.32	76.41 $\pm$ 0.42	80.19 $\pm$ 0.04	83.69 $\pm$ 0.17



(a) Neutral rollouts G (b) Capacity per level $K_{\max}$ (c) Frame composition

Figure 6: Parameter sensitivity on XD-Violence: (a) neutral rollouts, (b) capacity per experience level, and (c) WFS/top-score composition with 16 frames. Darker bars and bold ticks mark defaults.

fixed sixteen-frame budget, the default mixture of WFS and high-score frames achieves the highest AP among the tested compositions (panel (c)).

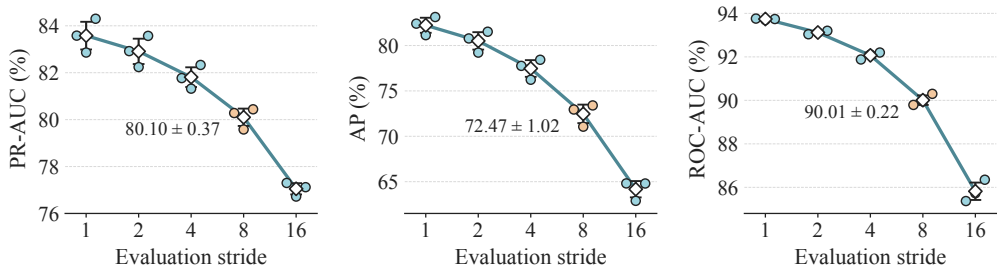
Evaluation stride. Temporal sampling determines which window predictions contribute to frame-level evaluation. We assess this sensitivity across three experience banks (Figure 7), reusing stride-1 predictions for strides 1/2/4 and subsampling stride-4 predictions for strides 8/16. Mean AP falls from 82.23% at stride 1 to 64.17% at stride 16, with ROC-AUC and trapezoidal PR-AUC following the same downward trend. The substantial variation highlights the importance of accounting for temporal evaluation resolution when comparing detection results.

C.5 Computational Cost and Test-time Selection

Comparison settings. Additional computation can be spent on learning reusable experience or selecting among test-time candidates. We compare these choices using Qwen3.5-9B on all 800 XD-Violence test videos. Self-selection generates six neutral candidates per window without experience, then uses the same frozen model to select an answer from the observations and candidate outputs. The oracle uses test-set ground truth to select the best of the same six candidates, serving as an empirical upper-bound reference for this fixed pool. Both learning methods use six trajectories per video over 140 training videos and keep their learned banks frozen for single-rollout testing.

Optimizing experience from conflicting judgments. The large gap between self-selection and oracle selection highlights the difficulty of choosing among generated alternatives. Anom-Ψ learns from reward-ranked conflicts between normal and abnormal interpretations during training and retains the resulting distinctions in a reusable experience bank. This enables it to reach 73.40% AP, only 1.15 percentage points below the oracle, using one rollout per window without consulting test-set ground truth at inference (Table 13). It also outperforms flat-bank learning and self-selection on both AP and AUC, supporting the value of optimizing experience before test-time decisions.

Runtime and measurement scope. The benefit of reusable experience also appears in test-time cost. Anom-Ψ takes an estimated 0.802 seconds per window, compared with 1.904 seconds for the oracle's six-candidate generation (Table 13). Its experience optimization takes 25.63 minutes. Runtime values are elapsed-time estimates normalized to a single GPU and the evaluation window



(a) PR-AUC (b) AP (c) ROC-AUC

Figure 7: Evaluation-stride sensitivity on XD-Violence. (a) trapezoidal PR-AUC, (b) AP, and (c) ROC-AUC across three experience banks per stride: diamonds indicate means, bars ± 1 population SD, and circles individual results. Peach highlights stride 8; adjacent labels report its mean$\pm$SD.

Table 13: Detection performance and computational cost on XD-Violence. Oracle selects the best of six candidates using test-set ground truth, providing an empirical upper-bound reference for this pool. Optim. reports experience optimization time; runtime and token counts are estimates per test window. Times are single-GPU equivalents; Total sums inference and selection.

Method	Performance (%)		Optim. (min)↓	Runtime (s)↓			Tokens↓	
	AP↑	AUC↑		Infer.	Select.	Total	Input	Output
Base (zero-shot)	57.14	85.57	Zero	0.434	Zero	0.434	3,223	153
Test-time self-selection	44.93	78.35	Zero	1.904	0.321	2.225	23,464	1,048
Test-time selection (oracle)	74.55	94.28	Zero	1.904	Zero	1.904	19,337	963
Flat-bank learning	65.36	86.52	32.13	0.744	Zero	0.744	4,859	301
Anom-Ψ (Ours)	**73.40**	**90.30**	25.63	0.802	Zero	0.802	5,627	331

count, with optimization reported separately; evaluation, offline scoring, and observation preparation are excluded.

Token estimates. Token counts provide a complementary measure of inference cost. We estimate input and output tokens with the local Qwen3.5-9B tokenizer, including visual tokens and, for self-selection, all candidate-generation and selection calls. These inference-only estimates exclude unlogged server overhead, failed requests, and retries. Anom-Ψ reduces mean input tokens per test window from 23,464 for self-selection to 5,627, with fewer output tokens as well. The single-rollout protocol therefore combines higher detection performance with lower test-time token use in this comparison.

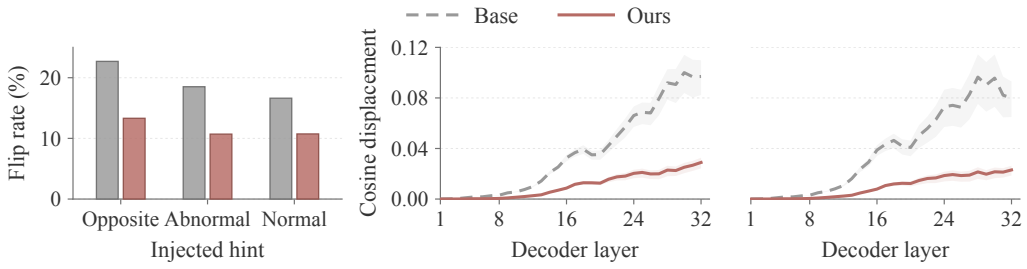
C.6 Sensitivity to Prompt Interference

A stable detection criterion should resist unsupported assertions in the prompt. We compare zero-shot inference and Anom-Ψ on XD-Violence with model weights and the learned bank fixed. Every window of a video receives the same hint: "This is a normal video.", "This is an anomalous video.", or the assertion opposite to its video-level label. Prediction changes are measured relative to each method's no-hint outputs, with a fixed threshold of 0.5 for flips and error transitions.

To relate output sensitivity to internal representations, we also use 2,000 paired windows, split equally between all-normal windows and windows containing anomalous frames. Each prompt condition receives the same sixteen frames. Within each full prompt configuration, we measure $d_\ell = 1 - \cos(h_\ell^{\text{clean}}, h_\ell^{\text{hint}})$ at decoder layer ℓ, using the hidden state of the last input token before generation. This captures the response to a hint relative to that configuration's own no-hint state.

Anom-Ψ has lower flip rates and smaller mean absolute score changes under all three hint types. Under label-opposite hints, the flip rate falls from 22.69% to 13.31% (Figure 8(a)), indicating more stable predictions in the presence of a conflicting assertion.

Error transitions reveal a class-dependent tradeoff. Under label-opposite hints, false-positive induction falls from 24.38% to 5.61%, while false-negative induction rises from 16.43% to 21.68%. Each rate is conditioned on the corresponding method's initially correct frames of that class. The



(a) Prediction flips (b) Normal hint (c) Abnormal hint

Figure 8: Prediction and representation sensitivity to injected hints on XD-Violence. (a) Frame prediction flip rates; (b,c) layer-wise cosine displacement under normal/abnormal hints, all relative to each configuration's no-hint state. Lines show means and bands the 25th–75th percentiles across windows.

other hint types show the same tradeoff: fewer normal frames are pushed toward anomaly, but more anomalous frames are pushed toward normality.

The learned configuration also has lower mean representation displacement at all 32 layers under both normal and abnormal assertions (Figure 8(b,c)). At the final layer, the reductions are 70.1% and 71.2%, respectively. These smaller shifts provide representation-level evidence consistent with the reduced prediction sensitivity.

D Visualizations and Case Studies

D.1 Experience Evolution and Final Bank

Similar visual cues can support different event interpretations, so useful experience must specify when each interpretation applies. Figure 9 shows how these conditions develop across learning. For vehicle events, intermediate experience emphasizes sudden loss of control and structural instability; the final bank combines these requirements with exclusions for predictable stunts in which the vehicle remains intact. For physical interactions, the evolving experience distinguishes uncontrolled force and retaliation from synchronized performances. Crowd-related entries use collective disorder and suppression context to determine when local violence belongs to a riot. Across these themes, the bank refines event requirements, exclusions, and the contextual basis for choosing among competing categories.

The final hierarchy connects these conditions to observable event evidence (Figure 10). A Principle distinguishes collision debris from evidence of an explosion; a Pattern specifies uncontrolled motion and structural failure; and a Clue links visible debris to loss of vehicle stability. Together, these entries make the classification depend on how debris arises and how the vehicle moves, encoding reusable distinctions between visually similar events.

D.2 Representation and Visual Attribution

Representation structure. Representations immediately before the first output token reveal how experience affects the state from which the model produces its answer. Base and Learned Experience use the same visual inputs, frozen backbone, and prompt instructions, differing only in the experience bank. On 2,000 paired XD-Violence windows, learned experience raises the high-dimensional silhouette score from 0.058 to 0.223, indicating greater separation between normal and anomalous observations. Figure 11 provides PHATE (Moon et al., 2019) views for XD-Violence and MSAD. The MSAD views use the same high-dimensional representations as the UMAP (McInnes et al., 2018) analysis in Figure 3. The increased separation provides complementary evidence that textual experience can improve discrimination without changing model weights.

Visual attribution under Base and Learned Experience contexts. Figure 12 aligns RGB video frames with two views of visual attribution. Inspired by FlashTrace (Pan et al., 2026), we use local IFR information-flow attribution $\mathcal{I}$, complemented by absolute Gradient $\times$ Activation $\mathcal{G}$. Both views

Step 20: $\mathcal{E}_{20}$

1. [...] If the vehicle survives the impact or the fireball is part of a predictable cinematic sequence without thermal energy release, classify as 'car_accident' or 'stunt' instead.
2. A transition from 'abuse' to 'fighting' is identified when the victim shifts from being pinned or controlled to actively grappling, punching, or resisting the aggressor. [...]
3. [...] Do not classify as 'normal' or 'march' solely because the crowd maintains directional flow or cohesion; the presence of environmental aggression overrides orderly movement.

...

Step 60: $\mathcal{E}_{60}$

1. A 'car_accident' pattern is confirmed by a sudden transition from stable driving to uncontrolled motion, collision, or structural failure. [...]
2. [...] The 'shooting' pattern is only triggered by the sudden, discrete transition to a visible discharge event (flash, smoke) distinct from the physical struggle.
3. [...] The collective nature and deployment of crowd-control equipment supersede the 'abuse' classification which applies only to isolated unilateral force without broader public order context.

...

Step 100: $\mathcal{E}_{100}$

1. A massive splash of mud, water, or debris combined with a vehicle's sudden deviation from the road surface and loss of structural stability serves as definitive evidence of a 'car_accident'. [...]
2. A 'choreographed performance' pattern is confirmed by synchronized physical maneuvers, uniformed groups executing coordinated collapses or stunts, and a celebratory atmosphere. [...]
3. [...] The collective context of public order operations supersedes the kinetic impact of a single vehicle unless the crash is the sole, isolated anomaly without broader crowd engagement.

...

Step 140: $\mathcal{E}_{140}$

1. A 'car_accident' pattern **requires** uncontrolled deviation resulting in structural failure or loss of stability. This pattern is **excluded if** the vehicle performs a jump, drift, or stunt where the trajectory is predictable and the vehicle remains structurally intact post-impact, even if dust or debris is generated.
2. A 'fighting' pattern is confirmed by uncontrolled mutual physical engagement [...]. This pattern supersedes 'choreographed performance' or 'normal' classifications if the actions involve unpredictable force, genuine retaliation, or chaotic struggle **rather than** synchronized stunts [...].
3. When an explosion occurs within a scene of active law enforcement suppression against a turbulent crowd, classify the event as 'riot'. [...]

...

Figure 9: Experience evolution for vehicle events, physical interactions, and crowd behavior at Steps 20, 60, 100, and 140. Blue highlights conditions or supporting evidence; orange highlights exceptions. Ellipses at the end of each column indicate additional entries; bracketed ellipses mark omitted text.

target explanatory text from each context's own generated response. For each case and method, each context's maps are normalized to sum to one over all frames and patches and displayed on a shared scale; Learned-minus-Base differences therefore indicate changes in relative attribution share.

The RGB sequence in Assault_11 (d) separates the earlier physical interaction near the left edge, where the person in blue lunges toward another person, from a dark-clothed person's later entrance into the corridor. Base describes the later movement and predicts positions 11–16. Learned Experience identifies the assault, while both attribution views allocate a larger share to the annotated event frames. The IFR difference row makes the temporal shift particularly clear: attribution increases during the earlier interaction and decreases in the later columns. The prediction now covers the previously missed portion of the assault, raising window IoU from 0.125 to 0.750. Assault_13 and Robbery_17 (e,f) also show increased anomalous-frame attribution share in both views, although their predicted intervals remain unchanged.

In the XD-Violence riot sequence (a), the input frames show a daytime crowd confrontation followed by an armored vehicle amid debris, before cutting to a dark street view at position 11. Learned Experience assigns more attribution share to the first ten frames in both views and confines the predicted event to positions 1–10, matching the annotation and improving IoU from 0.625 to 1.000. Here the advantage is retaining the visible event while avoiding extension into the subsequent shot.

Principles $\mathcal{E}_R^{\text{pri}}$

1. When a vehicle collision or crash generates smoke, dust, or debris, do not classify the event as 'explosion' unless a massive fireball or rapid thermal energy release is explicitly visible. Kinetic impact debris from a collision does not satisfy the mechanism of an explosion, even if the visual signature includes smoke or dust clouds.
2. When a video depicts a fictional narrative involving physical coercion or uncontrolled mutual combat, classify as 'fighting' or 'abuse' if the scene displays genuine physical force, structural damage, or chaotic struggle. Fictional physics and scripted dialogue do not override 'anomaly' classifications when the visual evidence confirms uncontrolled violence or physical harm inconsistent with a routine narrative event.
3. When an explosion occurs within a scene of active law enforcement suppression against a turbulent crowd, classify the event as 'riot'. The context of public disorder and police enforcement supersedes 'explosion' classifications, treating the blast as an aggravating factor of the riot rather than a standalone anomaly.

...

Patterns $\mathcal{E}_R^{\text{pat}}$

1. A 'car_accident' pattern **requires** uncontrolled deviation resulting in structural failure or loss of stability. This pattern is **excluded if** the vehicle performs a jump, drift, or stunt where the trajectory is predictable and the vehicle remains structurally intact post-impact, even if dust or debris is generated.
2. A 'fighting' pattern is confirmed by uncontrolled mutual physical engagement involving weapons or significant force, even within a cinematic narrative. This pattern supersedes 'choreographed performance' or 'normal' classifications if the actions involve unpredictable force, genuine retaliation, or chaotic struggle **rather than** synchronized stunts or routine procedural actions like arrests.
3. A 'riot' pattern is confirmed by uncontrolled public disorder involving uniformed personnel suppressing a crowd, even if the scene includes explosions or weapon discharge. This pattern supersedes 'explosion' or 'shooting' classifications when the primary visual signature is the chaotic suppression of a large group rather than isolated kinetic or projectile events.

...

Clues $\mathcal{E}_R^{\text{clu}}$

1. A massive splash of mud, water, or debris combined with a vehicle's sudden deviation from the road surface and loss of structural stability serves as definitive evidence of a 'car_accident'. This distinguishes real crashes from stunts where vehicles maintain predictable trajectories or remain structurally intact post-impact.
2. The presence of a weapon (e.g., bottle, bat) actively swung at a target, combined with physical grappling or choking, serves as definitive evidence of 'fighting' or 'abuse'. This visual signature overrides assumptions of 'normal' scripted events or 'choreographed performance' when the interaction involves uncontrolled physical force and potential harm.
3. The presence of uniformed officers actively suppressing a crowd alongside explosions or fire serves as definitive evidence of 'riot'. This visual signature overrides 'explosion' classifications when the blast is embedded within a broader context of public disorder and law enforcement engagement.

...

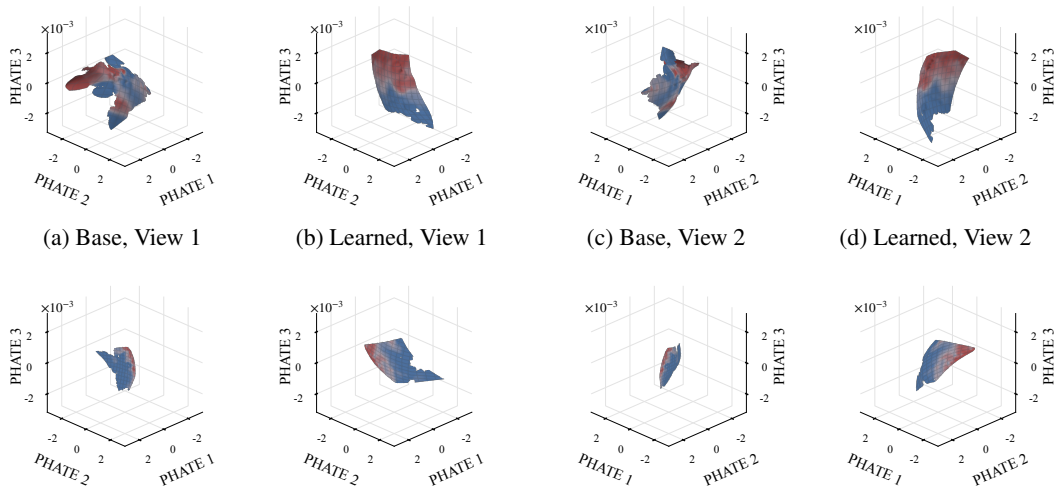
Figure 10: Selected Principles, Patterns, and Clues from the final Anom-Ψ experience bank $\mathcal{E}_R$. Blue highlights conditions or supporting evidence; orange highlights exceptions. Ellipses indicate additional entries.

The normal dialogue and flight sequences (b,c) show why appearance changes alone are insufficient evidence of an anomaly. In (b), the shot changes from a man speaking to a woman listening and briefly closing her eyes; Base flags the latter facial expressions as anomalous. In (c), a close view of aerial refueling changes to a distant view of aircraft flying in formation; Base treats the camera cut itself as an anomaly. With Learned Experience, both attribution views assign a smaller total share to the post-cut frames, and the responses interpret the sequences as an ordinary conversation and controlled flight, returning no anomalous interval. These examples connect reduced false alarms to event criteria that distinguish normal behavior and viewpoint changes from anomalous actions.

D.3 Prediction Curves and Response Comparisons

Effect of experience learning. A useful anomaly criterion should retain events while rejecting surrounding normal activity. Direct Optimization uses the same experience hierarchy as Anom-Ψ but learns without preference comparisons; this configuration is shared across the curve and response comparisons.

Figure 13 illustrates the resulting differences in temporal discrimination. In the all-normal clip (a), Anom-Ψ maintains low scores where Baseline and Direct Optimization produce elevated responses.



(a) Base, View 1 (b) Learned, View 1 (c) Base, View 2 (d) Learned, View 2

(e) Base, View 1 (f) Learned, View 1 (g) Base, View 2 (h) Learned, View 2

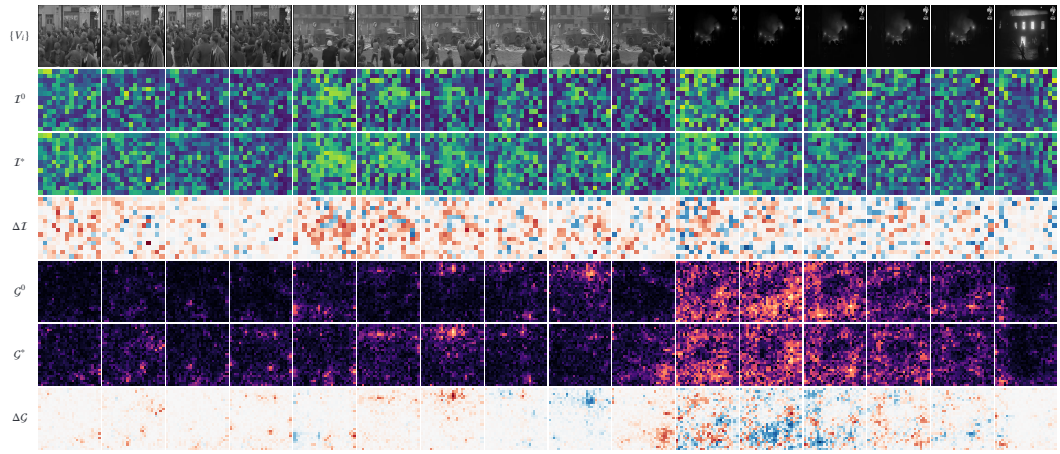
Figure 11: 3D PHATE visualizations under Base and Learned Experience on XD-Violence (top) and MSAD (bottom). Each row shows two views. Blue: normal; red: anomalous.

In the hockey clip (b), Direct Optimization reduces background responses but weakens part of the event response; Anom-Ψ preserves high scores during the event and suppresses them afterward. Its coverage of the three separated events in (c), with low scores between them, further illustrates that background suppression can coexist with sensitivity to recurring anomalies.

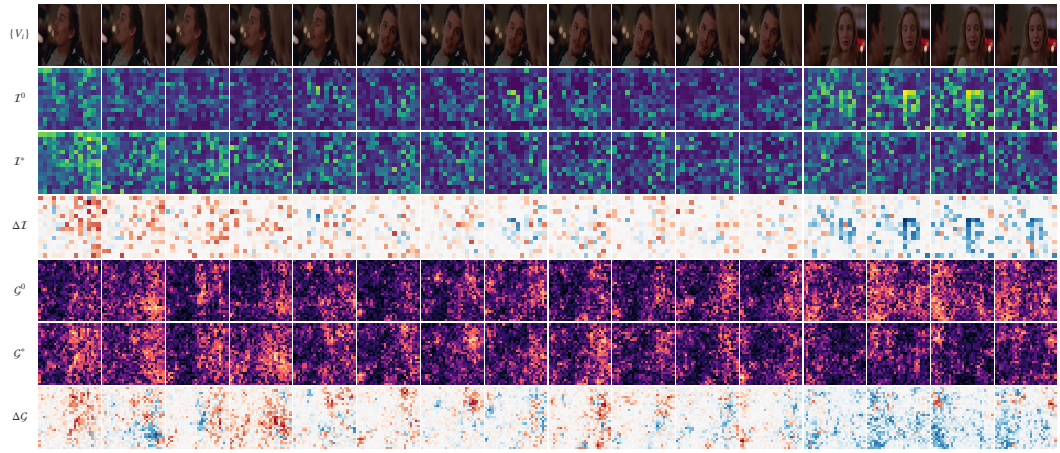
Comparison with external methods. Short events interspersed with normal content make the balance between event coverage and background suppression especially visible. In the field footage and *Fury* examples (Figure 14(c,d)), Anom-Ψ produces distinct responses near the annotated events while maintaining lower scores between them. On the field video, its AP reaches 71.4%, compared with the best external AP of 54.3% from DSRL. These curves complement the dataset-level comparison by showing how selective temporal responses contribute to the detection gain.

Response comparisons. The response comparisons examine how learned criteria distinguish similar event categories (Figures 15–20). In the car-accident case (Figure 18), Direct Optimization predicts an explosion, whereas Anom-Ψ identifies a car accident. This distinction is consistent with the bank's explicit separation of collision evidence from explosion mechanisms. In the riot cases (Figures 17 and 20), Anom-Ψ maintains the riot interpretation where Direct Optimization assigns explosion or abuse to parts of the scene, matching the learned emphasis on collective disorder and event context. With the same experience hierarchy in both methods, these cases illustrate the value of learning from conflicting judgments: the resulting conditions use event dynamics and context to resolve ambiguity in shared visual cues. The category distinctions complement the prediction curves, providing qualitative evidence that the learned criteria help identify both the event and its temporal extent.

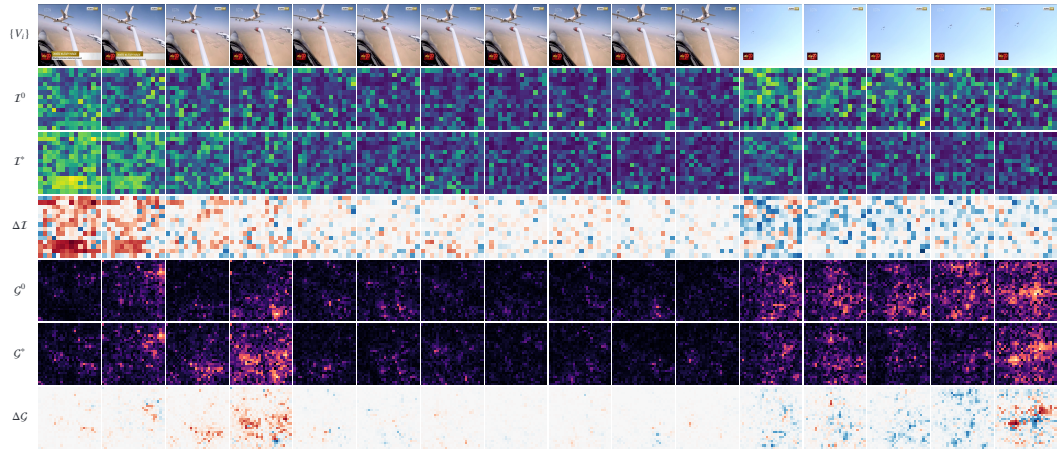
Green/red frame borders denote sampled normal/anomalous ground truth. Green/red comparison rows and like/dislike icons mark category agreement/disagreement with window ground truth. Temporal intersection over union (tIoU) measures localization over all sampled positions in the input window for matching categories. Figure 4 excerpts the *Fury* case shown in full in Figure 16. Ellipses mark omitted text and frames, but its ground-truth intervals, predicted intervals, and tIoU refer to all 16 input positions. Colored IDs and matching underlines link cited experience conditions to their application in reasoning. Red emphasis marks Direct Optimization's incorrect category decisions. Black emphasis highlights our category decisions and, in the explosion case (Figure 19), the distinction between the trigger and the anomaly. Emphasis is added for presentation without changing the quoted response or experience text.



(a) XD-Violence / ZCjNPgNu42g (riot)

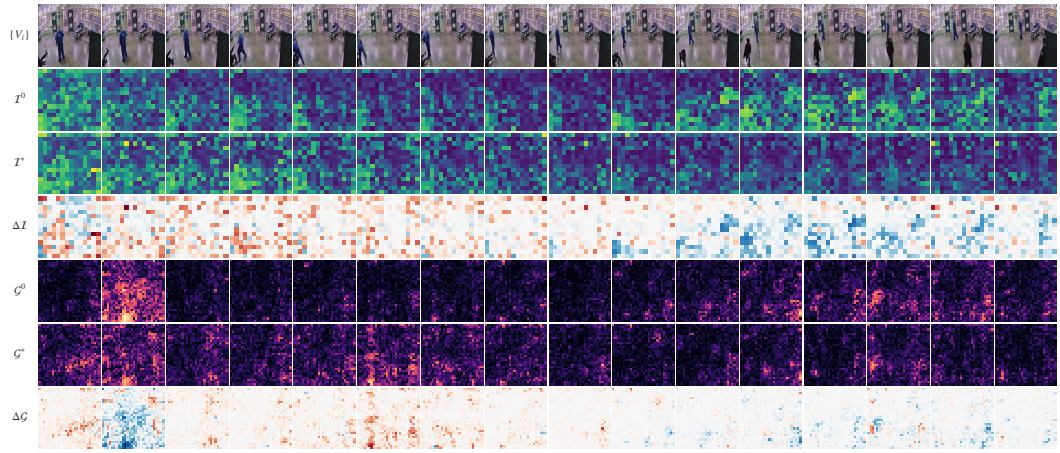


(b) XD-Violence / Before Sunrise (1995) (normal dialogue)

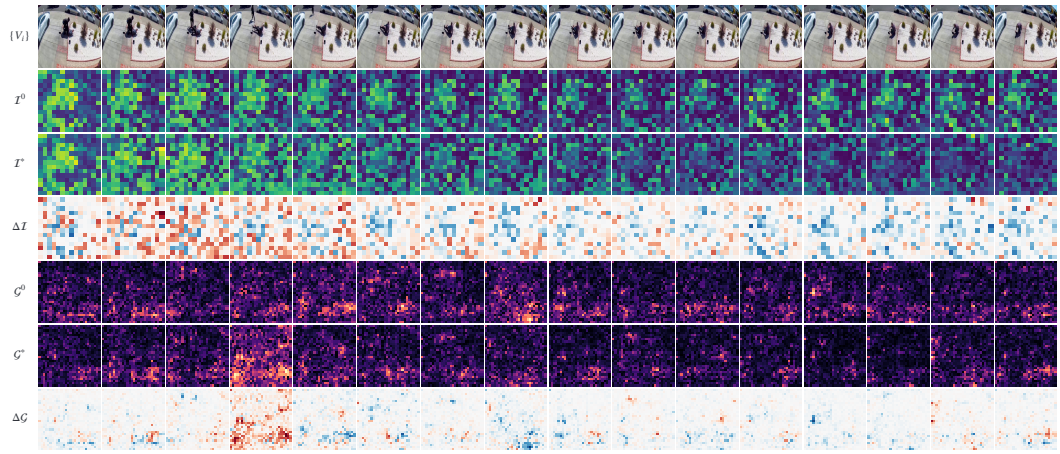


(c) XD-Violence / 0vVOXlgHC1U (normal flight)

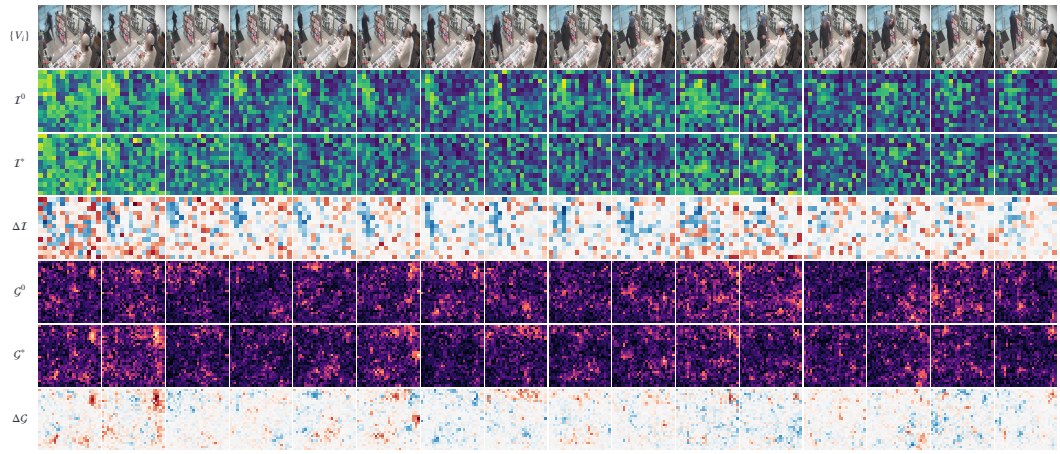
Figure 12: Selected visual-attribution examples from XD-Violence (a–c) and MSAD (d–f, continued). Each panel shows 16 frames, local IFR ($\mathcal{I}$) and absolute Gradient $\times$ Activation ($\mathcal{G}$) under Base (0) and Learned Experience ($*$), and Learned-minus-Base differences. Maps target each context's own explanation and are normalized over the full window, with shared scales per method and case. Red/blue denote increased/decreased attribution share.



(d) MSAD / Assault_11

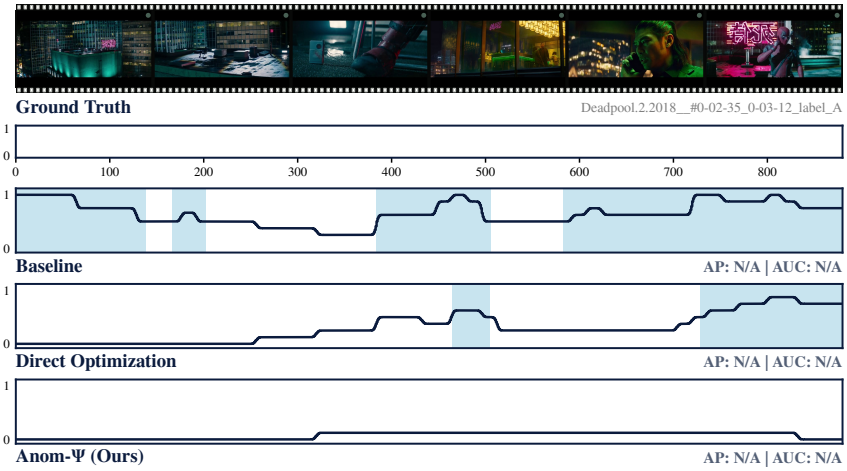


(e) MSAD / Assault_13

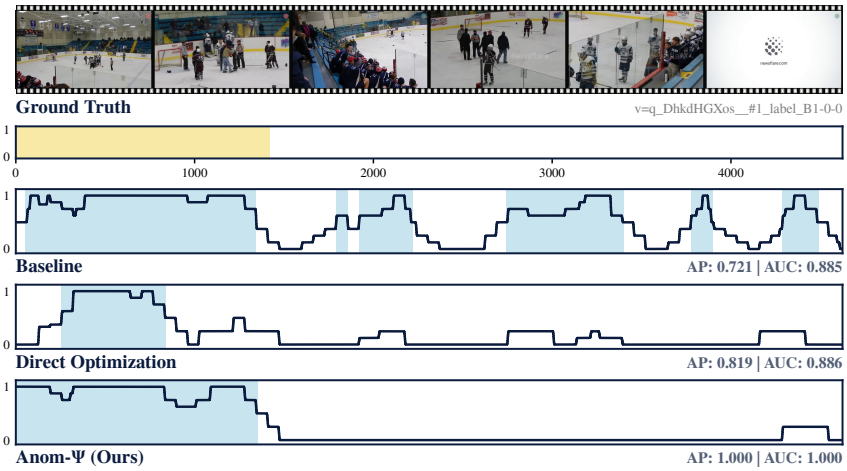


(f) MSAD / Robbery_17

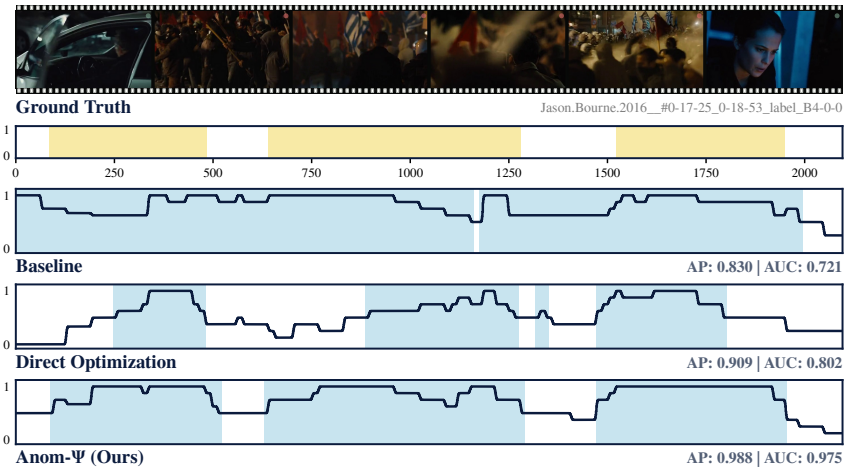
Figure 12: Visual attribution under Base and Learned Experience (continued): three MSAD examples. Rows and colors follow panels (a–c). In all three cases, both attribution methods assign a larger relative share to annotated anomalous frames under Learned Experience.



(a) Normal context

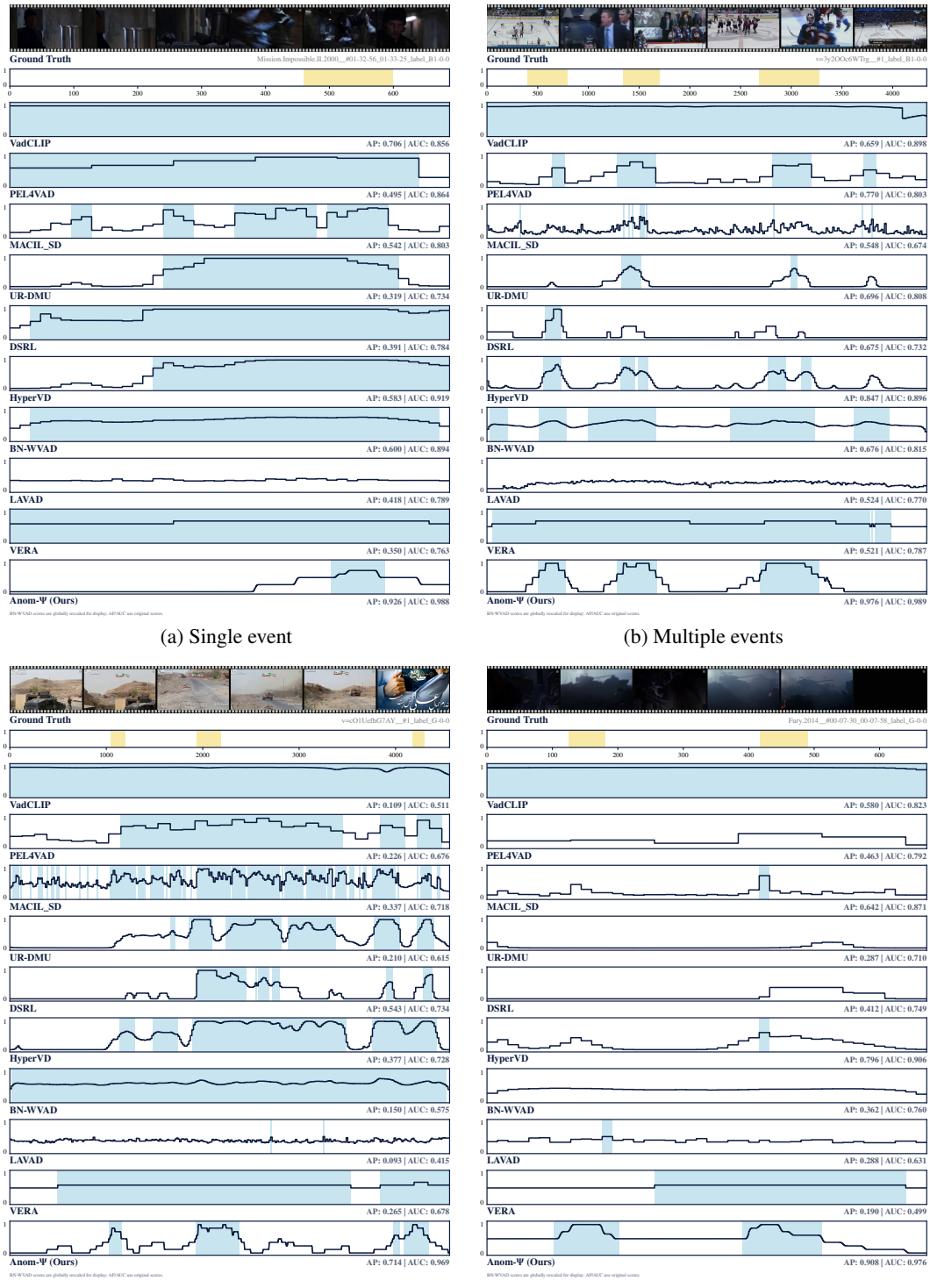


(b) Post-event discrimination



(c) Multi-event coverage

Figure 13: Effect of experience learning on temporal predictions. Baseline, Direct Optimization, and Anom-Ψ on three XD-Violence videos. AP and AUC are per-video scores on a 0-1 scale; N/A denotes the all-normal case.



(a) Single event (b) Multiple events

(c) Sparse events (d) Two short events

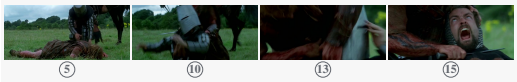
Figure 14: Temporal predictions of Anom-Ψ and nine external methods on XD-Violence. AP and AUC are per-video scores on a 0-1 scale. BN-WVAD curves are globally rescaled for display; its metrics use the original scores.

GT: fighting [6, 16] Video: Braveheart.1995__#02-07-00_02-08-15_label_B1-0-0 · 16 of 16 frames shown



Direct Optimization

Key evidence



MODEL OUTPUT

Observation

The sequence begins with a close-up of a knight's helmet, then cuts to a wide shot of a man lying injured on the grass while a knight on horseback stands over him. The knight then violently grabs the man, pulls him up, and stabs him in the neck with a dagger, causing the man to scream in agony. The sequence ends with a close-up of the man's bloody face.

Key positions: 5, 10, 13, 15

Reasoning

The sequence depicts a clear act of violence where one individual (the knight) inflicts severe harm upon another (the man on the ground). Frame 5 shows the man incapacitated on the ground. Frame 10 shows the knight physically grabbing and pulling the man up. Frame 13 shows the knight stabbing the man in the neck with a dagger. Frame 15 shows the man screaming in pain and blood on his face. **This constitutes 'abuse'** as it involves unilateral excessive force and victim incapacitation. The violence is not reciprocal (the man is **not fighting back),** and there is no evidence of a crowd or riot. The use of a weapon (dagger) to inflict harm is a form of abuse in this context.

Final prediction

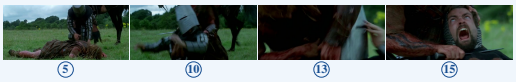
abuse [5, 15]

Comparison with ground truth tIoU

abuse [5, 15] —

Anom-Ψ (Ours)

Key evidence



MODEL OUTPUT

Observation

The sequence begins with a close-up of a knight's helmet, then cuts to a wide shot of a man lying on the ground in a field while a knight on horseback approaches. The knight dismounts and physically attacks the man on the ground, grappling and striking him. The man on the ground screams in agony as he is pinned and struck. The scene is clearly a physical confrontation involving weapons and significant force.

Key positions: 5, 10, 13, 15

Reasoning

The video depicts a clear **physical confrontation** between two individuals in a field. The principle [pr_010] states that fictional narratives involving physical coercion or uncontrolled mutual combat should be **classified as 'fighting'** or 'abuse' if the scene displays genuine physical force, structural damage, or chaotic struggle. The pattern [pt_010] confirms that 'fighting' is characterized by **uncontrolled mutual physical engagement** involving weapons or significant force, even within a cinematic narrative. The clue [cl_004] identifies the presence of a weapon (sword) actively swung at a target combined with physical grappling as definitive evidence of 'fighting'. The frames show the knight dismounting, attacking the man on the ground, and the man screaming in agony, which aligns with the definition of 'fighting'. The sequence is not a 'shooting' event as no projectile discharge is visible, and it is not a 'riot' as there is no crowd or law enforcement. Therefore, the event is classified as 'fighting'.

Final prediction

fighting [5, 16]

Comparison with ground truth tIoU

fighting [5, 16] 0.917

Direct Optimization · Experience bank

Cited experience

No experience entries cited.

Anom-Ψ (Ours) · Experience bank

Cited experience

[pr_010] When a video depicts a fictional narrative involving **physical coercion** or uncontrolled mutual combat, classify as 'fighting' or 'abuse' if the scene displays genuine physical force, structural damage, or chaotic struggle. Fictional physics and scripted dialogue do not override 'anomaly' classifications when the visual evidence confirms uncontrolled violence or physical harm inconsistent with a routine narrative event.

[pt_010] A 'fighting' pattern is confirmed by uncontrolled mutual physical engagement involving **weapons or significant force,** even within a cinematic narrative. This pattern supersedes 'choreographed performance' or 'normal' classifications if the actions involve unpredictable force, genuine retaliation, or chaotic struggle rather than synchronized stunts or routine procedural actions like arrests.

[cl_004] The presence of a weapon (e.g., bottle, bat) actively swung at a target, combined with physical grappling or choking, serves as definitive evidence of 'fighting' or 'abuse'. This visual signature overrides assumptions of 'normal' scripted events or 'choreographed performance' when the interaction involves uncontrolled physical force and potential harm.

Figure 15: Fighting: Braveheart. Direct Optimization predicts abuse, whereas Anom-Ψ predicts fighting. The baseline response cites no experience entries.

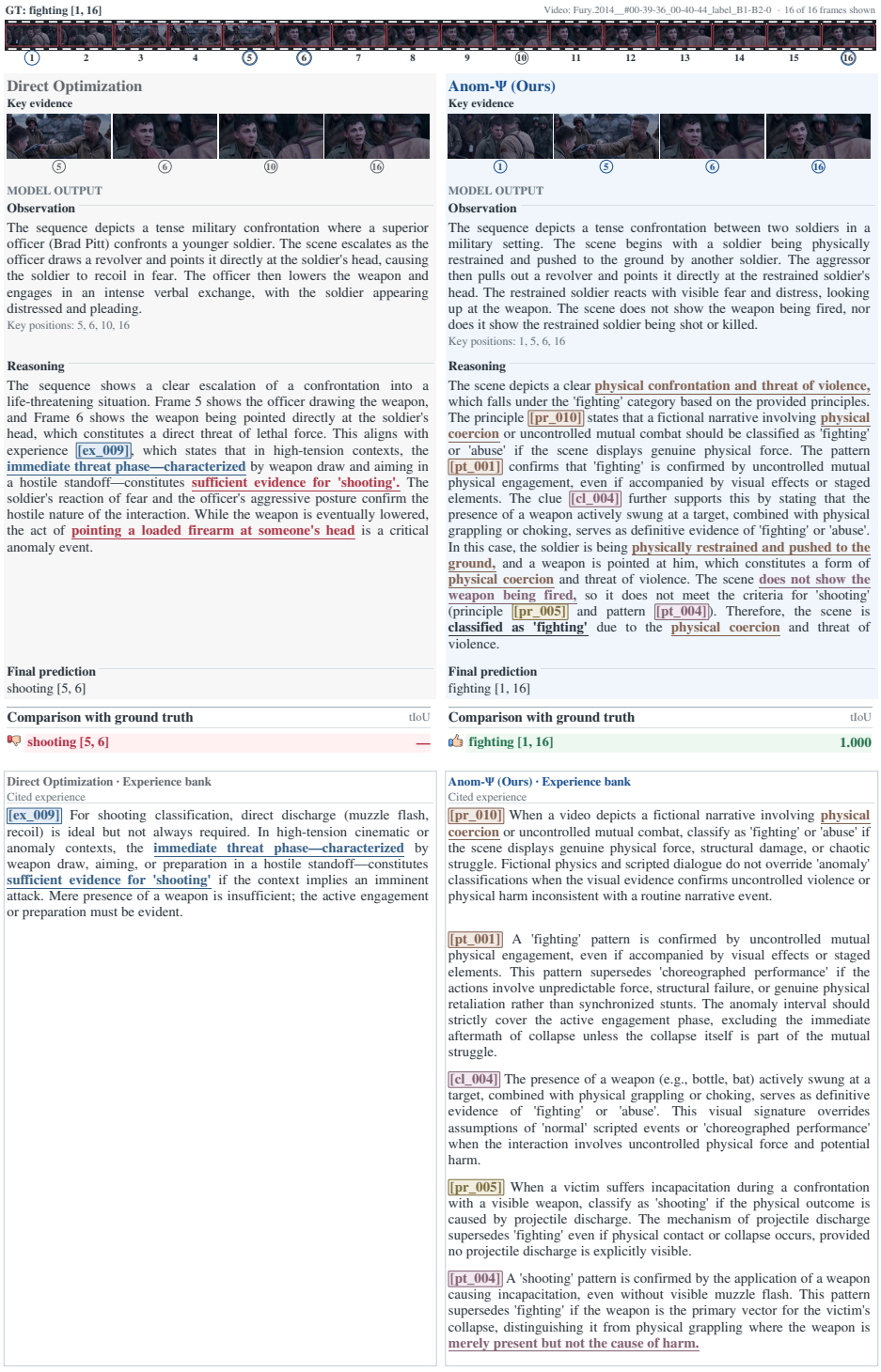
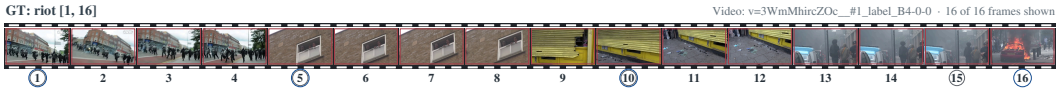
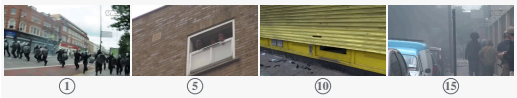


Figure 16: Fighting: Fury. Direct Optimization predicts shooting at positions 5-6, whereas Anom-Ψ predicts fighting over positions 1-16.



Direct Optimization

Key evidence



MODEL OUTPUT

Observation

The sequence begins with a large group of riot police in full gear moving down a city street, followed by civilians watching from a window. The scene then cuts to a damaged yellow storefront with debris on the sidewalk, and finally shows a red car engulfed in flames with heavy smoke billowing into the street, while people stand nearby observing.

Key positions: 1, 5, 10, 15

Reasoning

The sequence exhibits clear evidence of a riot and an explosion. Frames 1-4 show a large formation of riot police in full gear, indicating a loss of public order and a hostile environment consistent with a riot. Frames 5-8 show civilians observing from a window, suggesting the event is unfolding in a public space. Frames 9-12 show a damaged storefront with debris, indicating property destruction. Frames 13-16 show a red car completely engulfed in flames with thick smoke, which is a direct result of an explosion or fire. According to experience ex_002, the presence of a crowd in a hostile environment with loss of order constitutes a riot. According to experience ex_010 and ex_013, a vehicle engulfed in flames is **classified as an explosion,** and if a firearm discharge is visible, it would also be classified as shooting. In this case, the **car is clearly on fire,** indicating an explosion. The presence of riot police and the chaotic environment further supports the classification of a riot. Therefore, the sequence contains both a riot and an explosion.

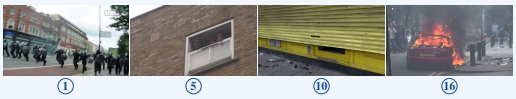
Final prediction

riot [1, 4]; explosion [13, 16]

Comparison with ground truth	tIoU
👍 riot [1, 4]	0.250
👎 explosion [13, 16]	—

Anom-Ψ (Ours)

Key evidence



MODEL OUTPUT

Observation

The video sequence begins with a large group of uniformed officers in riot gear advancing down a city street, followed by civilians looking out of windows. The scene then cuts to a damaged yellow storefront with debris on the sidewalk, and finally shows a red car engulfed in flames while people stand nearby. The progression indicates a public disorder event escalating into property damage and vehicle fire.

Key positions: 1, 5, 10, 16

Reasoning

The video depicts a clear escalation of public disorder. The initial frames show uniformed officers in riot gear actively suppressing a crowd, which establishes the context of a riot. The presence of debris and a damaged storefront further supports the chaotic nature of the event. The final frames show a car on fire, which, in the context of a riot, is **classified as part of the riot** rather than a standalone explosion. According to principle [pr_002], an explosion within a scene of **active law enforcement suppression** is classified as 'riot'. Pattern [pt_005] confirms that **uncontrolled public disorder involving uniformed personnel** supersedes 'explosion' classifications. Clue [cl_002] reinforces that **uniformed officers suppressing a crowd alongside fire** serves as definitive evidence of 'riot'. Therefore, the entire sequence is classified as a riot.

Final prediction

riot [1, 16]

Comparison with ground truth	tIoU
👍 riot [1, 16]	1.000

Direct Optimization · Experience bank

Cited experience

[ex_002] For riot classification, the crowd must exhibit active hostile interaction and loss of order (chaos, running, fighting). Mere standing, turning, or gathering in response to a threat does not constitute a riot. Arrests or detentions of individuals, even in tense settings, do not qualify as 'riot' unless they involve a broader breakdown of public order involving the crowd.

[ex_010] Classify vehicle destruction as 'explosion' **only if the primary mechanism is sudden energy release** (fireball, rapid expansion). Debris, dust, or smoke alone, even if intense, does not constitute 'explosion' if the mechanism is mechanical failure, collision, or uncontrolled deviation.

[ex_013] When a vehicle explosion occurs, strictly classify it as 'explosion' and never 'car_accident', even if the vehicle was moving or involved in a chase. Simultaneously, if a firearm discharge (muzzle flash, recoil) is visible, classify 'shooting' as a distinct concurrent category, regardless of the explosion's dominance. Do not omit 'shooting' if discharge evidence exists.

Anom-Ψ (Ours) · Experience bank

Cited experience

[pr_002] When an explosion occurs within a scene of active law enforcement suppression against a turbulent crowd, classify the event as 'riot'. The context of **public disorder and police enforcement** supersedes 'explosion' classifications, treating the blast as an aggravating factor of the riot rather than a standalone anomaly.

[pt_005] A 'riot' pattern is confirmed by **uncontrolled public disorder** involving uniformed personnel suppressing a crowd, even if the scene includes explosions or weapon discharge. This pattern supersedes 'explosion' or 'shooting' classifications when the primary visual signature is the chaotic suppression of a large group rather than isolated kinetic or projectile events.

[cl_002] The presence of uniformed officers **actively suppressing a crowd** alongside explosions or fire serves as definitive evidence of 'riot'. This visual signature overrides 'explosion' classifications when the blast is embedded within a broader context of public disorder and law enforcement engagement.

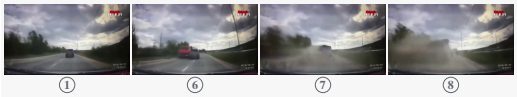
Figure 17: Riot: property damage and fire. Direct Optimization predicts riot followed by explosion, whereas Anom-Ψ predicts riot throughout the window.

GT: car_accident [6, 10] Video: v=X1arMZmYhsk__#00-13-00_00-14-09_label_B6-0-0 · 12 of 12 frames shown



Direct Optimization

Key evidence



MODEL OUTPUT

Observation

The sequence begins with a vehicle driving on a rural road, following another car. Suddenly, a red truck appears from the left and collides with the car ahead, causing it to explode violently. The explosion engulfs the camera's view in smoke and debris, obscuring the scene. The final frames show a dark, smoky interior, likely the aftermath of the explosion inside the recording vehicle.

Key positions: 1, 6, 7, 8

Reasoning

The sequence clearly depicts a violent vehicular collision followed by a **massive explosion.** Frames 1-5 show normal driving conditions. Frame 6 captures the critical moment where a red truck enters the frame and impacts the vehicle ahead. Frame 7 shows the immediate aftermath of the impact, with a large cloud of smoke and debris erupting from the struck vehicle, indicating a catastrophic failure or explosion. Frame 8 shows the camera view being completely obscured by the smoke and debris, consistent with a high-energy explosion. This aligns with experience [ex_004] for physical destruction events (car_accident moving to explosion) and [ex_010] which states that vehicle destruction via sudden energy release (explosion) should be classified as 'explosion' rather than 'car_accident'. The event is not a 'car_accident' in the sense of a simple crash, but a violent explosion caused by the collision. There is no evidence of 'shooting' (no muzzle flash or projectile trajectory), 'fighting' (no reciprocal physical contact), 'riot' (no crowd), or 'abuse' (no victim incapacitation or excessive force by an agent). Therefore, the **primary anomaly is an explosion.**

Final prediction

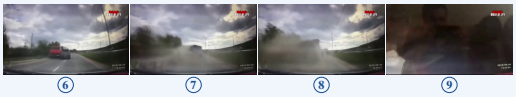
explosion [6, 8]

Comparison with ground truth tIoU

explosion [6, 8] —

Anom-Ψ (Ours)

Key evidence



MODEL OUTPUT

Observation

The video sequence begins with a vehicle driving on a rural road, following another car. Suddenly, a red truck appears from the left and collides with the car ahead, causing it to flip and generate a massive cloud of dust and debris. The camera vehicle is then engulfed in the dust cloud, obscuring visibility. The final frames show a completely different scene of heavy traffic on a winding mountain road, which appears to be a separate clip or a cutaway.

Key positions: 6, 7, 8, 9

Reasoning

The event progression shows a clear transition from normal driving to a violent collision. Principle [pr_003] establishes that smoke or dust generated by a vehicle collision does not constitute an 'explosion' unless a massive fireball or rapid thermal energy release is visible; here, the debris is clearly **kinetic impact dust.** Pattern [pt_003] defines a 'car_accident' as an **uncontrolled deviation resulting in structural failure** or loss of stability, which matches the **car flipping over.** Clue [cl_006] confirms this by stating that a massive splash of debris combined with a vehicle's sudden deviation and loss of stability is definitive evidence of a 'car_accident'. The visual evidence of the car flipping and the subsequent dust cloud overwhelming the camera vehicle supports the **classification of a crash.** The final frames showing traffic are likely a separate event or a cutaway and do not alter the classification of the primary anomaly.

Final prediction

car_accident [6, 9]

Comparison with ground truth tIoU

car_accident [6, 9] 0.800

Direct Optimization · Experience bank

Cited experience

[ex_004] For physical destruction events, shooting typically moves from a static weapon posture to visible discharge and consequences; car_accident moves from normal driving to uncontrolled deviation, collision, and vehicle damage; explosion moves from sudden energy release to rapid expansion, impact, debris, or structural rupture. Use the full causal chain rather than the aftermath alone to distinguish categories.

[ex_010] Classify vehicle destruction as 'explosion' only if the primary mechanism is sudden energy release (fireball, rapid expansion). Debris, dust, or smoke alone, even if intense, **does not constitute 'explosion'** if the mechanism is **mechanical failure, collision,** or uncontrolled deviation.

Anom-Ψ (Ours) · Experience bank

Cited experience

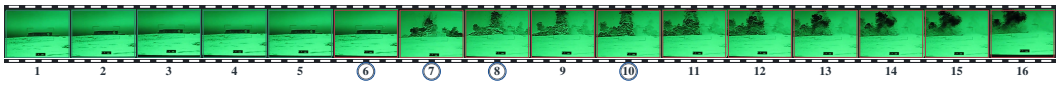
[pr_003] When a vehicle collision or crash generates smoke, dust, or debris, do not classify the event as 'explosion' unless a massive fireball or rapid thermal energy release is explicitly visible. **Kinetic impact debris** from a collision does not satisfy the mechanism of an explosion, even if the visual signature includes smoke or dust clouds.

[pt_003] A 'car_accident' pattern requires **uncontrolled deviation** resulting in structural failure or **loss of stability.** This pattern is excluded if the vehicle performs a jump, drift, or stunt where the trajectory is predictable and the vehicle remains structurally intact post-impact, even if dust or debris is generated.

[cl_006] A massive splash of mud, water, or debris combined with a vehicle's sudden deviation from the road surface and loss of structural stability serves as definitive evidence of a 'car_accident'. This distinguishes real crashes from stunts where vehicles maintain predictable trajectories or remain structurally intact post-impact.

Figure 18: Car accident. Direct Optimization predicts explosion at positions 6-8, whereas Anom-Ψ predicts car_accident at positions 6-9. The window contains 12 sampled frames.

GT: explosion [6, 16] — Video: v=abATVcjFumY__#1_label_G-0-0 · 16 of 16 frames shown



Direct Optimization

Key evidence



MODEL OUTPUT

Observation

The sequence begins with a static night-vision view of a dark, rectangular vehicle on a snowy surface. At frame 6, a projectile is seen flying towards the vehicle. By frame 7, the vehicle is engulfed in a massive explosion, followed by a large plume of smoke and debris that expands and dissipates through the remainder of the sequence.

Key positions: 6, 7, 8, 10

Reasoning

The sequence depicts a clear causal chain of violence. Frames 1-5 show a static scene. Frame 6 provides direct evidence of a projectile in flight, indicating a discharge event. Frame 7 shows the immediate consequence: a massive explosion and structural rupture of the vehicle. This aligns with experience [ex_004] (shooting moves to discharge/consequences) and [ex_006] (direct evidence for shooting includes projectile trajectory; direct evidence for explosion includes sudden expansion/structural rupture). The event is **classified as 'shooting'** due to the **visible projectile** and 'explosion' due to the blast and debris. It is not a 'car_accident' because the destruction is caused by an external blast, not mechanical failure or collision [ex_018].

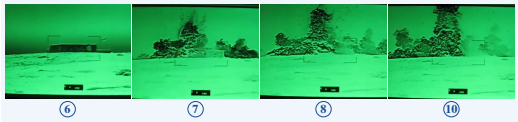
Final prediction

shooting [6, 6]; explosion [7, 16]

Comparison with ground truth	tIoU
shooting [6, 6]	—
explosion [7, 16]	0.909

Anom-Ψ (Ours)

Key evidence



MODEL OUTPUT

Observation

The video sequence begins with a static night-vision view of a rectangular object on a flat surface. At frame 6, a projectile appears and strikes the object. From frame 7 to 16, the object is obliterated by a massive explosion, generating a large fireball and debris cloud that expands and dissipates over time.

Key positions: 6, 7, 8, 10

Reasoning

The event progression clearly delineates a cause-and-effect sequence: a projectile impact followed by a massive energy release. Principle [pr_003] establishes that a vehicle collision or crash generating smoke or dust does not qualify as an 'explosion' unless a **massive fireball** or **rapid thermal energy release** is explicitly visible. In this case, the visual evidence from frames 7-16 confirms a massive fireball and rapid energy release, satisfying the boundary for an explosion. Pattern [pt_002] notes that an 'explosion' pattern is confirmed by massive energy release causing chaos, which is the dominant feature here. Clue [cl_005] identifies the presence of volatile infrastructure (the object) combined with sudden massive energy release as definitive evidence for 'explosion'. The event is not a 'car_accident' because the object is not a vehicle performing a stunt or crash; it is a stationary target being destroyed. The event is not 'shooting' because the **primary anomaly** is the **resulting explosion,** not the discharge itself (which is the trigger). Therefore, the classification is 'explosion' covering the interval from the impact to the dissipation of the smoke.

Final prediction

explosion [6, 16]

Comparison with ground truth	tIoU
explosion [6, 16]	1.000

Direct Optimization · Experience bank

Cited experience

[ex_004] For physical destruction events, shooting typically moves from a static weapon posture to visible discharge and consequences; car_accident moves from normal driving to uncontrolled deviation, collision, and vehicle damage; explosion moves from sudden energy release to rapid expansion, impact, debris, or structural rupture. Use the full causal chain rather than the aftermath alone to distinguish categories.

[ex_006] Direct evidence for shooting includes muzzle flash, recoil, **projectile trajectory,** or impact; direct evidence for car_accident includes abnormal trajectory, collision, vehicle deformation, or vehicle debris; direct evidence for explosion includes sudden expansion, a fireball, impact, debris, or structural rupture. A weapon, smoke, fire, or debris alone remains insufficient to determine the category.

[ex_018] A vehicle launch, flip, or violent deviation caused by loss of control, impact, or terrain is classified as 'car_accident', not 'explosion'. Classify as 'explosion' only if a distinct external blast source (fireball, shockwave, or rapid expansion from a non-vehicle origin) initiates the event. Mere vehicle destruction or airborne motion without an external blast source remains a 'car_accident'.

Anom-Ψ (Ours) · Experience bank

Cited experience

[pr_003] When a vehicle collision or crash generates smoke, dust, or debris, do not classify the event as 'explosion' unless a **massive fireball** or **rapid thermal energy release** is explicitly visible. Kinetic impact debris from a collision does not satisfy the mechanism of an explosion, even if the visual signature includes smoke or dust clouds.

[pt_002] An 'explosion' pattern is confirmed by **massive energy release** causing chaos, but this pattern is excluded if the scene is dominated by a large, uncontrolled crowd and police standoff. In such cases, the 'riot' pattern supersedes 'explosion' as the primary classification.

[cl_005] The presence of volatile infrastructure (pipes, tanks) combined with sudden massive energy release serves as definitive evidence for 'explosion'. This evidence overrides 'riot' or 'abuse' classifications unless uniformed crowd-control forces are visibly present and deploying suppression tools.

Figure 19: Explosion. Direct Optimization predicts shooting at position 6 followed by explosion at positions 7-16; Anom-Ψ predicts explosion at positions 6-16.



Figure 20: Riot: physical restraint. Direct Optimization predicts abuse at positions 12-16, whereas Anom-Ψ predicts riot over positions 1-16.