THINK BEFORE YOU LINK: ON-POLICY DISTILLATION ENHANCED GRAPH REASONING FOR MULTIMODAL KNOWLEDGE GRAPH COMPLETION

Anonymous authors
Paper under double-blind review

ABSTRACT

Multimodal knowledge graph completion (MKGC) predicts missing links from graph structure, entity images, and textual descriptions. Existing methods first construct static multimodal entity representations and then predict links by matching these representations. This representation–prediction procedure does not create a query-specific reasoning process to interpret the queried relation and assemble supporting multimodal evidence before prediction. We propose **MKGC-CoT** (Chain-of-Thought), which reformulates MKGC as reasoning followed by entity prediction. MKGC-CoT uses Multimodal Fusion to prepare the query context, MKG Chain-of-Thought (CoT) Reasoning to assemble relation-specific evidence, and Reasoning-Based Entity Prediction to rank all entities from the hidden state after the CoT. We construct Answer-Free MKG CoT Data, in which the student receives neither the gold edge nor the gold answer and an answer-aware teacher provides structured traces only during training. Two-Stage Training first uses Supervised Fine-Tuning (SFT) to learn the reasoning format, then uses Answer-Aware On-Policy Distillation (OPD) to strengthen reasoning on the student's own sampled CoT prefixes. At inference, the student receives only the answer-free query. MKGC-CoT leads on seven of eight dataset–metric combinations across DB15K and MKG-Y. It reaches **50.44%** MRR on DB15K, an **18.2%** relative improvement over the strongest listed baseline, DSoM, and **40.21%** MRR on MKG-Y, where it leads on MRR and most Hits@K metrics.

1 INTRODUCTION

Knowledge graphs (KGs) encode facts as $(\textit{head}, \textit{relation}, \textit{tail})$ triples and support search, recommendation, and question answering. Real KGs are incomplete, so *multimodal knowledge graph completion* (MKGC) predicts a missing head or tail from graph structure, entity images, and textual descriptions. A completion query often requires more than entity similarity. For $(h, r, ?)$, the model must identify the entity type, interpret relation r, gather supporting evidence from the graph and modalities, and rule out conflicting evidence. A correct completion therefore requires a reasoning process that assembles query-specific evidence before it predicts an entity (Figure 1(b)). The relevance of the same modality also changes with the query: an image, a textual attribute, or a graph neighbor can support one relation but be irrelevant or misleading for another. The model must therefore interpret multimodal evidence in the context of the queried relation rather than treat it as a fixed property of the entity.

Existing MKGC methods treat completion as a representation–prediction task (Figure 1(a)). They first encode structural, visual, and textual signals into entity vectors, then use geometric interactions to predict candidate triples (Bordes et al., 2013; Lu et al., 2022; Xu et al., 2022; Zhang et al., 2025; 2024a; Niu & Zhang, 2025; Huang et al., 2026; Li et al., 2026). Vision–language and language model methods improve these entity representations (Ma et al., 2025; Huang et al., 2025), but they retain the same decision rule: a scorer directly predicts from a pre-computed representation. The fundamental problem is that this procedure predicts a missing entity without reasoning about which evidence supports the current relation. Better fusion improves what the vector stores, but it does not tell the model which evidence to trust for the current relation. The model cannot explicitly use a

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.

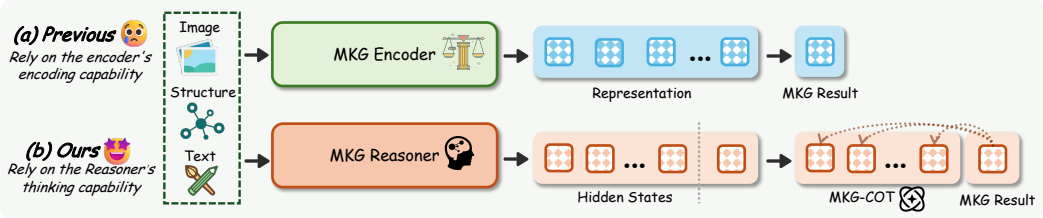

Figure 1: Connecting multimodal evidence to entity prediction. (a) Representation-based MKGC fuses evidence into embeddings for entity scoring. (b) MKGC-CoT jointly trains structured response generation and entity ranking, allowing response states to refine candidate scores.

relevant image, text attribute, or graph neighbor, and it cannot reject a conflicting one. It can then rank a similar but incorrect entity above the answer that the query evidence supports.

To address this limitation, we reformulate MKGC as a reasoning task followed by entity prediction and propose **MKGC-CoT**, a *reason-then-rank* framework (Figure 2). Multimodal Fusion prepares the structural, visual, and textual context for the query. MKG Chain-of-Thought (CoT) Reasoning (Wei et al., 2022) then makes the model identify the entity type, interpret the relation, and assemble supporting evidence in an explicit CoT. Reasoning-Based Entity Prediction ranks every candidate entity from the hidden state after the CoT, so the final prediction depends on the reasoning result instead of a static entity vector. We construct Answer-Free MKG CoT Data because standard MKGC benchmarks provide triples but not reasoning traces; the student receives neither the gold edge nor the gold answer, while an answer-aware teacher uses the answer only to create a structured target trace. Two-Stage Training first uses SFT to teach the required reasoning format, then uses Answer-Aware On-Policy Distillation (OPD) to strengthen reasoning on the student's own sampled CoT prefixes. During OPD, a frozen teacher that observes the gold entity provides next-token targets only during training (Agarwal et al., 2024; Jiang et al., 2026).

MKGC-CoT leads on seven of eight dataset–metric combinations (Table 1). On DB15K, it achieves **50.44%** MRR, exceeding the strongest listed baseline, DSoM (**42.68%**), by **7.76** percentage points (**18.2%** relative), and leads on all four metrics. On MKG-Y, Answer-Aware OPD raises MRR from **35.93%** after SFT to **40.21%**, a **4.28**-point gain, surpassing LBMKGC (**40.03%**). It leads on MRR and most Hits@K metrics.

Our contributions are:

- We propose MKGC-CoT (Chain-of-Thought), which turns MKGC from direct representation–prediction into reasoning–prediction. Multimodal Fusion, MKG Chain-of-Thought (CoT) Reasoning, and Reasoning-Based Entity Prediction make the final ranking depend on query-specific reasoning.
- To provide reasoning supervision without answer leakage, we construct CoT data in which the student never receives the gold edge or answer, while the teacher uses the answer only to create structured reasoning traces.
- To strengthen inference-time reasoning, we design Two-Stage Training with Supervised Fine-Tuning (SFT) followed by Answer-Aware On-Policy Distillation (OPD). OPD trains the student on its own sampled CoT prefixes.
- On DB15K and MKG-Y, MKGC-CoT leads on seven of eight dataset–metric combinations. It improves DB15K MRR by **18.2%** relative to the strongest listed baseline and leads on MRR and most Hits@K metrics on MKG-Y.

2 Related Work

Multimodal Knowledge Graph Completion. Knowledge graph completion usually embeds entities and relations and predicts a triple with a geometric interaction, as in TransE (Bordes et al., 2013). Multimodal KGC adds visual and textual entity information (Chen et al., 2024). Prior work improves robust multimodal embeddings, negative sampling, modality tokens, or expert routing (Lu et al., 2022; Xu et al., 2022; Niu & Zhang, 2025; Zhang et al., 2025; 2024a); recent approaches

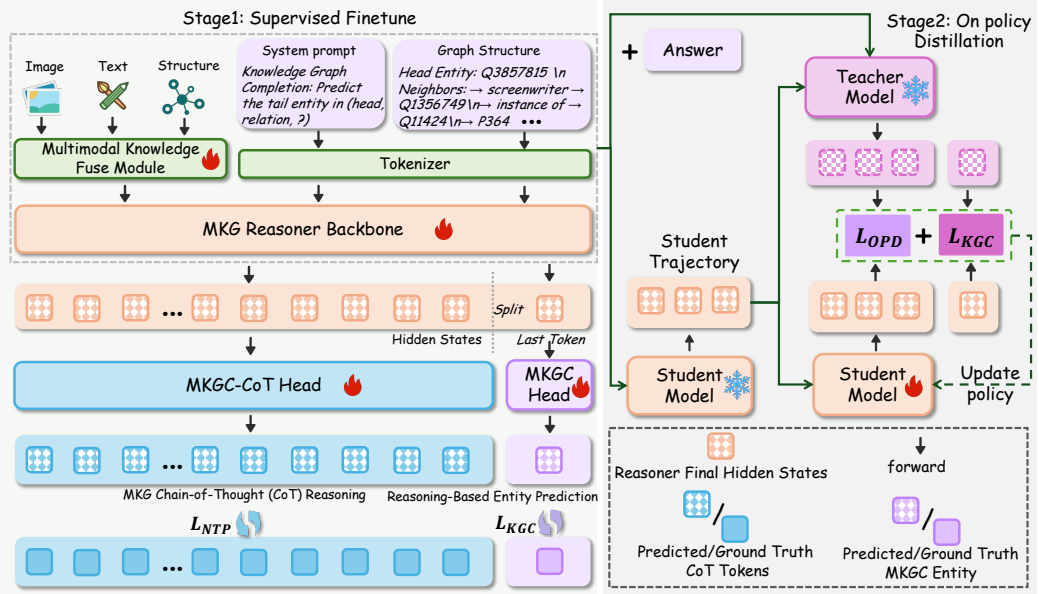

Figure 2: MKGC-CoT architecture and training. Left: joint SFT for token prediction and entity ranking. Right: answer-aware distillation on student trajectories with ranking supervision. The diagram uses $\mathcal{L}_{\text{NTP}}$ for $\mathcal{L}_{\text{LM}}$ and $\mathcal{L}_{\text{KGC}}$ for $\mathcal{L}_{\text{score}}$; its $\mathcal{L}_{\text{OPD}}$ denotes only the token-distillation term in Stage 2. Flames and snowflakes mark trainable and frozen components.

also study structure-aware generation, cross-modal complementarity, and vision–language representations (Huang et al., 2026; Li et al., 2026; Ma et al., 2025; Huang et al., 2025). Despite these differences, the dominant procedure remains representation–prediction: a method first constructs entity representations and then matches them to predict a link. This procedure does not explicitly determine which multimodal evidence supports the queried relation before it ranks a candidate.

Large Language Models for Knowledge Graph Completion. LLM-based KGC and KG reasoning use graph context beyond a single entity embedding. KICGPT ranks retrieved candidates from textual graph knowledge (Wei et al., 2023); neural-symbolic approaches compose observed paths (Liu et al., 2025); and Think-on-Graph and Reasoning on Graphs retrieve and traverse graph evidence step by step (Sun et al., 2024; Luo et al., 2024). These methods mainly reason over textual graphs and observed edges, and retrieval-based KGC is limited by candidate recall. They do not use graph, image, and text jointly to predict a missing edge over the full entity set.

Chain-of-Thought Reasoning and Distillation. Chain-of-thought (CoT) prompting elicits intermediate reasoning steps from language models (Wei et al., 2022); multimodal CoT methods use external knowledge to improve reasoning over multimodal inputs (Mondal et al., 2024). In these settings, CoT typically explains an answer or guides generation, rather than serving as the state for full-entity MKGC ranking. Standard SFT learns from fixed teacher traces, although inference continues from student-generated prefixes. On-policy distillation reduces this mismatch by training on the learner's rollouts (Agarwal et al., 2024; Jiang et al., 2026). These methods do not specify how to learn answer-free multimodal graph reasoning when the student must generate a trace and rank an entity without access to the gold answer.

3 Method

3.1 Overview

A multimodal knowledge graph $\mathcal{G} = (\mathcal{E}, \mathcal{R}, \mathcal{T})$ contains entities, relations, and observed triples, with structural, visual, and textual information for each entity. For $(h, r, ?)$, we rank all entities under the filtered protocol; head prediction uses the inverse query $(t, r^{-1}, ?)$. MKGC-CoT combines

multimodal fusion, CoT reasoning, and entity scoring in a large language model (LLM), trained by supervised fine-tuning (SFT) followed by answer-aware on-policy distillation (OPD) (Figure 2).

3.2 Multimodal Knowledge Fusion

For entity e, $\mathbf{x}_e^s$ is a learnable structural embedding, while $\mathbf{x}_e^v$ and $\mathbf{x}_e^t$ are visual and textual features from frozen encoders. Projections $\mathbf{W}_m$ and an attention vector $\mathbf{a}$ produce content-dependent fusion weights:

$$\alpha_e^m = \frac{\exp\big(\mathbf{a}^\top \tanh(\mathbf{W}_m \mathbf{x}_e^m)\big)}{\sum_{m' \in \{s,v,t\}} \exp(\mathbf{a}^\top \tanh(\mathbf{W}_{m'} \mathbf{x}_e^{m'}))}. \tag{1}$$

$$\mathbf{z}_e = \sum_{m \in \{s,v,t\}} \alpha_e^m \, \mathbf{W}_m \mathbf{x}_e^m, \tag{2}$$

The fused representation $\mathbf{z}_e$ supplies the query's multimodal prefix and the candidate embeddings used for ranking (Sec. 3.4). Sharing this space connects multimodal conditioning to entity prediction, while learned weights adjust each modality's contribution.

3.3 MKG Chain-of-Thought (CoT) Reasoning

The instruction specifies the known entity, relation and its description, and graph neighbors. We project $\mathbf{z}_h$ to the LLM hidden size and prepend it to the instruction. The LLM generates a CoT (Wei et al., 2022), forming $[\,\mathbf{z}_h\,][\,\text{instruction}\,][\,\text{CoT}\,]$. Four ordered fields—entity type, relation meaning, evidence, and answer—place relation interpretation and evidence synthesis before answer selection. The multimodal prefix remains available at every decoding step, conditioning reasoning alongside the graph context.

3.4 Reasoning-Based Entity Prediction

Let $\mathbf{h}_{\text{eoc}}$ be the last-layer hidden state at the end of the generated CoT. We project it to $\mathbf{q} = \mathbf{W}_q \mathbf{h}_{\text{eoc}}$ and score every candidate entity through its fused representation:

$$\phi(e \mid h, r) = \big(g_r(\mathbf{q})^\top \mathbf{z}_e\big) \cdot \tau_r + b_e, \tag{3}$$

where g_r is a relation-conditioned transform, τ_r a relation-specific scale, and b_e an entity bias. For multiple valid answers, $\tilde{y}_e$ distributes mass uniformly over the known tails of (h, r) in training triples only. We optimize multi-positive cross-entropy over all $|\mathcal{E}|$ candidates:

$$\mathcal{L}_{\text{score}} = -\sum_{e \in \mathcal{E}} \tilde{y}_e \, \log \text{softmax}_e\big(\phi(e \mid h, r)\big). \tag{4}$$

Ranking thus uses a reasoning-conditioned state, while multi-positive supervision avoids conflicting one-hot targets. Full-entity scoring also avoids restricting predictions to candidate names explicitly mentioned in the generated CoT or a retrieved shortlist.

3.5 Answer-Free MKG CoT Data

We construct CoT supervision from training triples (Figure 3). Each triple (h, r, t) yields tail and head queries, $(h, r, ?)$ and $(?, r, t)$. Context contains at most eight one-hop neighbors from the training split, verbalized as directed edges with readable relation labels; the gold edge (h, r, t) is removed. The teacher view appends the gold answer, a one-shot example, and a four-line format requirement. The student instruction omits the answer, matching the input format at inference.

A frozen teacher produces the four-field trace. We retain records only when both query directions are completed, remove text before "Entity type:", and check format (99.99% pass) and whether the Answer field matches the gold entity (97.0%). Deduplication and export yield DB15K-CoT (116,960 traces) and MKG-Y-CoT (42,620 traces). Each record stores an answer-free instruction, its trace, and the gold entity, which serves as teacher-side supervision and a ranking label. Answer-free refers to the student instruction: the supervised trace still contains an Answer field, and training triples provide multi-positive ranking supervision.

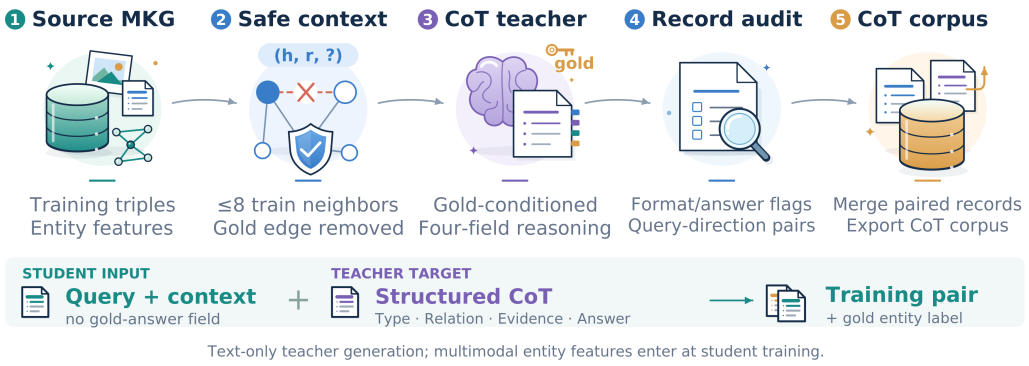

Figure 3: Answer-conditioned trace construction (schematic): the teacher receives a training answer; the student instruction omits it.

3.6 Two-Stage Training: SFT and Answer-Aware OPD

Stage 1: Supervised fine-tuning. We train lightweight backbone adapters with language modeling and entity ranking. For an answer-free query q and teacher CoT target $\mathbf{c}_{1:L}$, teacher forcing gives

$$\mathcal{L}_{\text{LM}} = -\frac{1}{L}\sum_{i=1}^{L} \log p_\theta\left(c_i \mid q, \mathbf{c}_{<i}\right). \tag{5}$$

The SFT objective is

$$\mathcal{L}_{\text{SFT}} = \mathcal{L}_{\text{LM}} + \alpha\, \mathcal{L}_{\text{score}}. \tag{6}$$

Here α weights ranking supervision. Joint training teaches response generation and makes the final CoT state useful for candidate discrimination.

Stage 2: Answer-aware on-policy distillation. To address the mismatch between teacher-forced SFT and inference on generated CoT, OPD trains on student trajectories (Agarwal et al., 2024; Jiang et al., 2026). An early generation error can change later prefixes, so teacher-written traces alone may not cover the contexts encountered during inference. Let $\mathbf{z}_{1:L} \sim \pi_\theta(\cdot \mid q)$ be a sampled CoT. At each prefix $\mathbf{z}_{<i}$, the frozen teacher additionally receives the gold entity, while the student uses the same prefix without the supplied answer (Figure 2, right). We minimize

$$\mathcal{L}_{\text{OPD}} = \mathbb{E}_{\mathbf{z}_{1:L}\sim\pi_\theta}\left[\frac{1}{L}\sum_{i=1}^{L} \text{KL}\left(p_{\text{T}}^{(i)} \,\|\, p_{\text{S}}^{(i)}\right)\right] + \beta\, \mathcal{L}_{\text{score}}, \tag{7}$$

where $p_{\text{T}}^{(i)} = p_{\text{teacher}}(\cdot \mid q, y^*, \mathbf{z}_{<i})$, $p_{\text{S}}^{(i)} = p_{\text{student}}(\cdot \mid q, \mathbf{z}_{<i})$, and y^* is the training gold entity. Gradients flow only through the student distribution; the ranking term, weighted by β, retains entity supervision during distillation. Neither the teacher nor the supplied gold answer is used at inference.

4 Experiments

4.1 Experimental Setup

Datasets and evaluation. We evaluate on DB15K (12,842 entities, 279 relations) and MKG-Y (15,000 entities, 28 relations) (Xu et al., 2022; Zhang et al., 2024a), using graph structure and precomputed visual/textual features. We report filtered MRR and Hits@K ($K \in \{1, 3, 10\}$), averaged over head and tail prediction. Training positives and query neighbors use training triples only; other known answers are filtered at evaluation. Table values are percentages; **bold** and underlining mark the best and second-best results, including ties. A dash (—) indicates that the original paper did not report the corresponding metric for the stated configuration.

Baselines. Table 1 compares six unimodal baselines: TransE (Bordes et al., 2013), DistMult (Yang et al., 2015), ComplEx (Trouillon et al., 2016), RotatE (Sun et al., 2019), GC-OTE (Tang et al.,

Table 1: Link prediction results (%) on DB15K and MKG-Y. The best result in each column is **bold**, and the second-best is underlined. Higher is better for all metrics.

	DB15K				MKG-Y			
Method	MRR	Hits@1	Hits@3	Hits@10	MRR	Hits@1	Hits@3	Hits@10
Unimodal KGC								
TransE	24.86	12.78	31.48	47.07	30.73	23.45	35.18	43.37
DistMult	23.03	14.78	26.28	39.59	28.71	22.26	27.80	35.95
ComplEx	27.48	18.37	31.57	45.37	25.04	19.33	32.12	40.93
RotatE	29.28	17.87	36.12	49.66	34.95	29.10	38.35	45.30
GC-OTE	31.85	22.11	36.52	51.18	32.95	26.77	36.44	44.08
TuckER	30.10	20.50	35.70	46.80	37.10	34.60	38.40	42.90
Multimodal KGC								
IKRL	26.82	14.09	34.93	49.09	33.22	30.37	34.28	38.26
TBKGC	28.40	15.61	37.03	49.86	33.99	30.47	35.27	40.07
TransAE	28.09	21.25	31.17	41.17	28.10	25.31	29.10	33.03
MMKRL	26.81	13.85	35.07	49.39	36.81	31.66	39.79	45.31
RSME	29.76	24.15	32.12	40.29	34.44	31.78	36.07	39.09
VBKGC	30.61	19.75	37.18	49.44	37.04	33.76	38.75	42.30
OTKGE	23.86	14.85	25.89	34.83	35.51	31.97	37.18	41.38
IMF	32.25	24.20	36.06	48.19	35.79	32.95	37.10	40.60
AdaMF-MAT	32.51	21.31	39.67	51.68	38.06	33.49	40.44	45.48
VISTA	30.42	22.49	33.56	45.94	30.45	24.87	32.40	41.50
MyGO	37.72	30.08	41.26	52.21	38.44	35.01	39.84	44.19
MoMoK	39.57	32.38	43.45	54.14	37.91	35.09	39.20	43.20
MCKGC	39.79	31.92	43.80	54.66	38.06	35.49	40.57	45.21
NativE (S+I+T)	36.74	26.87	—	54.65	39.04	34.79	—	—
LBMKGC	37.23	27.78	42.75	54.71	<u>40.03</u>	33.89	<u>43.11</u>	<u>50.81</u>
DSoM	42.68	34.45	46.62	58.12	40.02	35.67	41.70	47.71
HFR-MKGC	35.20	23.04	43.25	56.41	39.00	34.41	42.12	47.31
MDBGF	40.58	32.77	44.23	55.53	39.13	35.15	40.78	46.06
MEIF	40.49	33.95	45.08	57.57	39.87	**37.11**	42.13	47.68
M-Hyper	41.25	33.64	45.01	56.09	39.46	<u>36.02</u>	40.92	45.22
MKGC-CoT (SFT, ours)	<u>48.71</u>	<u>41.37</u>	<u>52.67</u>	<u>62.25</u>	35.93	34.51	39.67	44.41
MKGC-CoT (OPD, ours)	**50.44**	**43.46**	**54.02**	**63.78**	**40.21**	35.67	**44.88**	**52.72**

2020), and TuckER (Balažević et al., 2019). The twenty multimodal baselines are IKRL (Xie et al., 2017), TBKGC (Mousselly-Sergieh et al., 2018), TransAE (Wang et al., 2019), MMKRL (Lu et al., 2022), RSME (Wang et al., 2021), VBKGC (Zhang & Zhang, 2022), OTKGE (Cao et al., 2022), IMF (Li et al., 2023), AdaMF-MAT (Zhang et al., 2024c), VISTA (Lee et al., 2023), MyGO (Zhang et al., 2025), MoMoK (Zhang et al., 2024a), MCKGC (Gao et al., 2025), NativE (S+I+T) (Zhang et al., 2024b), LBMKGC (Guo et al., 2025), DSoM (Zhao et al., 2025), HFR-MKGC (Wang et al., 2026), MDBGF (Peng et al., 2026), MEIF (Liu et al., 2026a), and M-Hyper (Liu et al., 2026b).

Implementation and controlled comparisons. The controlled implementation uses a Qwen3.5-4B student, a frozen Qwen3.5-35B-A3B teacher, rank-16 LoRA, a 256-dimensional fusion space, and AdamW. Controlled variants share the feature pipeline, candidate set, and full-entity evaluator; checkpoints are selected by validation MRR. Query contexts remove the gold edge and include at most eight one-hop neighbors from the training graph, retaining other observed edges as context. Dataset splits and trace coverage appear in Appendix A.2; training-objective details are given in Appendix A.1 and Appendix A.3.

4.2 Main Results

Table 1 compares MKGC-CoT with unimodal and multimodal KGC baselines on DB15K and MKG-Y. For NativE, we use only structure, image, and text (S+I+T), excluding its additional numeric modality. LAFA evaluates FB15k-237-IMG and WN18-IMG rather than these two benchmarks (Shang et al., 2024), so it is not included in the quantitative table. Baseline-specific settings are summarized in Appendix A.4.

Table 2: MKG-Y ablations (%). The best values are **bold**, and the second-best are underlined. ✓/✗: enabled/disabled; no CoT also removes $\mathcal{L}_{\text{LM}}$.

CoT	SFT	OPD	$\mathcal{L}_{\text{score}}$	$\mathcal{L}_{\text{LM}}$	MRR ↑	H@1 ↑	H@3 ↑	H@10 ↑
✓	✓	✓	✓	✓	**40.21**	**35.67**	**44.88**	**52.72**
✗	✓	✓	✓	✗	23.42	20.56	25.09	29.96
✓	✗	✓	✓	✓	29.54	25.17	31.38	37.13
✓	✓	✗	✓	✓	<u>35.93</u>	<u>34.51</u>	<u>39.67</u>	<u>44.41</u>
✓	✓	✓	✗	✓	22.33	17.95	23.68	30.24
✓	✓	✓	✓	✗	28.33	23.58	30.54	37.07

DB15K. Since SFT already exceeds the listed baselines on all four metrics, the DB15K advantage cannot be credited entirely to OPD. Distillation refines a strong supervised ranker, improving broader coverage without sacrificing first-choice accuracy. This separates the contribution of the second stage from the strength of its starting system. Attributing the SFT advantage specifically to trace content, rather than backbone or feature choices, is not supported by the main-table comparison alone.

MKG-Y. MKG-Y separates broader coverage from first-choice discrimination: OPD leads in MRR and Hits@3/10 but trails MEIF and M-Hyper at Hits@1. Under filtered evaluation, the Hits@10–Hits@1 gap corresponds to about 17% of queries whose gold entities are in the top ten but not first. This motivates testing relation-specific discrimination among near-top candidates, rather than treating coverage as complete disambiguation. Unlike DB15K, SFT alone does not lead the baselines, indicating a dataset-dependent role for the second stage; paired rank transitions would identify which queries it repairs or regresses.

4.3 Ablation Study

Table 2 compares the full MKG-Y model with five ablated variants, using the same full-model result as Table 1.

Joint supervision. Without ranking supervision, MRR is weakest, consistent with token prediction not directly training the full-entity scoring objective. Conversely, removing LM supervision shows that ranking labels alone do not recover full performance in this configuration. This supports joint supervision but does not decide whether language targets help through trace semantics or shared-state representation learning.

Training stages and answer-aware OPD. Removing OPD yields 35.93% MRR, whereas removing SFT yields 29.54%, supporting SFT as the starting policy for rollout-based refinement. OPD improves the SFT model by 4.28 MRR points. Its student-prefix supervision addresses the mismatch between teacher-forced response states during SFT and student-generated states at inference, without requiring reproduction of a complete reference trajectory. The gains are consistent with making generated responses more useful for ranking, but do not establish greater explanation faithfulness or separately identify the effects of extra optimization, student-prefix supervision, and teacher access to the gold answer. The gold remains teacher-side training information and is unavailable to the student at test time.

Scope of the ablations. The no-CoT condition removes both LM supervision and the original scoring state, testing removal of the whole pathway rather than the causal value of reasoning text. The current ablations therefore support the pathway as a whole, without isolating the contribution of intermediate trace semantics.

Joint training can help while a model still exploits spurious visual or textual cues. The robustness settings studied by NativE and MDBGF (Zhang et al., 2024b; Peng et al., 2026) concern a different question from these component ablations: dependence on a modality does not establish rejection of its misleading content.

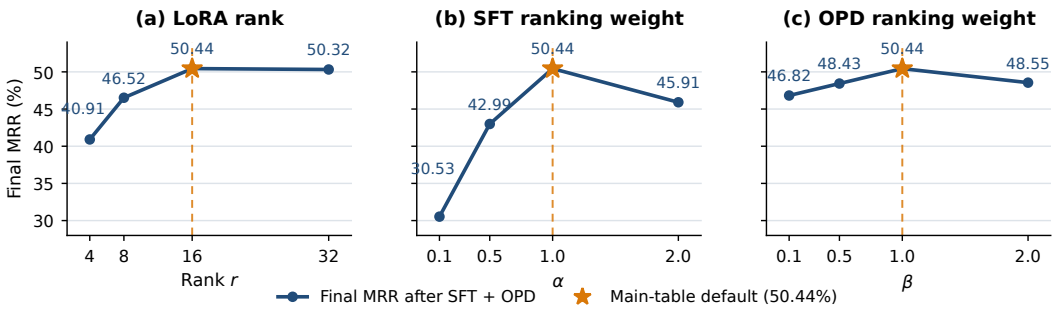

Figure 4: Final DB15K MRR after SFT+OPD, varying LoRA rank (left), SFT ranking-loss weight α (middle), and OPD ranking-loss weight β (right). Orange stars mark the main-table defaults: LoRA rank 16 and $\alpha = \beta = 1$, each yielding 50.44% MRR.

4.4 Parameter Sensitivity Analysis

Figure 4 evaluates adapter capacity and objective balance on DB15K using final MRR after SFT and OPD. The α sweep changes the entity-ranking loss weight during SFT, then completes OPD; β controls the ranking-loss weight during OPD. All three default points match the 50.44% MRR of MKGC-CoT (OPD) in Table 1.

LoRA rank. As shown in Figure 4(a), increasing the rank from 4 to 16 raises final MRR from 40.91% to 50.44%. Doubling the rank to 32 doubles the LoRA matrix parameters but yields 50.32%, a decrease of 0.12 percentage points. Rank 16 therefore achieves the highest MRR among the tested ranks while using fewer adapter parameters than rank 32.

SFT ranking-loss weight. As shown in Figure 4(b), weak ranking supervision during SFT lowers final MRR even after OPD: $\alpha = 0.1$ yields 30.53%, compared with 50.44% at $\alpha = 1$. Increasing α further to 2 lowers MRR to 45.91%, a drop of 4.53 percentage points. Both smaller and larger weights reduce final performance relative to $\alpha = 1$, supporting this default among the tested settings.

OPD ranking-loss weight. As shown in Figure 4(c), increasing β from 0.1 to 1 raises final MRR from 46.82% to 50.44%, while increasing it to 2 lowers MRR to 48.55%, a 1.89-point drop. This decrease is smaller than the 4.53-point drop in the α sweep, but both sweeps favor a weight of 1 among the tested values. These comparisons concern final ranking accuracy and do not establish the sensitivity of trace quality or optimization stability (see Appendix A.5 for the scope of this analysis).

4.5 Case Study

Figure 5 presents four held-out MKG-Y query/gold pairs, chosen to separate evidence-selection problems that aggregate ranking metrics cannot identify. These dataset-grounded schematics illustrate diagnostic situations, not verified student reasoning trajectories.

Misleading neighborhood. Two edges link Jonathan Bottinelli to the same club, Adana Demirspor: repeated association is not independent citizenship evidence. Transferring the club's location to the player would confuse the neighbor's relation with the queried relation. Neighbor usefulness thus depends on whether its semantics license the inference, not just entity association; rejecting this shortcut still supplies no positive evidence for Argentina.

Empty neighborhood. Discovery Channel separates structural isolation from informational isolation. With no incident training edges, a successful prediction could not be credited to a supplied neighborhood path, yet entity names, precomputed features, and pretrained knowledge remain potential sources. Zero training degree is therefore not a sufficient definition of a "no-evidence" test; attributing a correct Maryland prediction requires channel-specific evidence, not degree statistics alone.

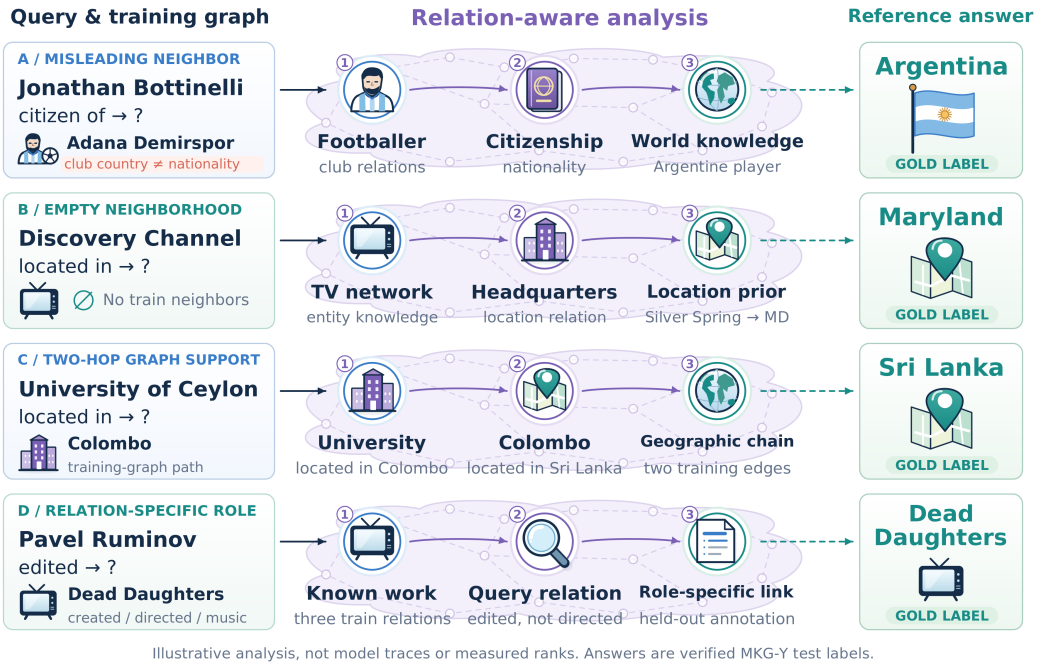

Figure 5: Four dataset-grounded MKG-Y schematics: (A) misleading affiliation, (B) an empty neighborhood, (C) a geographic training-graph path, and (D) distinct creative roles. Paths are illustrative and answers are test labels, not measured predictions. Icons are not model-input images.

Geographic graph support. The University of Ceylon–Colombo–Sri Lanka path distinguishes graph availability from input availability. Under gold-edge removal, the University of Ceylon–Colombo edge remains eligible for the university's one-hop context. The Colombo–Sri Lanka edge is a second hop, so the context rule does not guarantee that the full path is supplied. Checking explanations against the full graph can thus endorse evidence absent from the verbalized input. Assessing explicit path use requires tracking edges through filtering and prompt construction, not merely finding a supporting path after observing the answer.

Relation-specific roles. All three of Pavel Ruminov's neighboring roles point to *Dead Daughters*, making candidate identification easier than relation verification. A relation-blind matcher could select this sole associated work without understanding the held-out *edited* relation. Even replacing the query with *directed* would not distinguish those mechanisms: the same work is a known correct answer for both relations. A discriminating test needs comparably associated works with different verified roles; correctness on this example alone cannot establish that editing was distinguished from other creative relations.

Together, these cases separate source availability, relation validity, and evidence provenance. They illustrate why a correct entity alone does not establish a faithful explanation. Their dataset provenance and interpretive limits are detailed in Appendix A.6.

5 Conclusion

We propose MKGC-CoT, which changes MKGC from direct representation–prediction to *reason-then-rank*. Multimodal Fusion prepares the query context, MKG Chain-of-Thought (CoT) Reasoning produces an explicit trace, and Reasoning-Based Entity Prediction ranks candidates from the post-CoT hidden state. Answer-Free MKG CoT Data provides reasoning supervision without exposing the gold answer, while SFT and Answer-Aware On-Policy Distillation (OPD) strengthen the student's inference-time reasoning. MKGC-CoT achieves the best results on DB15K and MKG-Y, including a **26.8%** relative MRR improvement on DB15K. Ablations confirm the contribution of reasoning, ranking, SFT, and OPD. Future work will extend this framework to additional modalities and larger open-domain graphs.

AI USE STATEMENT

We used AI tools for translation, language polishing, and grammar correction. The authors reviewed and take responsibility for all content.

REFERENCES

Rishabh Agarwal, Nino Vieillard, Yongchao Zhou, Piotr Stanczyk, Sabela Ramos Garea, Matthieu Geist, and Olivier Bachem. On-policy distillation of language models: Learning from self-generated mistakes. In *International Conference on Learning Representations*, volume 2024, pp. 21246–21263, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/5be69a584901a26c521c2b51e40a4c20-Abstract-Conference.html.

Ivana Balažević, Carl Allen, and Timothy Hospedales. TuckER: Tensor factorization for knowledge graph completion. In *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP)*, pp. 5185–5194, 2019.

Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. Translating embeddings for modeling multi-relational data. *Advances in neural information processing systems*, 26, 2013.

Zongsheng Cao, Qianqian Xu, Zhiyong Yang, Yuan He, Xiaochun Cao, and Qingming Huang. OTKGE: Multi-modal knowledge graph embeddings via optimal transport. *Advances in neural information processing systems*, 35:39090–39102, 2022.

Zhuo Chen, Yichi Zhang, Yin Fang, Yuxia Geng, Lingbing Guo, Xiang Chen, Qian Li, Wen Zhang, Jiaoyan Chen, Yushan Zhu, et al. Knowledge graphs meet multi-modal learning: A comprehensive survey. *arXiv preprint arXiv:2402.05391*, 2024.

Yuxiao Gao, Fuwei Zhang, Zhao Zhang, Xiaoshuang Min, and Fuzhen Zhuang. Mixed-curvature multi-modal knowledge graph completion. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 11699–11707, 2025.

Yuan Guo, Qian Ma, Hui Li, Qiao Ning, Furui Zhan, Yu Gu, Ge Yu, and Shikai Guo. LBMKGC: Large model-driven balanced multimodal knowledge graph completion. In *Advances in Neural Information Processing Systems*, volume 38, 2025. doi: 10.52202/085713-3736. URL https://papers.neurips.cc/paper_files/paper/2025/hash/a254abbfdd029d388fc35fd850bbc551-Abstract-Conference.html.

Wei Huang, Peining Li, Meiyu Liang, Xu Hou, Junping Du, Yingxia Shao, Guanhua Ye, Wu Liu, Kangkang Lu, and Yang Yu. ELMM: Efficient lightweight multimodal large language models for multimodal knowledge graph completion. *arXiv preprint arXiv:2510.16753*, 2025.

Wei Huang, Peining Li, Meiyu Liang, Xu Hou, Yawen Li, Junping Du, Zhe Xue, and Zeli Guan. DiffusionCom: Structure-aware multimodal diffusion model for multimodal knowledge graph completion. In *ICASSP 2026-2026 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 12452–12456. IEEE, 2026.

Dengyang Jiang, Xin Jin, Dongyang Liu, Zanyi Wang, Mingzhe Zheng, Ruoyi Du, Xiangpeng Yang, Qilong Wu, Zhen Li, Peng Gao, et al. D-OPSD: On-policy self-distillation for continuously tuning step-distilled diffusion models. *arXiv preprint arXiv:2605.05204*, 2026.

Jaejun Lee, Chanyoung Chung, Hochang Lee, Sungho Jo, and Joyce Whang. VISTA: Visual-textual knowledge graph representation learning. In *Findings of the association for computational linguistics: EMNLP 2023*, pp. 7314–7328, 2023.

Lijian Li, Yuanpeng He, Yandong Yan, Linyu Li, Wenpin Jiao, Zhi Jin, and Chi-Man Pun. Complementarity-driven representation learning for multi-modal knowledge graph completion. *Pattern Recognition*, pp. 113327, 2026.

Xinhang Li, Xiangyu Zhao, Jiaxing Xu, Yong Zhang, and Chunxiao Xing. IMF: interactive multimodal fusion model for link prediction. In *Proceedings of the ACM web conference 2023*, pp. 2572–2580, 2023.

Haiyan Liu, Xin Song, Ye Wang, Kai Chen, Yuying Liu, Bin Zhou, Hongkui Tu, Liqun Gao, and Yue Qian. Robust multi-modal knowledge graph completion via modality-specific experts. In *Proceedings of the 2026 International Conference on Multimedia Retrieval*, pp. 728–737. ACM, 2026a. doi: 10.1145/3805622.3810765. URL https://doi.org/10.1145/3805622.3810765.

Lihui Liu, Zihao Wang, and Hanghang Tong. Neural-symbolic reasoning over knowledge graphs: A survey from a query perspective. *ACM SIGKDD Explorations Newsletter*, 27(1):124–136, 2025.

Zhiqiang Liu, Yichi Zhang, Mengshu Sun, Lei Liang, and Wen Zhang. Collaboration of fusion and independence: Hypercomplex-driven robust multi-modal knowledge graph completion. In *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 27968–27983. Association for Computational Linguistics, 2026b. doi: 10.18653/v1/2026.acl-long.1289. URL https://aclanthology.org/2026.acl-long.1289/.

Xinyu Lu, Lifang Wang, Zejun Jiang, Shichang He, and Shizhong Liu. MMKRL: A robust embedding approach for multi-modal knowledge graph representation learning. *Applied Intelligence*, 52 (7):7480–7497, 2022.

Linhao Luo, Yuan-Fang Li, Reza Haffari, and Shirui Pan. Reasoning on graphs: Faithful and interpretable large language model reasoning. In *International Conference on Learning Representations*, volume 2024, pp. 14400–14423, 2024.

Haodi Ma, Dzmitry Kasinets, and Daisy Zhe Wang. Transformer-based multimodal knowledge graph completion with link-aware contexts. *arXiv preprint arXiv:2501.15688*, 2025.

Debjyoti Mondal, Suraj Modi, Subhadarshi Panda, Rituraj Singh, and Godawari Sudhakar Rao. KAM-CoT: Knowledge augmented multimodal chain-of-thoughts reasoning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, pp. 18798–18806, 2024.

Hatem Mousselly-Sergieh, Teresa Botschen, Iryna Gurevych, and Stefan Roth. A multimodal translation-based approach for knowledge graph representation learning. In *Proceedings of the seventh joint conference on lexical and computational semantics*, pp. 225–234, 2018.

Guanglin Niu and Xiaowei Zhang. Diffusion-based hierarchical negative sampling for multimodal knowledge graph completion. In *International Conference on Database Systems for Advanced Applications*, pp. 479–495. Springer, 2025.

Jiashun Peng, Fu Zhang, Hongzhi Chen, Jingwei Cheng, Yingsong Ning, and Xiaoke Wang. Not all modalities at once: Dynamic dropout and bidirectional fusion for robust multi-modal knowledge graph completion. In *Findings of the Association for Computational Linguistics: ACL 2026*, pp. 17928–17940. Association for Computational Linguistics, 2026. doi: 10.18653/v1/2026.findings-acl.890. URL https://aclanthology.org/2026.findings-acl.890/.

Bin Shang, Yinliang Zhao, Jun Liu, and Di Wang. LAFA: Multimodal knowledge graph completion with link aware fusion and aggregation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 8957–8965, 2024. doi: 10.1609/aaai.v38i8.28744. URL https://ojs.aaai.org/index.php/AAAI/article/view/28744.

Jiashuo Sun, Chengjin Xu, Lumingyuan Tang, Saizhuo Wang, Chen Lin, Yeyun Gong, Lionel Ni, Heung-Yeung Shum, and Jian Guo. Think-on-Graph: Deep and responsible reasoning of large language model on knowledge graph. In *The Twelfth International Conference on Learning Representations*, 2024.

Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, and Jian Tang. RotatE: Knowledge graph embedding by relational rotation in complex space. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net, 2019. URL https://openreview.net/forum?id=HkgEQnRqYQ.

Yun Tang, Jing Huang, Guangtao Wang, Xiaodong He, and Bowen Zhou. Orthogonal relation transforms with graph context modeling for knowledge graph embedding. In *Proceedings of the 58th annual meeting of the association for computational linguistics*, pp. 2713–2722, 2020.

Théo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, and Guillaume Bouchard. Complex embeddings for simple link prediction. In *International conference on machine learning*, pp. 2071–2080. PMLR, 2016.

Di Wang, Junping Du, Zhe Xue, Meiyu Liang, Guanhua Ye, Yingxia Shao, and Haisheng Li. HFR-MKGC: Hierarchical fusion reasoning with MLLMs for multi-modal knowledge graph completion. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 15815–15823, 2026. doi: 10.1609/aaai.v40i18.38613. URL https://ojs.aaai.org/index.php/AAAI/article/view/38613.

Meng Wang, Sen Wang, Han Yang, Zheng Zhang, Xi Chen, and Guilin Qi. Is visual context really helpful for knowledge graph? a representation learning perspective. In *Proceedings of the 29th ACM international conference on multimedia*, pp. 2735–2743, 2021.

Zikang Wang, Linjing Li, Qiudan Li, and Daniel Zeng. Multimodal data enhanced representation learning for knowledge graphs. In *2019 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8. IEEE, 2019.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed H. Chi, Quoc V. Le, and Denny Zhou. Chain-of-thought prompting elicits reasoning in large language models. In *Advances in Neural Information Processing Systems*, volume 35, pp. 24824–24837, 2022. URL https://papers.nips.cc/paper_files/paper/2022/file/9d5609613524ecf4f15af0f7b31abca4-Paper-Conference.pdf.

Yanbin Wei, Qiushi Huang, Yu Zhang, and James Kwok. KICGPT: Large language model with knowledge in context for knowledge graph completion. In *Findings of the association for computational linguistics: EMNLP 2023*, pp. 8667–8683, 2023.

Ruobing Xie, Zhiyuan Liu, Huanbo Luan, and Maosong Sun. Image-embodied knowledge representation learning. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, pp. 3140–3146. International Joint Conferences on Artificial Intelligence Organization, 2017.

Derong Xu, Tong Xu, Shiwei Wu, Jingbo Zhou, and Enhong Chen. Relation-enhanced negative sampling for multimodal knowledge graph completion. In *Proceedings of the 30th ACM international conference on multimedia*, pp. 3857–3866, 2022.

Bishan Yang, Scott Wen-tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. Embedding entities and relations for learning and inference in knowledge bases. In *Proceedings of the International Conference on Learning Representations (ICLR) 2015*, 2015.

Yichi Zhang and Wen Zhang. Knowledge graph completion with pre-trained multimodal transformer and twins negative sampling. *arXiv preprint arXiv:2209.07084*, 2022.

Yichi Zhang, Zhuo Chen, Lingbing Guo, Yajing Xu, Binbin Hu, Ziqi Liu, Wen Zhang, and Huajun Chen. Mixture of modality knowledge experts for robust multi-modal knowledge graph completion. *CoRR, abs/2405.16869*, 2024a.

Yichi Zhang, Zhuo Chen, Lingbing Guo, Yajing Xu, Binbin Hu, Ziqi Liu, Wen Zhang, and Huajun Chen. NativE: Multi-modal knowledge graph completion in the wild. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 91–101. ACM, 2024b. doi: 10.1145/3626772.3657800. URL https://arxiv.org/abs/2406.17605.

Yichi Zhang, Zhuo Chen, Lei Liang, Huajun Chen, and Wen Zhang. Unleashing the power of imbalanced modality information for multi-modal knowledge graph completion. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pp. 17120–17130, 2024c.

Yichi Zhang, Zhuo Chen, Lingbing Guo, Yajing Xu, Binbin Hu, Ziqi Liu, Wen Zhang, and Huajun Chen. Tokenization, fusion, and augmentation: Towards fine-grained multi-modal entity representation. In *AAAI*, pp. 13322–13330. AAAI Press, 2025.

Yu Zhao, Ying Zhang, Xuhui Sui, Baohang Zhou, Haoze Zhu, Jeff Z. Pan, and Xiaojie Yuan. Dark side of modalities: Reinforced multimodal distillation for multimodal knowledge graph reasoning. In *Proceedings of the 33rd ACM International Conference on Multimedia*. ACM, 2025. doi: 10.1145/3746027.3754576. URL https://arxiv.org/abs/2507.20738.

A Supplementary Method and Experimental Details

This appendix expands the prediction head, data construction, and two-stage training used by MKGC-CoT, followed by baseline configurations and the provenance of the qualitative examples. Definitions and references are self-contained so that the appendix can be read separately from the main text.

A.1 Prediction Head and Ranking Supervision

Let $\mathcal{E}$ be the entity set, $N = |\mathcal{E}|$, and $\mathbf{z}_e \in \mathbb{R}^d$ the fused structural, visual, and textual representation of candidate e. The query entity's fused representation is projected to the LLM hidden size and prepended to its instruction. The response contains four ordered fields: entity type, relation meaning, evidence, and answer. Let $\mathbf{h}_{\text{eoc}} \in \mathbb{R}^H$ be the last-layer hidden state at the end of this CoT. The prediction head reads this state and ranks every entity:

$$\mathbf{q} = \mathbf{W}_q \mathbf{h}_{\text{eoc}}, \qquad \mathbf{W}_q \in \mathbb{R}^{d \times H},$$
$$\phi(e \mid h, r) = \tau_r \, g_r(\mathbf{q})^\top \mathbf{z}_e + b_e. \tag{8}$$

Here $g_r : \mathbb{R}^d \to \mathbb{R}^d$ is a relation-conditioned transform, τ_r scales the interaction for relation r, and b_e is an entity bias. The dependence on the response enters through $\mathbf{h}_{\text{eoc}}$. This common query vector is compared with every candidate embedding, so ranking does not require matching the generated Answer string to an entity name. Head prediction uses the inverse query $(t, r^{-1}, ?)$ with the same scoring procedure.

Multi-positive ranking. Let $\mathcal{T}_{\text{train}}$ contain the training triples, including their inverse representation for head queries, and define $P(h, r) = \{e : (h, r, e) \in \mathcal{T}_{\text{train}}\}$. The target distributes probability uniformly over the valid training tails of the current query:

$$\widetilde{y}_e = \frac{\mathbf{1}[e \in P(h,r)]}{|P(h,r)|}, \qquad \mathcal{L}_{\text{score}} = -\sum_{e \in \mathcal{E}} \widetilde{y}_e \log \frac{\exp \phi(e \mid h, r)}{\sum_{e' \in \mathcal{E}} \exp \phi(e' \mid h, r)}. \tag{9}$$

The positive set is obtained from training triples rather than from the teacher's Answer field. Thus, separate traces for the same query can name different valid answers while sharing the same ranking target. If p_e denotes the softmax probability of candidate e, then $\partial \mathcal{L}_{\text{score}} / \partial \phi(e \mid h, r) = p_e - \widetilde{y}_e$; all valid training answers receive positive target mass instead of being treated as mutually exclusive labels. Validation and test answers do not enter this target. They are used for filtered evaluation, which removes other known correct answers from a query's candidate ranking while retaining the answer being evaluated.

A.2 Data Construction and Implementation Settings

DB15K has 12,842 entities, 279 relations, and 79,222/9,904/9,902 train/validation/test triples in the local split files. MKG-Y has 15,000 entities, 28 relations, and 21,310/2,665/2,663 triples. The CoT corpora contain 116,960 and 42,620 directional records, respectively. Before filtering and deduplication, DB15K has 158,444 directional training records from the two prediction directions of each training triple; this raw count differs from the exported trace count.

Each training triple (h, r, t) produces tail and head queries, $(h, r, ?)$ and $(?, r, t)$. For each query, the instruction verbalizes at most eight one-hop neighbors from the training split using directed edges and readable relation labels. The gold edge (h, r, t) is removed from the context in both query directions. Other training edges may provide supporting context, including edges involving another valid answer. Training positives for ranking are collected separately from the unfiltered training triples, so removing an edge from the instruction does not remove its entity from the positive target in Equation 9.

The teacher view appends the canonical gold answer, a one-shot example, and a four-line format requirement to the query instruction. A frozen teacher produces the fields `Entity type`, `Relation means`, `Evidence`, and `Answer`. The student view stores the instruction without the appended answer. Answer-free therefore describes the student's query input; its supervised response still contains the Answer field, and its ranking label is obtained from the training graph.

We retain records only when both query directions are completed. Text before `Entity type:` is removed, followed by checks of the four-field format and whether the Answer field names the gold entity. The reported format-pass and answer-agreement rates are 99.99% and 97.0%, respectively. Completed records are deduplicated and exported as instruction–trace–entity tuples. These format and agreement checks describe corpus construction; they do not by themselves establish the faithfulness of every intermediate reasoning statement.

The student is Qwen3.5-4B, and the frozen teacher used for trace generation and distillation is Qwen3.5-35B-A3B. The experimental configuration uses rank-16 LoRA, a 256-dimensional fusion space, and AdamW. Visual and textual encoders remain frozen, while structural embeddings, fusion, and the prediction head support the shared reasoning and ranking objectives. The ranking-loss weights are $\alpha = 1$ in SFT and $\beta = 1$ in OPD. Checkpoints are selected by validation MRR; final evaluation ranks all entities and averages filtered metrics over head and tail prediction.

A.3 Teacher Conditioning and Distillation

We first jointly learn CoT generation and entity ranking by supervised fine-tuning (SFT), then apply answer-aware on-policy distillation (OPD) on student-generated traces. Let q denote an answer-free query for $(h, r, ?)$, and let y^* be its gold entity from the training split. Head prediction uses the inverse query $(t, r^{-1}, ?)$. The gold entity provides supervision during training; it is not appended to the student query. A generated or supervised trace may itself contain an Answer field, which is distinct from supplying the gold in the input.

Ranking after the CoT. For a CoT sequence $\mathbf{u}_{1:L}$, let $\mathbf{h}_{\text{eoc}}$ be the final-layer hidden state at its end. The prediction head defined in Equation 8 reads this state and scores every entity $e \in \mathcal{E}$, including entities not mentioned in the trace. Its multi-positive loss $\mathcal{L}_{\text{score}}$ is defined in Equation 9, with target mass distributed uniformly over the query's training answers.

Stage 1: joint supervised fine-tuning. Given a teacher-written target trace $\mathbf{c}_{1:L}$, the language modeling loss uses teacher forcing:

$$\mathcal{L}_{\text{LM}} = -\frac{1}{L}\sum_{i=1}^{L} \log p_\theta(c_i \mid q, \mathbf{c}_{<i}), \qquad \mathcal{L}_{\text{SFT}} = \mathcal{L}_{\text{LM}} + \alpha\mathcal{L}_{\text{score}}. \tag{10}$$

The ranking state is taken after this supervised CoT. The coefficient α balances token prediction and entity ranking.

Stage 2: answer-aware on-policy distillation. Starting from the SFT student, we draw a CoT $\widehat{\mathbf{c}}_{1:L} \sim \pi_\theta(\cdot \mid q)$. At each position, the frozen teacher and the student evaluate the same student-generated prefix, with the gold entity supplied only to the teacher (Figure 6):

$$\begin{aligned} p_{\text{T}}^{(i)} &= p_{\text{teacher}}(\cdot \mid q, y^*, \widehat{\mathbf{c}}_{<i}), \\ p_{\text{S}}^{(i)} &= p_\theta(\cdot \mid q, \widehat{\mathbf{c}}_{<i}). \end{aligned} \tag{11}$$

For the token vocabulary $\mathcal{V}$, define the forward-KL term

$$\mathcal{L}_{\text{tok}}(q, \widehat{\mathbf{c}}, y^*) = \frac{1}{L}\sum_{i=1}^{L}\sum_{v\in\mathcal{V}} p_{\text{T}}^{(i)}(v) \log \frac{p_{\text{T}}^{(i)}(v)}{p_{\text{S}}^{(i)}(v)}. \tag{12}$$

The complete second-stage objective is

$$\mathcal{L}_{\text{OPD}} = \mathbb{E}_{\widehat{\mathbf{c}}_{1:L}\sim\pi_\theta(\cdot|q)}\left[\mathcal{L}_{\text{tok}}(q, \widehat{\mathbf{c}}, y^*)\right] + \beta\mathcal{L}_{\text{score}}. \tag{13}$$

Here the ranking term is evaluated after the sampled student CoT, and β retains training-entity supervision during distillation. Teacher distributions are detached, and the sampled token identities are treated as fixed when differentiating the update. Gradients flow through the student token probabilities and its trainable ranking path, not through the teacher or the discrete sampling operation. This exposes training to prefixes produced by the student rather than only to teacher-written prefixes.

1. Student rollout
Answer-free query $q \to \widehat{\mathbf{c}}_{1:L} \sim \pi_\theta(\cdot \mid q)$.

2. Answer-aware teacher guidance
Frozen teacher: $q + y^* + \widehat{\mathbf{c}}_{<i} \to p_{\mathrm{T}}^{(i)}$.
Student: $q + \widehat{\mathbf{c}}_{<i} \to p_{\mathrm{S}}^{(i)}$.

3. Joint student update
Forward token KL on the shared prefixes, plus entity-ranking supervision from the state after $\widehat{\mathbf{c}}_{1:L}$. Only the student is updated.

Figure 6: Answer-aware on-policy distillation. The teacher receives the training gold and evaluates the same prefixes as the student. The student learns from token distributions and post-CoT entity-ranking supervision; neither a teacher nor a supplied gold answer is needed at inference.

Loss notation and inference. In the main architecture diagram, $\mathcal{L}_{\mathrm{NTP}}$ denotes $\mathcal{L}_{\mathrm{LM}}$, $\mathcal{L}_{\mathrm{KGC}}$ denotes $\mathcal{L}_{\mathrm{score}}$, and the diagram's $\mathcal{L}_{\mathrm{OPD}}$ label denotes only the token-distillation term in Equation 12. In contrast, Equation 13 defines $\mathcal{L}_{\mathrm{OPD}}$ as the complete second-stage objective, including ranking supervision. At inference, the student generates its own CoT from the answer-free query and ranks all entities using the post-CoT state; no teacher or supplied gold answer is used.

A.4 Baseline Configurations

The original main-table baselines are TransE, DistMult, ComplEx, RotatE, GC-OTE, and TuckER (Bordes et al., 2013; Yang et al., 2015; Trouillon et al., 2016; Sun et al., 2019; Tang et al., 2020; Balažević et al., 2019); and IKRL, TBKGC, TransAE, MMKRL, RSME, VBKGC, OTKGE, IMF, AdaMF-MAT, VISTA, MyGO, MoMoK, and MCKGC (Xie et al., 2017; Mousselly-Sergieh et al., 2018; Wang et al., 2019; Lu et al., 2022; Wang et al., 2021; Zhang & Zhang, 2022; Cao et al., 2022; Li et al., 2023; Zhang et al., 2024c; Lee et al., 2023; Zhang et al., 2025; 2024a; Gao et al., 2025). The added NativE row uses its S+I+T configuration from Table 2 (Zhang et al., 2024b), not the richer-modality "All" variant; DB15K Hits@3 and MKG-Y Hits@3/10 remain unreported. The LBMKGC, DSoM, HFR-MKGC, MDBGF, MEIF, and M-Hyper rows use each method's own Table 1 (Guo et al., 2025; Zhao et al., 2025; Wang et al., 2026; Peng et al., 2026; Liu et al., 2026a;b). Scores are converted to percentages where necessary and reordered as MRR/Hits@1/Hits@3/Hits@10; no other paper's reproduction replaces an original baseline entry.

Identical dataset names do not establish identical protocols. For example, DSoM, M-Hyper, MDBGF, and HFR-MKGC list DB15K validation/test counts of 9,902/9,904, whereas our loaded split statistics are 9,904/9,902 (Section A.2); file-level alignment remains necessary. LBMKGC augments missing images with SDXL, and MDBGF uses multiple image/token inputs, unlike a fixed common-feature comparison. Filtering details, text coverage, and feature encoders also require alignment. A matched rerun must fix splits, features, entity candidates, tie/filter rules, training budgets, and validation selection. The local adapter is named `local_RelMoE_ol`, not a certified full official MoMoK reproduction. Importing rank files does not itself establish a reproduction.

LAFA's original evaluation uses FB15k-237-IMG and WN18-IMG, with MR and Hits@K (Shang et al., 2024); neither MR nor scores on these datasets can substitute for DB15K/MKG-Y MRR. It is therefore discussed in related work but not included in the quantitative comparison on these two datasets.

A.5 Scope of the Parameter Sensitivity Results

The parameter-sensitivity study evaluates DB15K ranking accuracy as a function of LoRA rank and the entity-ranking loss weights in SFT and OPD. All values are final MRR after SFT followed by OPD; the α sweep changes the SFT weight before completing both stages. The default settings (LoRA rank 16 and $\alpha = \beta = 1$) each yield 50.44% MRR. The fourth points are 50.32% at rank 32, 45.91% at $\alpha = 2$, and 48.55% at $\beta = 2$, respectively 0.12, 4.53, and 1.89 percentage points below the defaults. Each default is the highest reported point within its tested sweep. These measurements concern the evaluated configurations; they do not establish robustness to missing modalities, reasoning faithfulness, or variability across training seeds.

A.6 Qualitative Evidence and Response Diagnostics

The four main-text cases use MKG-Y test triples at rows 60, 1679, 1798, and 392, respectively; all four triples are absent from training. For Jonathan Bottinelli's citizenship, two training edges connect him only to Adana Demirspor; affiliation does not establish the test answer, Argentina. Discovery Channel has no incident training edges, and Maryland is its dataset location target, not a claim about its current headquarters. University of Ceylon has one incident training edge, to Colombo; a second training edge connects Colombo to Sri Lanka. Gold-edge removal leaves the first edge eligible for the university's one-hop context, but does not guarantee that the second-hop edge is supplied. The full path's presence in the training graph therefore does not establish its presence in the actual instruction or its use by the model. Pavel Ruminov has exactly three incident training edges, all to *Dead Daughters*, for creation, direction, and music; none states the held-out editing relation. All four panels are curated data-grounded schematics, not verified generated traces, top-ranked predictions, or evidence of visual grounding. Their icons are decorative, not entity images recovered from the model's input features.