SCMRAG 2.0: Efficient and Scalable Multi-hop Graph RAG with Multimodal Knowledge-Graphs and Agentic Self-Correction

Extended Abstract

Rishabh Agrawal
University of Western Ontario
London, Canada
ragrawa9@uwo.ca

Uday Devulapalli
University of Western Ontario
London, Canada
udevulap@uwo.ca

Apurva Narayan
University of Western Ontario
London, Canada
apurva.narayan@uwo.ca

ABSTRACT

We present SCMRAG 2.0, a next-generation retrieval-augmented generation framework that unifies text, image, and structured data into a Multimodal Knowledge Graph. Unlike traditional graph RAG systems, SCMRAG 2.0 introduces dual linkages via symbolic relations and cross-modal embeddings, an optimized graph-retrieval algorithm, and a multimodal agentic self-correction loop. By aligning language-level structure with vector-space signals and enabling agentic critique and repair, SCMRAG 2.0 mitigates outdated context, incomplete reasoning chains, and hallucinations common in text-only graph RAG systems. Experiments on MMLU and MRAG-Bench demonstrate that SCMRAG 2.0 significantly outperforms strong baselines like LightRAG in retrieval precision and factuality while maintaining computational efficiency.

KEYWORDS

Retrieval-Augmented Generation; Multimodal Knowledge Graph; Agentic Self-Correction; Graph Retrieval; Multi-hop Reasoning

ACM Reference Format:
Rishabh Agrawal, Uday Devulapalli, and Apurva Narayan. 2026. SCMRAG 2.0: Efficient and Scalable Multi-hop Graph RAG with Multimodal Knowledge-Graphs and Agentic Self-Correction: Extended Abstract. In *Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), Paphos, Cyprus, May 25 – 29, 2026*, IFAAMAS, 3 pages. https://doi.org/10.65109/GGJL7344

1 INTRODUCTION

Large Language Models (LLMs) and multimodal foundation models have transformed generative tasks but remain prone to hallucinations and weak grounding when reasoning over dynamic or complex contexts [1]. While Retrieval-Augmented Generation (RAG) addresses this by fetching external evidence, most systems rely on flat, text-centric vector stores [2, 11]. These approaches fail to capture the compositional and cross-modal relationships essential for robust reasoning—for instance, linking a visual chart in one document to a textual summary in another.

Existing graph-based RAG extensions attempt to introduce structure but often produce surface-level symbolic links without semantic depth, or suffer from scalability issues due to heavy post-hoc optimization [5, 6, 8, 12]. To address these gaps, we propose **SCMRAG 2.0**, a next-generation framework that rethinks both representation and retrieval. Our contributions are: **(i) Multimodal Knowledge Graph (MMKG):** We introduce a novel graph representation that integrates text, image, and structured data into claim-anchored nodes, enabling coherent reasoning and retrieval across heterogeneous modalities. **(ii) Dual-link graph connectivity:** Nodes are connected via both *symbolic relations* (explicit entities) and *embedding edges* (latent cross-modal similarity), capturing both logical dependencies and vector-space correspondences. **(iii) Optimized multimodal graph retrieval algorithm:** We design an efficient retrieval and pruning mechanism that jointly explores graph structure and embedding proximity, achieving significant gains in retrieval performance. **(iv) Agentic multimodal self-correction:** A reasoning loop that audits retrieved evidence for grounding gaps and issues modality-aware follow-up queries.

This work is licensed under a Creative Commons Attribution International 4.0 License.

Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), C. Amato, L. Dennis, V. Mascardi, J. Thangarajah (eds.), May 25 – 29, 2026, Paphos, Cyprus. © 2026 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). https://doi.org/10.65109/GGJL7344

2 PROPOSED METHOD

The SCMRAG 2.0 pipeline (Figure 1) consists of three stages: Evidence Processing, Graph Construction, and Multihop Retrieval via a Self-Corrective Agent.

2.1 Multimodal Knowledge Graph Construction

Unlike standard graphs that treat images as isolated nodes, SCMRAG 2.0 processes raw inputs (documents, images, tables) into *claim tuples*: $E_i = (claim_i, target_i, topic_i, z_i)$, where z_i represents a joint embedding vector of the claim, target entity, and topic. The graph construction utilizes a **dual-linkage mechanism**:

i) **Symbolic Edges:** Established via shared semantic topic, targets or explicit references to extracted metadata information.
ii) **Embedding Edges:** Established via high cosine similarity between node embedding vector tuples (v_{target}, v_{topic}) exceeding relevant threshold hyper-parameters $(\tau_{target}, \tau_{topic})$.

This dual structure allows the retriever to perform multi-hop traversal across modalities, jumping from a textual claim to a visually similar image node even if no explicit symbolic link exists.

2.2 Agentic Self-Corrective Retrieval

To mitigate the retrieval of outdated or incomplete context, SCMRAG 2.0 employs a Self-Correction Agent that integrates reasoning with multi-hop traversal. This module operationalizes a dynamic retrieval-correction cycle defined by the following process: **(i)** The system encodes the query q and retrieves the top-k nodes from the MMKG to generate a preliminary answer A_0. **(ii)** A verification step calculates consistency scores (s_t) for alignment between the current answer and the retrieved evidence. **(iii)** If s_t falls below a threshold

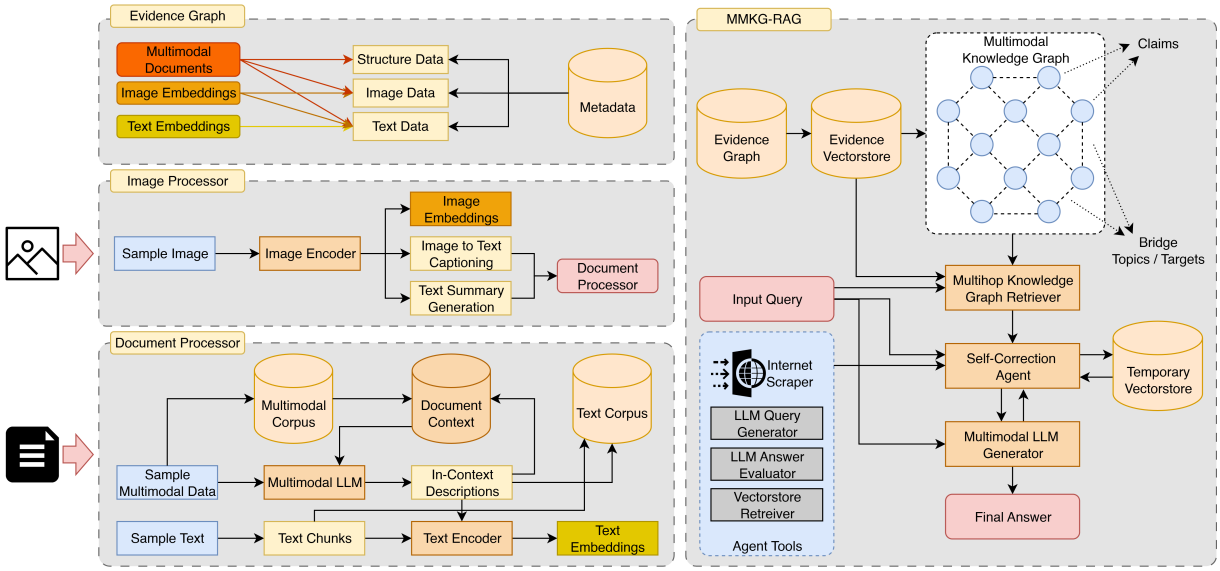

Figure 1: Overview of the SCMRAG 2.0 framework. Multimodal inputs are processed into an Evidence Graph and unified into a dual-linked MMKG. A Self-Correction Agent performs multi-hop retrieval and refinement to generate the final answer.

Table 1: Performance comparison of SCMRAG 2.0 against baseline methods on MMLU and MRAG-Bench datasets.

Method	Acc %	Prec (w)	Rec (w)	F1 (w)
MMLU				
LightRAG	75.40	0.774	0.754	0.764
SCMRAG 2.0 (Ours)	**86.00**	**0.869**	**0.860**	**0.864**
Improvement	*+10.60*	*+0.095*	*+0.106*	*+0.100*
MRAG				
RAG-Anything	45.52	0.481	0.455	0.451
SCMRAG 2.0 (Ours)	**60.15**	**0.608**	**0.601**	**0.603**
Improvement	*+14.63*	*+0.127*	*+0.146*	*+0.152*

δ, the agent identifies specific modality gaps, refines the query to q_t, and expands the evidence set $\mathcal{E}_q$ via targeted graph traversal and/or using external knowledge sources. **(iv)** The cycle repeats up to T_{max} iterations or until sufficient grounding is achieved, producing the final grounded answer A_t.

3 EVALUATION AND FINDINGS

We evaluated SCMRAG 2.0 on two very large retrieval and generation benchmarks: **MMLU** (Massive Multitask Language Understanding) [9] for text-based reasoning and **MRAG-Bench** [10] for multimodal retrieval. We compared our method against **LightRAG** [5], a graph-enhanced RAG baseline, and its state-of-the-art multimodal implementation **RAG-Anything** [4]. We utilized *Qwen2.5-VL-32B-Instruct* [3] as the generator and tool-reasoning backbone, and *jina-embeddings-v4* [7] for multimodal encoding.

SCMRAG 2.0 demonstrates consistent and substantial improvements across all evaluated metrics. As shown in Table 1, our method achieves an aggregate accuracy of **86.00%** in the MMLU benchmark, outperforming LightRAG by **+10.60%**. The most significant performance gains were observed in mathematically intensive and reasoning-heavy domains. SCMRAG 2.0 outperformed the baseline in *College Mathematics* by **+29.0%** and *College Physics* by **+22.1%**.

On MRAG-Bench dataset, SCMRAG 2.0 achieved a **+14.63%** increase in accuracy. This confirms that our dual-linked MMKG effectively bridges the semantic gap between textual and visual modalities better than state-of-the-art RAG-anything baseline. Our analysis reveals the reason behind this is the distinct roles for each component: **(i) Agentic Self-Correction** excels in multi-step symbolic reasoning, recovering from initial retrieval misses in abstract domains.**(ii) The Dual-Linked MMKG** is critical for taxonomy-dense and cross-referential domains, where aligning symbolic relations with embedding edges helps consolidate dispersed facts.

To further understand the source of these improvements, we ablated the system into *Retrieval Only* (MMKG traversal without agent), *Self-Correction Only* (agent without MMKG), and the *Full Pipeline* on the MMLU dataset. The Full Pipeline consistently outperformed individual modules, yielding a **+3.58%** accuracy gain over Retrieval Only and **+2.82%** over Self-Correction Only. This confirms that the dual-linked graph and the agentic loop provide complementary benefits: the graph improves coverage, while the agent actively repairs retrieval failures.

4 CONCLUSION

SCMRAG 2.0 addresses the limitations of graph-based, unimodal RAG systems by integrating a novel dual-linked Multimodal Knowledge Graph with agentic self-correction. Our results demonstrate that aligning symbolic structure with dense cross-modal embeddings, combined with iterative critique, yields state-of-the-art performance in knowledge-intensive and multimodal tasks.

REFERENCES

[1] Mohammad Mahdi Abootorabi, Amirhosein Zobeiri, Mahdi Dehghani, Mohammadali Mohammadkhani, Bardia Mohammadi, Omid Ghahroodi, Mahdieh Soleymani Baghshah, and Ehsaneddin Asgari. 2025. Ask in Any Modality: A Comprehensive Survey on Multimodal Retrieval-Augmented Generation. arXiv:2502.08826 [cs.CL] https://arxiv.org/abs/2502.08826

[2] Akari Asai, Zeqiu Wu, Yizhong Wang, Avirup Sil, and Hannaneh Hajishirzi. 2023. Self-rag: Learning to retrieve, generate, and critique through self-reflection. *arXiv preprint arXiv:2310.11511* (2023).

[3] Shuai Bai, Keqin Chen, Xuejing Liu, Jialin Wang, Wenbin Ge, Sibo Song, Kai Dang, Peng Wang, Shijie Wang, Jun Tang, Humen Zhong, Yuanzhi Zhu, Mingkun Yang, Zhaohai Li, Jianqiang Wan, Pengfei Wang, Wei Ding, Zheren Fu, Yiheng Xu, Jiabo Ye, Xi Zhang, Tianbao Xie, Zesen Cheng, Hang Zhang, Zhibo Yang, Haiyang Xu, and Junyang Lin. 2025. Qwen2.5-VL Technical Report. arXiv:2502.13923 [cs.CV] https://arxiv.org/abs/2502.13923

[4] Zirui Guo, Xubin Ren, Lingrui Xu, Jiahao Zhang, and Chao Huang. 2025. RAG-Anything: All-in-One RAG Framework. arXiv:2510.12323 [cs.AI] https://arxiv.org/abs/2510.12323

[5] Zirui Guo, Lianghao Xia, Yanhua Yu, Tu Ao, and Chao Huang. 2025. LightRAG: Simple and Fast Retrieval-Augmented Generation. arXiv:2410.05779 [cs.IR] https://arxiv.org/abs/2410.05779

[6] Bernal Jiménez Gutiérrez, Yiheng Shu, Yu Gu, Michihiro Yasunaga, and Yu Su. 2025. HippoRAG: Neurobiologically Inspired Long-Term Memory for Large Language Models. arXiv:2405.14831 [cs.CL] https://arxiv.org/abs/2405.14831

[7] Michael Günther, Saba Sturua, Mohammad Kalim Akram, Isabelle Mohr, Andrei Ungureanu, Bo Wang, Sedigheh Eslami, Scott Martens, Maximilian Werk, Nan Wang, and Han Xiao. 2025. jina-embeddings-v4: Universal Embeddings for Multimodal Multilingual Retrieval. arXiv:2506.18902 [cs.AI] https://arxiv.org/abs/2506.18902

[8] Haoyu Han, Yu Wang, Harry Shomer, Kai Guo, Jiayuan Ding, Yongjia Lei, Mahantesh Halappanavar, Ryan A. Rossi, Subhabrata Mukherjee, Xianfeng Tang, Qi He, Zhigang Hua, Bo Long, Tong Zhao, Neil Shah, Amin Javari, Yinglong Xia, and Jiliang Tang. 2025. Retrieval-Augmented Generation with Graphs (GraphRAG). arXiv:2501.00309 [cs.IR] https://arxiv.org/abs/2501.00309

[9] Dan Hendrycks, Collin Burns, Steven Basart, Andy Zou, Mantas Mazeika, Dawn Song, and Jacob Steinhardt. 2021. Measuring Massive Multitask Language Understanding. arXiv:2009.03300 [cs.CY] https://arxiv.org/abs/2009.03300

[10] Wenbo Hu, Jia-Chen Gu, Zi-Yi Dou, Mohsen Fayyaz, Pan Lu, Kai-Wei Chang, and Nanyun Peng. 2025. MRAG-Bench: Vision-Centric Evaluation for Retrieval-Augmented Multimodal Models. arXiv:2410.08182 [cs.CV] https://arxiv.org/abs/2410.08182

[11] Shi-Qi Yan, Jia-Chen Gu, Yun Zhu, and Zhen-Hua Ling. 2024. Corrective retrieval augmented generation. *arXiv preprint arXiv:2401.15884* (2024).

[12] Xu Yuan, Liangbo Ning, Wenqi Fan, and Qing Li. 2025. mKG-RAG: Multimodal Knowledge Graph-Enhanced RAG for Visual Question Answering. arXiv:2508.05318 [cs.CV] https://arxiv.org/abs/2508.05318