INTEREVOLVE: SELF-EVOLVING AGENTIC SEARCH VIA RETRIEVAL–UNDERSTANDING INTERACTION

Anonymous authors
Paper under double-blind review

ABSTRACT

State assessment in self-evolving agentic search poses significant challenges, as inaccurate evaluations can mislead the evolutionary trajectory and trap the system in redundant or mis-evolved loops. In this paper, we observe that humans mitigate such failures via retrieval-understanding interaction. During multi-turn search, they use prior understanding to form expectations about retrieval results and assess the consistency between results and expectations to resist misdirection from search outputs. Inspired by this behavior, we propose InterEvolve, an agentic search framework that improves the accuracy of state assessment without requiring additional model training or task-specific fine-tuning. Specifically, InterEvolve comprises two main components: the Retrieval–Understanding Progress Assessor (RUPA) and the Query Refinement Validator (QRV). RUPA leverages retrieval–understanding consistency to assess whether each search iteration progresses in a productive direction. Moreover, QRV constrains retrieval discrepancies between the original and refined queries, ensuring reliable query refinement and preventing misdirected evolution. Extensive experiments on the MMEB-v2 benchmark (78 datasets) across 4 popular retrieval backbones demonstrate that InterEvolve consistently outperforms the naive agentic search baseline in all model combinations. Notably, agentic search relying on weaker models like Qwen3.8-27B for state assessment suffers from misdirected evolution, degrading from 74.96 to 74.14 below single-turn retrieval. InterEvolve recovers this degradation and further improves it to 75.48, demonstrating its robustness safeguard. The code will be released if the paper is accepted.

1 INTRODUCTION

Agentic search harnesses the reasoning capabilities of large language models (LLMs) Bai et al. (2025); Qwen Team (2026); Kimi Team (2026) to drive iterative retrieval. Across multiple turns, the system progressively refines its understanding of a complex query. In this way, the system is able to address information needs that cannot be satisfied in a single retrieval pass, yielding substantial gains over single-shot baselines. Consequently, agentic search has emerged as a promising paradigm for real-world tasks requiring multi-step information seeking.

Previous studies on improving agentic systems for information search have primarily focused on two perspectives. Gou et al. (2026) Several works focus on enhancing the retrieval backbone, Lin et al. (2026); Zeng et al. (2026) such as adopting stronger embedding models or introducing relevance-aware re-ranking, to improve the quality of retrieved results at each step. Other methods concentrate on refining the query formulation process, Li et al. (2026c); Shi et al. (2026) employing techniques such as query rewriting, expansion, or decomposition to better align user intent with retrievable content. However, these approaches largely assume that each iteration moves the system in a productive direction, without explicitly assessing whether the current retrieval state reflects genuine progress. Inaccurate state assessment can mislead the evolutionary trajectory, causing the system to fall into redundant retrieval loops or diverge toward irrelevant topics.

In this paper, we observe that humans mitigate such failures through retrieval–understanding interaction. For instance, given an image of an uncommon fruit, a person may not know its name before searching. Nevertheless, they already possess rich prior knowledge about fruits — such as typical shapes, colors, and textures — which shapes their expectations about what relevant retrieval

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

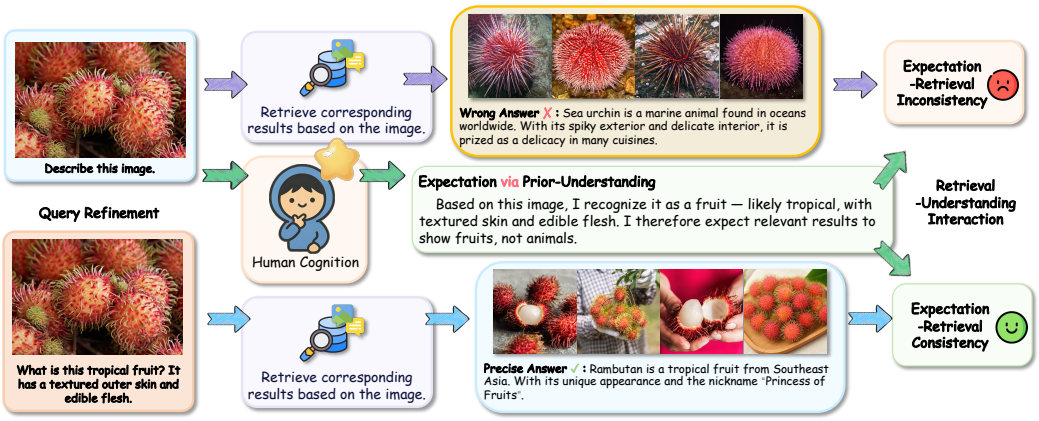


Figure 1: **Retrieval-understanding interaction in human multi-turn search.** A person uses prior understanding to form expectations about retrieval results, and assesses consistency between results and expectations to detect misdirection.

results should look like. When the returned results deviate from these expectations, the person can promptly filter out irrelevant items and reflect on whether the search tool itself has misinterpreted the query. In this way, humans use prior understanding to form expectations about retrieval results. They then assess the consistency between results and expectations to resist misdirection from search outputs. Such retrieval–understanding interaction enables early misdirection detection and real-time trajectory correction, ensuring efficient and reliable multi-turn retrieval.

Inspired by this observation, we propose InterEvolve, a training-free agentic search framework that improves the accuracy of state assessment by explicitly modeling retrieval–understanding consistency. Specifically, InterEvolve comprises two main components: the Retrieval–Understanding Progress Assessor and the Query Refinement Validator. **Higher retrieval–understanding consistency indicates more reliable retrieval.** At each iteration, the Retrieval–Understanding Progress Assessor (RUPA) measures this consistency to evaluate whether the iteration moves the system toward a correct state or into a misdirected one. In this way, RUPA mirrors the human ability to detect misdirection through expectation–evidence comparison. Moreover, **queries sharing semantic information tend to yield similar retrieval results**. The Query Refinement Validator (QRV) analyzes the discrepancy between the retrieval sets of the original and refined queries to verify whether query rewriting is correct. Therefore, QRV constrains retrieval discrepancies to ensure reliable query refinement and prevent misdirected evolution. We conduct comprehensive experiments on MMEB-v2, a large-scale multimodal embedding benchmark comprising 78 datasets spanning classification, visual question answering, retrieval, and grounding tasks. Evaluated across 4 retrieval backbones and 3 understanding models, InterEvolve consistently outperforms all baselines, achieving a new state-of-the-art of 82.10 Overall on WeMM-VL-9B Zhou et al. (2026). Our contributions can be summarized as follows:

- We identify the critical role of state assessment in self-evolving agentic search and reveal that inaccurate evaluations can mislead the evolutionary trajectory.
- We propose InterEvolve, a training-free agentic search framework that improves state assessment accuracy by explicitly modeling retrieval–understanding interaction.
- Extensive experiments demonstrate that InterEvolve effectively mitigates misdirected evolution caused by erroneous state assessments.

2 Related Work

2.1 Agentic Search

Recent work has replaced static retrieval pipelines with autonomous agents that interleave reasoning, query formulation, and multi-turn search. By casting retrieval as a sequential decision process,

agents learn to decompose complex information needs, issue targeted sub-queries, and aggregate evidence over multiple rounds (Chen et al., 2025; Li et al., 2025; Jin et al., 2025). Query rewrite has emerged as a central capability in this paradigm. Reasoning-conditioned queries have been generated and progressively refined across turns to accumulate more relevant evidence (Li et al., 2025), while multi-hop questions have been decomposed into atomic sub-queries with dynamic retrieval timing (Guan et al., 2026). Concise yet diverse query rewrite have further been encouraged through reward shaping during training (Li et al., 2026a). Self-evolution mechanisms have further reduced reliance on curated training data through proposer-solver co-evolution (Yue et al., 2026) and memory-guided dense reward signals (Li et al., 2026a). However, these methods require non-trivial parameter updates and target text-only question answering rather than multimodal retrieval. In the multimodal domain, visual search tools have been learned through reinforcement learning for question answering (Wu et al., 2025; Jia et al., 2026a), and training-free hierarchical memory has been shown to support multi-round image search in personal photo collections Lyu et al. (2026). Despite these advances, iterative query reformulation has not been systematically explored as a mechanism for directly improving multimodal retrieval performance. In this work, we propose a training-free agent search framework where an LLM iteratively rewrites the retrieval query across multiple rounds.

2.2 Multimodal Retrieval

Multimodal retrieval aims to match queries against a heterogeneous corpus of images, text, and interleaved documents. The quality of retrieval is fundamentally bounded by the underlying representation model. Early approaches relied on modality-specific dual encoders to achieve cross-modal alignment through contrastive learning on large-scale image–text pairs (Radford et al., 2021; Jia et al., 2021; Zhai et al., 2023; Jia et al., 2026b). While these models provide efficient and scalable retrieval, they struggle with jointly representing interleaved inputs and capturing fine-grained cross-modal interactions. To bridge this gap, subsequent works integrated pretrained visual encoders as tokenizers into language models (Chen et al., 2024; Wang et al., 2025; Bai et al., 2025), broadening the input scope but still constrained by the encoder's representational capacity. The advent of multimodal large language models (MLLMs) has catalyzed a new generation of universal embedding models that natively handle arbitrary interleaved combinations of text and images. By adapting MLLM hidden states into retrieval representations through continued contrastive training (Jiang et al., 2024a; Zhang et al., 2024), and further scaling with synthetic data pipelines (Zhou et al., 2025; Jiang et al., 2024b) and refined training recipes (Li et al., 2026b; Zhou et al., 2026), these models have established MLLM-based embedding as the dominant paradigm for multimodal retrieval. However, even the most advanced embedding models operate in a single-turn, query-to-document matching fashion, leaving the potential of multi-turn agentic search strategies largely untapped. In this work, we demonstrate that our proposed agentic search framework consistently improves retrieval performance across a comprehensive set of such state-of-the-art multimodal embedding backbones.

3 Method

We introduce InterEvolve, a training-free framework for reliable test-time query evolution in multimodal retrieval. The framework uses only two frozen models: a multimodal retrieval model $\mathcal{R}$ and an LVLM $\mathcal{L}$. The same retrieval model performs Top-K retrieval and subsequent answer–candidate consistency scoring, while the LVLM handles candidate assessment, query refinement, and direct answer generation. As illustrated in Fig. 2, InterEvolve separates query proposal from transition acceptance through two complementary checks: the Retrieval–Understanding Progress Assessor (RUPA) evaluates consistency gains, and the Query Refinement Validator (QRV) constrains query-induced distributional shifts over the currently accepted Top-10 candidates. Both checks operate on the outputs of $\mathcal{R}$ and $\mathcal{L}$ without introducing additional models.

3.1 InterEvolve Overview

Given an initial query $Q_0 = (I_q, u, q_0)$ and a candidate corpus $\mathcal{C} = \{c_i\}_{i=1}^{N}$, the objective is to retrieve a candidate that satisfies the original information need. Here, I_q denotes the query image,

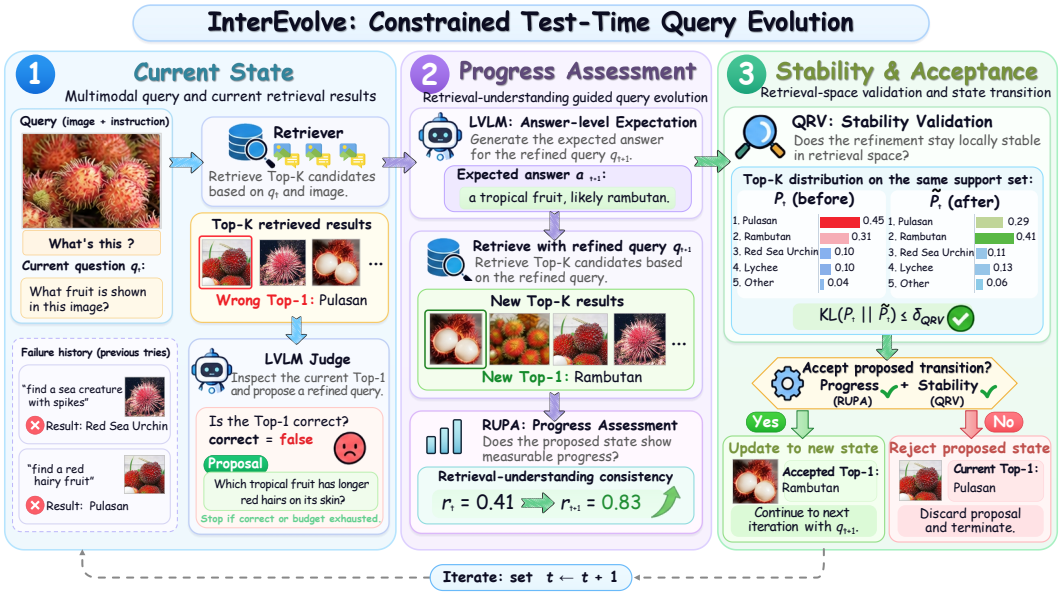


Figure 2: Overview of InterEvolve. One LVLM proposes refinements and answer-level expectations, while the same multimodal retriever supplies retrieval results and consistency scores. RUPA checks consistency gains; QRV compares distributions induced by both queries over the currently accepted Top-10 candidates. Only proposals passing both checks are accepted.

u the task instruction, q_0 the original question, and c_i the i-th candidate among N corpus entries. InterEvolve keeps I_q and u fixed and iteratively refines only the textual question.

At iteration t, the currently accepted question q_t defines $Q_t = (I_q, u, q_t)$. The retrieval model $\mathcal{R}$ scores the corpus candidates and returns a ranked list:

$$s_i^t = \mathrm{sim}_{\mathcal{R}}(Q_t, c_i), \qquad \mathcal{C}_t^K = \mathrm{TopK}_{c_i \in \mathcal{C}}\, s_i^t, \tag{1}$$

where $\mathrm{sim}_{\mathcal{R}}$ is the native similarity function of $\mathcal{R}$, and s_i^t is the score assigned to corpus candidate c_i under Q_t. The operator TopK returns the K highest-scoring candidates in descending order. We use $K = 10$, with $N \geq K$. The candidate at rank $j \in \{1, \ldots, K\}$ is denoted by c_t^j, with c_t^1 representing the Top-1 result. Retrieval also retains a candidate-ID-to-score map S_t, where $S_t[i] = s_i^t$ for each retained candidate. This map allows scores from different queries to be aligned by candidate identity rather than by retrieval rank.

The current question and its retrieval results constitute the retrieval state. The LVLM assesses this state and proposes a refinement when necessary; the proposed query is then executed by the same retrieval model before RUPA and QRV determine whether to accept the transition. Quantities indexed by $t + 1$ remain provisional until both checks pass. The initial retrieval is indexed by $t = 0$, and the index increases only after an accepted refinement, up to a budget of $T_{\max}$ refinements. Algorithm 1 summarizes this procedure, with $\mathrm{Retrieve}_{\mathcal{R}}(Q_t, \mathcal{C}, K)$ returning the ranked candidates $\mathcal{C}_t^K$ and score map S_t.

3.2 Retrieval–Understanding Guided Query Evolution

The LVLM assesses whether the current Top-1 satisfies the original request and, when necessary, proposes a refined question. We express these outputs as

$$(b_t, q_{t+1}) = \mathcal{L}^{\mathrm{judge}}\left(Q_0, q_t, c_t^1, \mathcal{C}_t^K, c_{t-1}^1, \mathcal{H}_t\right), \tag{2}$$

where $\mathcal{L}^{\mathrm{judge}}$ denotes $\mathcal{L}$ prompted for candidate assessment and query refinement. The binary output $b_t \in \{0, 1\}$ indicates whether the LVLM considers c_t^1 satisfactory, and $\mathcal{H}_t$ records previous questions and their judged-incorrect candidates. The previous accepted Top-1, c_{t-1}^1, provides additional feedback about the retrieval trajectory. At $t = 0$, the history is empty and the previous candidate is

Table 1: Main results on MMEB-v2 with Qwen3.8-Max.

Model	Size	AVG	Image					Video					VisDoc
			Overall	Cls	QA	Ret	GD	Overall	Cls	QA	V-Ret	M-Ret	Overall
# datasets →		78	36	10	10	12	4	18	5	5	5	3	24
VLM2Vec Jiang et al. (2024b)	2B	47.8	59.7	58.7	49.3	65.0	72.9	29.0	33.4	30.5	20.6	33.0	44.0
GME Zhang et al. (2024)	2B	55.4	51.9	54.4	29.9	66.9	55.5	33.9	34.9	42.0	25.6	32.4	76.8
VLM2Vec-V2 Meng et al. (2025)	2B	59.3	64.9	62.9	56.3	69.5	77.3	34.9	39.3	34.3	28.8	38.5	69.2
Ops-MM-embedding-v1	2B	64.6	69.0	68.1	65.1	69.2	80.9	47.6	53.6	55.7	41.8	33.7	70.8
RzenEmbed Jian et al. (2025)	2B	67.2	72.3	68.5	66.3	74.5	90.3	47.3	50.4	49.7	46.6	38.9	74.5
Qwen3-VL-Embedding Li et al. (2026b)	2B	73.4	75.0	70.2	74.4	74.9	88.5	62.4	73.0	66.3	53.9	52.3	79.2
+Ours	2B	**75.4**	**77.5**	**73.3**	**77.5**	**77.0**	**89.6**	**65.5**	**75.3**	**72.5**	**54.0**	**56.9**	**79.8**
WeMM-Embedding Zhou et al. (2026)	2B	77.8	79.6	74.4	79.3	79.4	94.2	70.2	84.8	71.7	**62.1**	56.9	80.8
+Ours	2B	**78.9**	**80.8**	**75.4**	**80.7**	**80.5**	**95.0**	**71.9**	**85.0**	**74.7**	**62.1**	**62.0**	**81.3**
VLM2Vec Jiang et al. (2024b)	8B	53.2	65.5	62.7	56.9	69.4	82.2	34.0	39.1	30.0	29.0	40.6	49.1
GME Zhang et al. (2024)	8B	59.2	56.0	57.7	34.7	71.2	59.3	38.6	37.4	50.4	28.4	38.2	79.3
Ops-MM-embedding-v1	8B	68.9	72.7	69.7	69.6	73.1	87.2	53.8	59.7	62.2	45.7	43.2	74.4
RzenEmbed Jian et al. (2025)	8B	72.9	75.9	70.6	71.7	78.5	92.1	55.7	58.8	63.5	51.0	45.5	81.3
IFM-TTE Cui et al. (2026)	8B	74.1	77.9	76.7	78.5	74.6	89.3	59.2	60.5	67.9	51.7	54.9	79.5
Qwen3-VL-Embedding Li et al. (2026b)	8B	77.8	80.1	74.3	81.2	80.1	92.4	66.7	78.7	70.4	58.6	54.0	82.5
+Ours	8B	**78.7**	**80.9**	**75.3**	**82.0**	**80.6**	**93.2**	**68.8**	**79.6**	**74.5**	**58.7**	**58.0**	**82.8**
WeMM-Embedding Zhou et al. (2026)	9B	80.3	81.6	75.7	82.2	81.5	95.6	73.7	**87.4**	**77.7**	**66.3**	56.5	83.4
+Ours	9B	**80.7**	**82.1**	**76.3**	**82.5**	**81.9**	**96.1**	**74.0**	86.0	77.6	**66.3**	**60.6**	**83.5**

absent. The question output is unused when $b_t = 1$; a proposal is pursued only when $b_t = 0$ and $t < T_{\max}$.

The original query anchors the retrieval objective, while the current results and failure history inform the refinement. The LVLM uses this feedback to identify ambiguities or missing constraints rather than selecting an answer exclusively from the current Top-10. When a proposal is pursued, the current failure is recorded as $\mathcal{H}_{t+1} = \mathcal{H}_t \cup \{(q_t, c_t^1)\}$ for subsequent judge calls. The proposed question is then passed to $\mathcal{R}$ to obtain its ranked candidates $\mathcal{C}_{t+1}^K$ and score map S_{t+1}.

RUPA assesses the resulting transition by comparing each retrieved Top-1 with an answer-level expectation. For $v \in \{t, t+1\}$, indexing the current or proposed state, the same LVLM directly answers the corresponding query:

$$a_v = \mathcal{L}^{\text{answer}}(I_q, u, q_v), \qquad r_v = \cos\big(E_{\mathcal{R}}(a_v), E_{\mathcal{R}}(c_v^1)\big)\,, \tag{3}$$

where $\mathcal{L}^{\text{answer}}$ denotes a separate prompting role of the same $\mathcal{L}$, receiving the image, instruction, and question without retrieved candidates in this answering call. Its output a_v serves as an answer-level expectation rather than a ground-truth label. The notation $E_{\mathcal{R}}(\cdot)$ denotes the encoding interface of the retrieval model $\mathcal{R}$ for textual or multimodal inputs, using the input formatting appropriate to each role. In particular, $E_{\mathcal{R}}(c_v^1)$ is the candidate representation used for retrieval, and $E_{\mathcal{R}}(a_v)$ encodes the generated answer with that same model. The operator cos computes the normalized inner product of these representations, yielding the retrieval–understanding consistency score r_v.

Thus, the two LVLM roles share the same model, and retrieval and consistency scoring share the same multimodal backbone. RUPA introduces neither an auxiliary answer encoder nor a separate learned evaluator. It uses the consistency change as a progress surrogate:

$$\Delta r_t = r_{t+1} - r_t, \qquad \text{RUPA}_t = \mathbb{I}[\Delta r_t \geq \delta_{\text{RUPA}}]\,, \tag{4}$$

where Δr_t is the consistency gain induced by the proposal, and δ_{RUPA} is the minimum required gain. The indicator $\mathbb{I}[\cdot]$ equals 1 when its condition holds and 0 otherwise, so $\text{RUPA}_t = 1$ indicates a passed progress check. This criterion measures relative agreement with the LVLM's understanding, not correctness with respect to the original query. Because the question and its expected answer can both change, we complement RUPA with a constraint on how strongly the refinement changes the retriever's local candidate preferences.

3.3 Retrieval-Space Validation and State Transition

QRV evaluates query refinement over a shared candidate support rather than comparing two independently ranked Top-10 lists. Let $\mathcal{I}_t = \{i_1, \ldots, i_K\}$ contain the corpus IDs of the currently

accepted Top-10 candidates, with $c_{i_j} = c_t^j$. For the transition from Q_t to Q_{t+1}, we fix this support and compare the score of each candidate under both queries. Specifically, $s_{i_j}^t$ is its current score, whereas $s_{i_j}^{t+1} = \mathrm{sim}_{\mathcal{R}}(Q_{t+1}, c_{i_j})$ is its score under the proposed query. These values are read from S_t and S_{t+1}, which contain all entries required for the comparison. A candidate need not remain in the proposed Top-10 to be included. Both score vectors follow the candidate order in $\mathcal{C}_t^K$; the proposed scores are not independently re-sorted.

We convert the aligned score vectors into local distributions using a temperature-scaled softmax:

$$p_j^t = \frac{\exp(s_{i_j}^t/\tau)}{\sum_{\ell=1}^K \exp(s_{i_\ell}^t/\tau)}, \qquad \widetilde{p}_j^t = \frac{\exp(s_{i_j}^{t+1}/\tau)}{\sum_{\ell=1}^K \exp(s_{i_\ell}^{t+1}/\tau)}. \tag{5}$$

Here, $\tau > 0$ is the shared temperature controlling distribution concentration, and ℓ indexes the candidates in the normalization sum. The quantities p_j^t and $\widetilde{p}_j^t$ represent the relative preference for the same candidate c_{i_j} under the current and proposed queries, respectively. We denote the resulting vectors by $P_t = (p_j^t)_{j=1}^K$ and $\widetilde{P}_t = (\widetilde{p}_j^t)_{j=1}^K$. Both distributions are normalized over $\mathcal{I}_t$, with the tilde indicating the proposed query evaluated on the current support. After an accepted transition, the next comparison uses the updated support $\mathcal{I}_{t+1}$ from $\mathcal{C}_{t+1}^K$. Accordingly, $\widetilde{P}_t$ is not identified with the next iteration's current distribution P_{t+1}.

QRV measures the discrepancy between these distributions using the Kullback–Leibler divergence:

$$d_t = D_{\mathrm{KL}}(P_t \parallel \widetilde{P}_t) = \sum_{j=1}^K p_j^t \log \frac{p_j^t}{\widetilde{p}_j^t}, \qquad \mathrm{QRV}_t = \mathbb{I}[d_t \le \delta_{\mathrm{QRV}}], \tag{6}$$

where d_t is the local distributional discrepancy, log is the natural logarithm, and $\delta_{\mathrm{QRV}} \ge 0$ is the maximum permitted discrepancy. A proposal passes QRV when it changes the relative preferences over the current Top-10 candidates by no more than this bound. Using the current distribution as the reference makes the check sensitive to reductions in the relative probability of currently prominent candidates. QRV therefore provides a local, trust-region-like constraint on retrieval behavior, discouraging overly aggressive changes within the current neighborhood. The constraint is local to $\mathcal{I}_t$ and does not by itself establish semantic fidelity to the original query or retrieval correctness. It uses aligned scores from the same retrieval model without introducing an additional learned evaluator.

InterEvolve accepts the proposed transition only when both checks pass:

$$\mathrm{Accept}_t = \mathrm{RUPA}_t \wedge \mathrm{QRV}_t, \tag{7}$$

where $\mathrm{Accept}_t \in \{0, 1\}$ indicates acceptance and $\wedge$ denotes logical conjunction. An accepted proposal commits the new question, retrieval results, and updated history, after which the same LVLM assesses the newly accepted Top-1. If either check fails, InterEvolve discards the proposal, terminates the search, and returns c_t^1 from the last accepted state. The process also returns the current Top-1 when $b_t = 1$ or $t \ge T_{\max}$. The resulting interaction reuses $\mathcal{L}$ to propose and interpret queries and $\mathcal{R}$ to retrieve and assess consistency, while the joint validation rule requires each accepted transition to achieve sufficient consistency gain with controlled changes in local retrieval preferences.

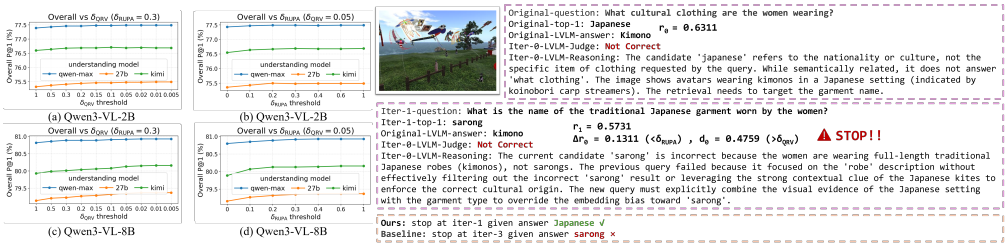
Algorithm 1: InterEvolve test-time search

Require: $Q_0 = (I_q, u, q_0)$, corpus $\mathcal{C}$, models $\mathcal{R}, \mathcal{L}$
Require: Budget $T_{\max}$, $K = 10$, τ, thresholds $\delta_{\mathrm{RUPA}}, \delta_{\mathrm{QRV}}$
Ensure: Final Top-1 candidate

$t \leftarrow 0$, $\mathcal{H}_0 \leftarrow \varnothing$, $c_{-1}^1 \leftarrow \varnothing$
$(\mathcal{C}_0^K, S_0) \leftarrow \mathrm{Retrieve}_{\mathcal{R}}(Q_0, \mathcal{C}, K)$
while true **do**
  $(b_t, q_{t+1}) \leftarrow \mathcal{L}^{\mathrm{judge}}(Q_0, q_t, c_t^1, \mathcal{C}_t^K, c_{t-1}^1, \mathcal{H}_t)$
  if $b_t = 1$ **or** $t \ge T_{\max}$ **then**
    return c_t^1
  end if
  $\mathcal{H}_{t+1} \leftarrow \mathcal{H}_t \cup \{(q_t, c_t^1)\}$; $Q_{t+1} \leftarrow (I_q, u, q_{t+1})$
  $(\mathcal{C}_{t+1}^K, S_{t+1}) \leftarrow \mathrm{Retrieve}_{\mathcal{R}}(Q_{t+1}, \mathcal{C}, K)$
  for $v \in \{t, t+1\}$ **do**
    $a_v \leftarrow \mathcal{L}^{\mathrm{answer}}(I_q, u, q_v)$
    $r_v \leftarrow \cos\big(E_{\mathcal{R}}(a_v), E_{\mathcal{R}}(c_v^1)\big)$
  end for
  $\mathcal{I}_t \leftarrow$ candidate IDs in $\mathcal{C}_t^K$
  Construct $P_t, \widetilde{P}_t$ on $\mathcal{I}_t$ using Eq. (5)
  $\Delta r_t \leftarrow r_{t+1} - r_t$; $d_t \leftarrow D_{\mathrm{KL}}(P_t \parallel \widetilde{P}_t)$
  if $\Delta r_t < \delta_{\mathrm{RUPA}}$ **or** $d_t > \delta_{\mathrm{QRV}}$ **then**
    return c_t^1
  end if
  $t \leftarrow t + 1$
end while

Table 2: Ablation study on Qwen3-VL-2B retrieval model with Kimi-K-2.6 as understanding model.

Method	Cls	QA	Ret	GD	Overall
Retrieval Only	70.16	74.39	74.92	88.53	74.96
Baseline	70.19	74.41	75.39	87.33	75.00
+RUPA	71.74	76.22	76.71	89.17	76.58
+QRV	71.43	76.25	76.68	88.70	76.44
InterEvolve	**71.86(↑1.67)**	**76.35(↑1.94)**	**76.82(↑1.43)**	**89.25(↑1.92)**	**76.69(↑1.69)**



(a) Hyperparameter ablation (b) Show case

Figure 3: The left image shows Overall P@1 under different hyperparameters of RUPA and QRV modules. The right image shows one case of iteration details in InterEvolve(Ours) and Baseline.

4 Experiments

4.1 Setup

Dataset. The experiments are conducted on the large-scale multimodal embedding benchmark MMEB-v2 Meng et al. (2025). This benchmark aims to comprehensively evaluate the general representation capabilities of multimodal models, comprising a total of 78 datasets divided into three core sub-task categories. Since the Image sub-task encompasses the broadest range of visual understanding scenarios and offers comprehensive metrics, *in the subsequent extensive ablation studies, we will only present the results on the **MMEB-v2 Image** task.*

Experimental Setup. The overall experimental setup primarily follows the official evaluation protocols of Qwen3-VL-Embedding Li et al. (2026b) and WeMM-Embedding Zhou et al. (2026), and the understanding models are Qwen3.8-Max Qwen Team (2026), Kimi-K-2.6 Kimi Team (2026), Qwen3.8-27B Qwen Team (2026). In our framework, the maximum number of iterations for all experiments is strictly limited to 4, where $\delta_{\text{RUPA}} = 0.3$ and $\delta_{\text{QRV}} = 0.05$.

Overview of Baselines. ***To the best of our knowledge, InterEvolve is the first work to introduce training-free agentic search into large-scale multimodal retrieval.*** Therefore, to validate the effectiveness of InterEvolve framework, we compare it with the following two categories of methods:

(1) **Retrieval Methods**: Several classic state-of-the-art retrieval methods.

(2) **Naive Agentic Search Baseline**: A naive multi-turn iterative search strategy. This method relies on a LVLM to judge the correctness of the Top-1 result returned by the retrieval model; if judged as correct, it triggers "Correct Early Stopping", otherwise, it continues to rewrite the query.

4.2 Overall Evaluations

As shown in Table 1, the results demonstrate that InterEvolve yields consistent and significant performance gains, establishing new state-of-the-art (SOTA) performance. InterEvolve exhibits strong generalizability and scalability regardless of the base model's capacity. It boosts the average score of Qwen3-VL-2B from 73.4 to **75.4**, and WeMM-2B from 77.8 to **78.9**. For larger models, the Qwen3-VL-8B improves to **78.7**, while the WeMM-9B equipped with our method achieves the highest overall score of **80.7** on the entire benchmark. Notably, the enhanced WeMM-2B (78.9) even surpasses several 8B models, fully demonstrating that our approach can substantially unlock

Table 3: Ablation study on different understanding models with Qwen3-VL-2B as retrieval model.

Model	Method	Cls	QA	Ret	GD	Overall
Qwen3.8-27B Qwen Team (2026)	Baseline	68.24	74.01	74.32	88.67	74.14
	Ours	**69.47(↑1.23)**	**75.72(↑1.71)**	**75.76(↑1.44)**	**89.05(↑0.38)**	**75.48(↑1.34)**
Kimi-K-2.6 Kimi Team (2026)	Baseline	70.19	74.41	75.39	87.33	75.00
	Ours	**71.86(↑1.67)**	**76.35(↑1.94)**	**76.82(↑1.43)**	**89.25(↑1.92)**	**76.69(↑1.69)**
Qwen3.8-Max Qwen Team (2026)	Baseline	72.98	77.05	75.70	**89.70**	76.88
	Ours	**73.27(↑0.29)**	**77.49(↑0.44)**	**76.96(↑1.26)**	89.60(↓0.10)	**77.49(↑0.61)**

the potential of multimodal embedding models in a highly parameter-efficient manner, providing orthogonal benefits complementary to parameter scaling.

Beyond overall improvements, our method delivers substantial breakthroughs across specific sub-tasks, which yields the most notable gains in Video tasks. Taking the WeMM-2B as an example, the MRet score surges from 56.9 to **62.0** (+5.1), and the Video overall increases from 70.2 to **71.9**. Even for the highly competitive 9B baseline, M-Ret sees a remarkable +4.1 improvement (56.5 $\rightarrow$ **60.6**). *Across all Image sub-tasks and VisDoc, our method consistently sets new records for each respective baseline.*

4.3 Ablation Study

4.3.1 Ablation of Components

Core Component Ablation. Table 2 reports the experimental results of adding consistency gain guidance (+RUPA), adding distribution shift constraint (+QRV), and combining both (InterEvolve).

As shown in Table 2, Baseline brings only marginal gains compared to Retrieval Only (overall increases from 74.96 to 75.00). Moreover, due to the lack of constraints, its GD metric even experiences degradation. This indicates that the simple Correct Early Stopping strategy easily falls into local optima or introduces noisy rewrites. When the RUPA module is introduced alone, guiding the query evolution direction through consistency gain, the overall performance significantly increase to 76.58. Meanwhile, introducing the QRV module alone also boosts the Overall to 76.44. Ultimately, InterEvolve achieves the global optimum across all dimensions to **76.69**. This proves that RUPA is responsible for providing positive evolutionary momentum, while QRV acts as safeguard to prevent semantic drift; both are indispensable.

Hyperparameter Sensitivity Ablation. Figure 3a illustrates the curves of Overall P@1 varying with hyperparameters and different understanding models.

As shown in Figure 3a(a)(c), as δ_{QRV} gradually decreases from 1 to 0.05, the distributional constraint of QRV tightens, effectively filtering out erroneous rewrites generated by weak understanding models, and overall P@1 exhibits a steady upward trend; whereas when δ_{QRV} further decreases to 0.005, the performance tends to saturate, indicating that excessively strict verification does not bring additional gains and may instead increase computational overhead. Similarly, Figure 3a(b)(d) shows that as δ_{RUPA} increases from 0, the consistency gain guidance provides stronger evolutionary momentum, and the performance saturates around $\delta_{\text{RUPA}} = 0.3$.

4.3.2 Ablation of Understanding Models

This section fixes the retrieval model as Qwen3-VL-2B and integrate Qwen3.8-27B (weak), Kimi-K-2.6 (medium), and Qwen3.8-Max (strong) as understanding models for verification experiments. Observing Table 3, the following three key conclusions can be drawn:

(1) Catastrophic misguidance of weak understanding models on the naive baseline. When using the weakest Qwen3.8-27B, Baseline not only fails to bring gains but leads to a comprehensive collapse in performance. Its overall score plummets from 74.96 to **74.14**, and the Cls metric suffers severe degradation. This intuitively proves that: in the absence of strict distributional constraints, weak understanding models easily introduce semantic drift during multi-turn query rewrites, thereby steering the search trajectory in the wrong direction.

Table 4: Ablation study on different retrieval models with Qwen3.8-Max as understanding model.

Model	Method	Cls	QA	Ret	GD	Overall
Qwen3-VL-2B	Baseline	72.98	77.05	75.70	**89.70**	76.88
	Ours	**73.27(↑0.29)**	**77.49(↑0.45)**	**76.96(↑1.26)**	89.60(↓0.10)	**77.49(↑0.61)**
WeMM-2B	Baseline	75.02	79.54	78.62	94.88	79.68
	Ours	**75.39(↑0.37)**	**80.74(↑1.20)**	**80.52(↑1.90)**	**95.02(↑0.14)**	**80.77(↑1.09)**
Qwen3-VL-8B	Baseline	74.83	80.15	78.97	92.92	79.70
	Ours	**75.32(↑0.49)**	**81.95(↑1.80)**	**80.62(↑0.56)**	**93.22(↑1.65)**	**80.92(↑1.22)**
WeMM-9B	Baseline	75.97	81.27	80.09	95.78	81.02
	Ours	**76.28(↑0.31)**	**82.53(↑1.26)**	**81.94(↑1.85)**	**96.05(↑0.27)**	**82.10(↑1.08)**

(2) Robustness recovery of our method under weak models. Although Qwen3.8-27B has limited inherent capability and is highly prone to misguidance, after introducing InterEvolve, its overall performance is successfully pulled back to **75.48**, which not only completely repairs the performance collapse. This result strongly proves that: even under adverse conditions with weak understanding model capabilities, our method still possesses extreme robustness, capable of preventing system performance degradation and ensuring positive gains.

(3) Scaling Law of understanding model capability and comprehensive optimality. As the capability of the understanding model improves (27B $\rightarrow$ Kimi $\rightarrow$ Max), the overall performance of our method exhibits a clear and stable step-wise ascent (75.48 $\rightarrow$ 76.69 $\rightarrow$ **77.49**). When equipped with Qwen3.8-Max, our method achieves the global optimum on whole tasks. This indicates that our method can both downwardly accommodate the robustness safeguard for weak models and upwardly unleash the performance upper bound of strong models.

4.3.3 Ablation of Retrieval Models

Fixing the understanding model as Qwen3.8-Max, this section evaluates our framework across diverse retrieval backbones at varying scales, as detailed in Table 4.

(1) Architecture-agnostic generalization vs. Baseline instability. InterEvolve yields consistent gains across all four backbones, demonstrating strong generalization independent of specific retrieval architectures. Conversely, Baseline exhibits severe instability: on stronger models, its unconstrained iterations degrade the performance and yield widespread negative gains in QA and Ret, indicating that unguided multi-turn rewrites easily disrupt high-quality initial representations.

(2) Robustness safeguard and performance potential maximization. Across all configurations, our method rectifies the degradation of Baseline and consistently surpasses the Retrieval Only baseline, peaking at **82.10** on WeMM-9B. This confirms that RUPA and QRV adaptively regulate the search process, injecting positive momentum for weaker models while preventing over-rewriting for stronger ones, thereby seamlessly capitalizing on the underlying retrieval scaling law.

5 Conclusion

In this paper, we present InterEvolve, a training-free agentic search framework that improves the reliability of state assessment in self-evolving multimodal retrieval. Our key observation is that humans naturally mitigate misdirected search by forming expectations from prior understanding and assessing the consistency between retrieval results and these expectations. Building on this insight, we propose a retrieval–understanding interaction mechanism that provides two complementary strategies: the Retrieval–Understanding Progress Assessor (RUPA) and the Query Refinement Validator (QRV). Extensive experiments on the MMEB-v2 benchmark across 4 retrieval backbones and 3 understanding models demonstrate that InterEvolve consistently outperforms all baselines. Overall, this work highlights the critical role of reliable state assessment in self-evolving agentic systems and offers a new perspective on leveraging retrieval–understanding interaction for robust multi-turn search. We hope the proposed framework provides insights for future research in both agentic search and test-time adaptation of multimodal models.

REFERENCES

Shuai Bai, Yuxuan Cai, Ruizhe Chen, Keqin Chen, Xionghui Chen, Zesen Cheng, Lianghao Deng, Wei Ding, Chang Gao, Chunjiang Ge, et al. Qwen3-vl technical report. *arXiv preprint arXiv:2511.21631*, 2025.

Zehui Chen, Kuikun Liu, Qiuchen Wang, Jiangning Liu, Wenwei Zhang, Kai Chen, and Feng Zhao. Mindsearch: Mimicking human minds elicits deep ai searcher. In *International Conference On Learning Representations*, volume 2025, pp. 90007–90029, 2025.

Zhe Chen, Jiannan Wu, Wenhai Wang, Weijie Su, Guo Chen, Sen Xing, Muyan Zhong, Qinglong Zhang, Xizhou Zhu, Lewei Lu, et al. Internvl: Scaling up vision foundation models and aligning for generic visual-linguistic tasks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 24185–24198, 2024.

Xuanming Cui, Jianpeng Cheng, Hong-you Chen, Satya Narayan Shukla, Abhijeet Awasthi, Xichen Pan, Chaitanya Ahuja, Shlok Mishra, Taipeng Tian, Qi Guo, et al. Think then embed: Generative context improves multimodal embedding. In *International Conference on Learning Representations*, volume 2026, pp. 2690–2709, 2026.

Boyu Gou, Zanming Huang, Yuting Ning, Yu Gu, Michael Lin, Weijian Qi, Andrei Kopanev, Botao Yu, Bernal Jimenez Gutierrez, Yiheng Shu, et al. Mind2web 2: Evaluating agentic search with agent-as-a-judge. *Advances in Neural Information Processing Systems*, 38, 2026.

Xinyan Guan, Jiali Zeng, Fandong Meng, Chunlei Xin, Yaojie Lu, Hongyu Lin, Xianpei Han, Le Sun, and Jie Zhou. Deeprag: Thinking to retrieve step by step for large language models. In *International Conference on Learning Representations*, volume 2026, pp. 117056–117071, 2026.

Chao Jia, Yinfei Yang, Ye Xia, Yi-Ting Chen, Zarana Parekh, Hieu Pham, Quoc Le, Yun-Hsuan Sung, Zhen Li, and Tom Duerig. Scaling up visual and vision-language representation learning with noisy text supervision. In *International conference on machine learning*, pp. 4904–4916. PMLR, 2021.

Haonan Jia, Shichao Dong, Xin Dong, Zenghui Sun, Jin Wang, Jinsong Lan, Xiaoyong Zhu, Bo Zheng, and Kaifu Zhang. Cross-modal identity mapping: Minimizing information loss in modality conversion via reinforcement learning. *arXiv preprint arXiv:2603.01696*, 2026a.

Haonan Jia, Shichao Dong, Zenghui Sun, Jiawen Zheng, Ziqi Miao, Gege Shi, Qiuyu Zhao, Jinsong Lan, Xiaoyong Zhu, and Bo Zheng. Recap: Retrieval-guided refinement for image captioning enhancement via reinforcement learning. *arXiv preprint arXiv:2608.21305*, 2026b.

Weijian Jian, Yajun Zhang, Dawei Liang, Chunyu Xie, Yixiao He, Dawei Leng, and Yuhui Yin. Rzenembed: Towards comprehensive multimodal retrieval. *arXiv preprint arXiv:2510.27350*, 2025.

Ting Jiang, Minghui Song, Zihan Zhang, Haizhen Huang, Weiwei Deng, Feng Sun, Qi Zhang, Deqing Wang, and Fuzhen Zhuang. E5-v: Universal embeddings with multimodal large language models. *arXiv preprint arXiv:2407.12580*, 2024a.

Ziyan Jiang, Rui Meng, Xinyi Yang, Semih Yavuz, Yingbo Zhou, and Wenhu Chen. Vlm2vec: Training vision-language models for massive multimodal embedding tasks. *arXiv preprint arXiv:2410.05160*, 2024b.

Bowen Jin, Hansi Zeng, Zhenrui Yue, Jinsung Yoon, Sercan Arik, Dong Wang, Hamed Zamani, and Jiawei Han. Search-r1: Training llms to reason and leverage search engines with reinforcement learning. *arXiv preprint arXiv:2503.09516*, 2025.

Kimi Team. Kimi k2.6: Advancing open-source coding, April 2026. URL `https://www.kimi.com/en/blog/kimi-k2-6`.

Jian Li, Yizhang Jin, Dongqi Liu, Hang Ding, Jiafu Wu, Dongsheng Chen, Yunhang Shen, Yulei Qin, Ying Tai, Chengjie Wang, et al. Se-search: Self-evolving search agent via memory and dense reward. *arXiv preprint arXiv:2603.03293*, 2026a.

Mingxin Li, Yanzhao Zhang, Dingkun Long, Keqin Chen, Sibo Song, Shuai Bai, Zhibo Yang, Pengjun Xie, An Yang, Dayiheng Liu, et al. Qwen3-vl-embedding and qwen3-vl-reranker: A unified framework for state-of-the-art multimodal retrieval and ranking. *arXiv preprint arXiv:2601.04720*, 2026b.

Xiaoxi Li, Guanting Dong, Jiajie Jin, Yuyao Zhang, Yujia Zhou, Yutao Zhu, Peitian Zhang, and Zhicheng Dou. Search-o1: Agentic search-enhanced large reasoning models. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pp. 5420–5438, 2025.

Zhuofeng Li, Haoxiang Zhang, Seungju Han, Sheng Liu, Jianwen Xie, Yu Zhang, Yejin Choi, James Y Zou, and Pan Lu. In-the-flow agentic system optimization for effective planning and tool use. In *International Conference on Learning Representations*, volume 2026, pp. 50524–50570, 2026c.

Bin Lin, Zhenyu Tang, Yang Ye, Jinfa Huang, Junwu Zhang, Yatian Pang, Peng Jin, Munan Ning, Jiebo Luo, and Li Yuan. Moe-llava: Mixture of experts for large vision-language models. *IEEE Transactions on Multimedia*, 2026.

Kailin Lyu, Zhiqiang Yuan, Jianwei He, Qiwei Yan, Xuanbo Su, Nanxing Hu, Yang Liu, Ce Hao, Shengqian Qin, Lianyu Hu, et al. Photocraft: Agentic reasoning with hierarchical self-evolving memory for deep image search. *arXiv preprint arXiv:2606.03099*, 2026.

Rui Meng, Ziyan Jiang, Ye Liu, Mingyi Su, Xinyi Yang, Yuepeng Fu, Can Qin, Zeyuan Chen, Ran Xu, Caiming Xiong, et al. Vlm2vec-v2: Advancing multimodal embedding for videos, images, and visual documents. *arXiv preprint arXiv:2507.04590*, 2025.

Qwen Team. Qwen3.8-max: A new bar for coding and cowork, August 2026. URL `https://qwen.ai/blog?id=qwen3.8`.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pp. 8748–8763. PmLR, 2021.

Zhengliang Shi, Lingyong Yan, Dawei Yin, Suzan Verberne, Maarten Rijke, and Zhaochun Ren. Iterative self-incentivization empowers large language models as agentic searchers. *Advances in Neural Information Processing Systems*, 38:135679–135726, 2026.

Weiyun Wang, Zhangwei Gao, Lixin Gu, Hengjun Pu, Long Cui, Xingguang Wei, Zhaoyang Liu, Linglin Jing, Shenglong Ye, Jie Shao, et al. Internvl3. 5: Advancing open-source multimodal models in versatility, reasoning, and efficiency. *arXiv preprint arXiv:2508.18265*, 2025.

Jinming Wu, Zihao Deng, Wei Li, Yiding Liu, Bo You, Bo Li, Zejun Ma, and Ziwei Liu. Mmsearch-r1: Incentivizing lmms to search. *arXiv preprint arXiv:2506.20670*, 2(4), 2025.

Zhenrui Yue, Kartikeya Upasani, Xianjun Yang, Suyu Ge, Shaoliang Nie, Yuning Mao, Zhe Liu, and Dong Wang. Dr. zero: Self-evolving search agents without training data. *arXiv preprint arXiv:2601.07055*, 2026.

Aohan Zeng, Xin Lv, Zhenyu Hou, Zhengxiao Du, Qinkai Zheng, Bin Chen, Da Yin, Chendi Ge, Chenghua Huang, Chengxing Xie, et al. Glm-5: from vibe coding to agentic engineering. *arXiv preprint arXiv:2602.15763*, 2026.

Xiaohua Zhai, Basil Mustafa, Alexander Kolesnikov, and Lucas Beyer. Sigmoid loss for language image pre-training. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 11941–11952. IEEE, 2023.

Xin Zhang, Yanzhao Zhang, Wen Xie, Mingxin Li, Ziqi Dai, Dingkun Long, Pengjun Xie, Meishan Zhang, Wenjie Li, and Min Zhang. Gme: Improving universal multimodal retrieval by multimodal llms. *arXiv preprint arXiv:2412.16855*, 2024.

Junjie Zhou, Yongping Xiong, Zheng Liu, Ze Liu, Shitao Xiao, Yueze Wang, Bo Zhao, Chen Jason Zhang, and Defu Lian. Megapairs: Massive data synthesis for universal multimodal retrieval. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 19076–19095, 2025.

Junjie Zhou, Ke Mei, Lei Li, Tianyi Wang, Fengyun Rao, and Jing Lyu. Wemm-embedding: Wechat multi-modal embedding technical report. *arXiv preprint arXiv:2608.24053*, 2026.

A Appendix

A.1 Check Lists

A.1.1 AI use statement

In this work, we utilized generative AI tools to polish the paper and assist with programming. We did not employ generative AI for any other tasks requiring disclosure, such as retrieving related works, generating synthetic datasets, proving mathematical claims, drafting sections of paper and so on. All AI-assisted content has been thoroughly reviewed by the authors. Specifically, LLM-generated code was verified and tested for correctness by 2 authors. We take full responsibility for the final content of this work, including all text, claims, and artifacts produced with the assistance of generative AI.