Cross-modal Reasoning-based Unsupervised Multi-modal Entity Linking

Yongtao Tang[1], Shasha Li[1] *, Jun Ma[1]*, Bin Ji[1], Xiaodong Liu[1], and Jie Yu[1]*

National University of Defense Technology, China
{tangyt,shashali,majun,jibin,liuxiaodong,yj}@nudt.edu.cn

Abstract. Entity Linking (EL) plays a crucial role in mapping textual mentions to corresponding entities in structured knowledge bases, while Multi-modal Entity Linking (MEL) extends this task by incorporating both textual and visual information. A key challenge in MEL is effectively utilizing multi-modal contextual data to improve entity disambiguation, particularly when candidate entities are highly similar. In this paper, we propose a novel approach that leverages the cross-modal attention and reasoning capabilities of multi-modal large language models (MLLM) to enhance MEL in an unsupervised setting. Our model consists of an optimized re-ranking stage that reduces computational cost by narrowing down the number of candidate entities, and a comparative selection stage that improves entity mention identification accuracy by fully leveraging indirect and latent information to distinguish highly similar candidates. Experimental results on three benchmark datasets show significant improvements in top-1 accuracy, with our model achieving 94.83%, 94.19%, and 91.25% on the Wiki-MEL, Richpedia-MEL, and WikiDiverse datasets, respectively, surpassing state-of-the-art approaches by up to 4.95% or more.

Keywords: Multi-modal Entity Linking · Knowledge Graphs · Multi-modal · Large Language Model

1 Introduction

In the field of Natural Language Processing (NLP), Entity Linking (EL) plays a critical role in mapping textual entity mentions to their corresponding entity in structured knowledge bases. Multi-modal Entity Linking (MEL) extends traditional EL by leveraging both textual and visual information to improve the accuracy and robustness of entity linking tasks. A key challenge in MEL is fully exploiting multi-modal contextual information to accurately identify entity mentions. Existing models, which primarily rely on information embedding, typically use only the directly relevant data associated with entity mentions. As a result, they often fail to incorporate indirect, potential evidence and reasoning that could enhance entity disambiguation. This limitation becomes particularly apparent when distinguishing between candidate entities with high similarity.

* Co-corresponding authors

While current methods achieve over 98% accuracy in top-10 predictions and over 95% in top-5 predictions, the accuracy of top-1 predictions remains relatively low, with state-of-the-art approaches stagnating around 86%.

In this paper, we present a novel state-of-the-art approach for multi-modal entity linking by leveraging MLLM to extract, integrate, and infer cross-modal information in an unsupervised setting (without relying on domain-specific or dataset-specific corpora). Specifically, we introduce a comparative selection stage within the existing multi-modal entity linking framework, utilizing the cross-modal attention capabilities of large language models (LLMs) to extract semantic information related to entity mentions from visual modalities (e.g., images). Additionally, we exploit the reasoning abilities of these models to compare the semantic differences between entity mentions and candidate entities, ultimately selecting the most appropriate entity as the prediction. To address the high computational cost associated with MLLM, we propose modifications to the re-ranking stage of the framework, streamlining the candidate entity set and thus reducing the computational burden of the subsequent comparative selection stage, achieving a balance between efficiency and performance.

Experimental results demonstrate that our model significantly outperforms existing multi-modal entity linking approaches, even in an unsupervised setting. On the **Wiki-MEL** dataset, our model achieves a top-1 accuracy of 94.83%, which represents a 5.86% improvement over the previous state-of-the-art (SOTA) of 88.97%. On the **Richpedia-MEL** dataset, our model attains a top-1 accuracy of 94.19%, yielding a 10.89% improvement over the existing SOTA of 83.3%. Additionally, on the **WikiDiverse** dataset, our model achieves a top-1 accuracy of 91.25%, surpassing the current SOTA of 86.3% by 4.95%.

The main contributions of this work are threefold:

1. We leverage the cross-modal attention and reasoning capabilities of MLLM to more effectively exploit multi-modal contextual information associated with entity mentions.
2. We propose a novel three-stage framework that strikes an optimal balance between model efficiency and performance.
3. We conduct extensive experiments on three benchmark datasets, demonstrating substantial performance improvements across all datasets.

2 Related Work

The limitations of current methods have motivated researchers to explore the potential of multi-modal Entity Linking, which incorporates additional modalities beyond text, such as visual information, to enhance the entity linking process. Moon et al. [7] pioneered the concept of MEL, leveraging both text and image data to disambiguate entities in noisy social media posts. Zhang et al. [17] proposed an Interactive Learning Network to fully leverage multi-modal information at the knowledge level. Xing et al. [15] focused on exploiting fine-grained and dynamic alignment relations between entities and mentions. Luo et al. [6]

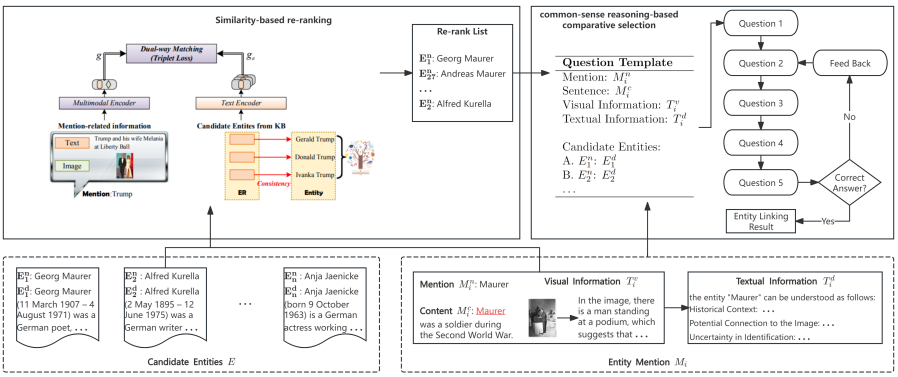

Fig. 1. Overview framework of the proposed multi-modal entity linking model.

proposed a new approach to understand the comprehensive expression of abbreviated textual context and implicit visual indications. Gan et al. [4] introduced a novel method for MEL by modeling the alignment of textual and visual mentions as a bipartite graph matching problem and published the M3EL dataset. Wang et al. [12] enhanced the feature extraction process by utilizing a multi-modal co-attention mechanism to extract hierarchical features of text and visual co-attention. Shi et al. [10] applied LLMs to the MEL task, while Song et al. [11] refined queries with multi-modal data. Zhang et al. [19] formulate the correlation assignment problem as an optimal transport (OT) problem, and propose a novel MEL framework, namely OT-MEL, with OT-guided correlation assignment.

Adjali et al. [1] contributed to the field by creating annotated datasets for evaluating MEL methods, including the release of the Twitter MEL dataset. This initiative was further expanded by Zhou et al. [20], who not only released three MEL datasets but also proposed a construction approach for such datasets. Wang et al. [13] presented Wikidiverse, a high-quality, human-annotated MEL dataset from Wikinews, known for its diversified contextual topics and entity types.

3 Methodology

3.1 Overall Framework

Existing multi-modal entity linking models typically re-rank candidate entities based on similarity scores, which are computed by comparing the embeddings of entity mentions with those of candidate entities. Embedding representation generation generally relies on integrating directly relevant information associated with entity mentions. However, this approach often overlooks indirect or potential evidence and fails to incorporate reasoning, which could enhance entity disambiguation.

We propose a three-stage multi-modal entity linking framework that leverages the cross-modal attention and reasoning capabilities of MLLM to more effectively extract, integrate, and utilize multi-modal context, thereby improving the accuracy of entity linking predictions. Additionally, the framework strikes a balance between efficiency and performance by progressively narrowing the set of candidate entities to reduce the computational cost of multi-modal inference processes.

Our proposed framework consists of three stages: multi-modal entity retrieval, similarity-based re-ranking, and cross-modal reasoning-based comparative selection. An overview of the framework is provided in Figure 1. We follow the multi-modal entity retrieval method used by Song et al. [11]. In the following sections, we detail the functions and specific implementations of the **similarity-based re-ranking** and **cross-modal reasoning-based comparative selection** stages.

3.2 Similarity-based Re-ranking

The similarity-based re-ranking stage serves two main objectives: first, to refine the candidate set in order to reduce the computational complexity of subsequent stages, and second, to enhance the ranking accuracy of candidate entity sets, thereby improving top-N accuracy and minimizing error propagation. To achieve this, we employ a scaling embedding model that transforms entity mentions and each candidate entity's information into embedding vectors. These vectors are then used for similarity calculations. Next, the candidate entities are ranked based on their similarity to the entity mentions, and the top-K candidates are selected to refine the candidate entity set.

Experimental results demonstrate the effectiveness of this approach. By encoding text information using the Thousand Question Embedding model and applying cosine similarity for sorting, we achieved top-10 accuracy rates of 94.02% and 96.18% on the WikiMEL and RichpediaMEL datasets, respectively. Furthermore, the mixed similarity method proposed by DWE+ for candidate entity screening yielded even higher top-10 accuracy rates of 98.8% and 98.9% respectively. These impressive top-10 accuracy values suggest that the error propagation introduced by the re-ranking stage is minimal and can be considered negligible.

3.3 Cross-modal Reasoning-based Comparative Selection

The cross-modal reasoning-based comparative selection stage is the core of our proposed model, designed to leverage the cross-modal attention and reasoning capabilities of MLLM to overcome the limitations of existing entity linking technologies through fully leveraging indirect and latent information to distinguish highly similar candidates, and achieve more accurate entity linking. We optimize the use of MLLM from three key dimensions to fully harness their potential in multi-modal entity linking tasks:

- **Depth of Thinking**: We employ multi-rounds dialogue to guide the LLM in simulating human-like reasoning, enabling it to incrementally generate answers rather than providing them all at once which mimics the iterative thought process of human decision-making.
- **Instruction Compliance**: We use controlled decoding techniques [8] to ensure that the generated answers follow the required format, thereby mitigating issues caused by the inherent randomness in LLM outputs.
- **Model Hallucination**: To reduce errors caused by the model's potential "hallucination" or overconfidence, we introduce a self-validation step that helps identify and correct incorrect answers, improving overall reliability.

In this stage, we format the multi-modal entity linking task as a single-choice task, rather than a re-ranking task, to more effectively leverage the understanding and reasoning capabilities of LLMs.

Our approach begins by using a MLLM to generate a detailed textual description T_i^v for the image associated with each entity mention. We then integrate all available information about the entity mention—such as its name, context, and the textual description of its corresponding image—to produce a comprehensive description T_i^d of the entity mention.

It is important to note that, based on empirical findings, we incorporate explicit instructions to prevent the LLM from making speculative inferences about the meaning of entity mentions based on external knowledge. This precaution is essential to avoid the influence of erroneous speculations, which could adversely affect the subsequent candidate comparison and selection process.

Subsequently, we organize the information of entity mentions and candidate entities into a single-choice question format.

To efficiently guide the LLM in discerning semantic differences between entity mentions and candidate entities, we have devised a multi-rounds dialogue process that emulates human reasoning. This process involves using single-choice questions to select the optimal candidate entity. The following instructions are utilized throughout the multi-rounds dialogue session:

- **System Role Definition:** Establishes the role of the system to align with task requirements.
- **Initial Question Analysis (Q_1):** Analyzes each option in a single-choice question to assist in decision-making, increasing the likelihood of selecting the correct answer.
- **Answer Selection (Q_2):** Selects the correct answer based on the analysis of the options in the single-choice question.
- **Answer Formatting (Q_3):** Generates clearly formatted, easy-to-parse single-choice answers using formatting instructions through controlled decoding techniques.
- **Self-validation (Q_4):** Removes the conversation history from Q_2 and Q_3 and asks the LLM to verify the correctness of the answer, enhancing the quality of the response.
- **Final Judgment (Q_5):** Generates a final judgment answer with clear formatting instructions through controlled decoding techniques [8].

In the multi-rounds dialogue process, we use the LLM's default text generation strategy to generate answers for questions Q_1, Q_2, and Q_4. For questions Q_3 and Q_5, we employ a controlled decoding strategy [8] to ensure that the LLM generates responses that conform to the predefined format.

If the answer generated by Q_3 fails the self-validation process in Q_5, we append the error reason to the historical context of Q_1 and restart the process from Q_2. If the number of restarts exceeds a predefined threshold, we produce a compromise result based on the available predictions, ensuring robustness even in the face of repeated failures.

4 Experiments

4.1 Datasets

In the experiments, we selected three public MEL datasets Richpedia-MEL [20] with 17K samples, Wiki-MEL [20] with 25K samples and WikiDiverse [13] with 15K samples to verify the effectiveness of our proposed method.

4.2 Baselines

We compare our method with various competitive baselines in three groups: 1) the text-based methods, which utilize textual information only to achieve EL, including ARNN [3], BERT [2], and Blink [14]. 2) the vision-and-language pre-training (VLP) methods including HieCoAtt [5], DZMNED [7], JMEL [1], MEL-HI [18], CLIP [9], GHMFC [12], MMEL [16], Drin [15], MIMIC [6], DWE [11], GEMEL [10], and OT-MEL [19]. Among them, DWE uses large model memory to obtain mention related knowledge, GEMEL uses fine-tuned LLM as a decoder, and OT-MEL distills large-scale pre trained models to transfer OT allocation knowledge to the attention mechanism.

4.3 Evaluation Metrics

To evaluate the efficacy of our multi-modal entity linking framework, we evaluate the top-1 accuracy predicted from 10 candidate entities, consistent across the Richpedia-MEL, Wiki-MEL, and Wikidiverse datasets, aligning our assessment with established benchmarks in the domain.

4.4 Experimental Setup

To ensure both fairness and reproducibility in our evaluations, we have meticulously detailed our experimental setup. Firstly, for equitable comparisons, we harness entity search results from the openly accessible dataset in DWE [11], which provides 100 candidate entities for each instance within the Richpedia-MEL and Wiki-MEL datasets. Secondly, re-ranking stage employs the qw-embedding model from the Qianwen platform as the primary text encoder. For the visual

Modality	Model	Richpedia-MEL	Wiki-MEL	Wikidiverse
Text-only	BERT	31.6	31.7	22.2
	blink	30.8	30.8	-
	ARNN	31.2	32.0	22.4
Visual-text	DZMNED	29.5	30.9	-
	JMEL	29.6	31.3	21.9
	MEL-HI	34.9	38.7	27.1
	HieCoAtt	37.2	40.5	28.4
	GHMFC	38.7	43.6	-
	MMEL	-	71.5	-
	CLIP	60.4	36.1	42.4
	Drin	-	65.5	-
	MIMIC	81.02	87.98	63.51
	DWE	67.6	44.7	47.5
	DWE+	72.5	72.8	51.2
	GEMEL	-	82.6	86.3
	OT-MEL	83.3	88.97	66.07
Visual-text	Ours	**94.19**	**94.83**	**91.25**

Table 1. Performance comparison between the selection stage and existing models on the wikiDiverse, wiki-MEL, and richpedia-MEL datasets, measured by top-1 accuracy

information extraction required in the selection stage, we utilize the MiniCPM-LLaMA3-V2.5 model, and for the multi-rounds dialog process, the LLaMA3-8B model is deployed. Thirdly, to circumvent the excessive duration on unresolvable samples, we restrict the maximum number of restarts during the answer self-validation process to three.

4.5 Experimental Results

We evaluated the effectiveness of the selection stage using top-1 accuracy across multiple datasets. The results, presented in Table 1, demonstrate significant improvements over previous SOTA methods on all three datasets. On the Wiki-MEL dataset, our model achieves a top-1 accuracy of 94.83%, representing a 5.86% improvement over the previous SOTA of 88.97%. On the Richpedia-MEL dataset, our model attains a top-1 accuracy of 94.19%, yielding a 10.89% improvement over the existing SOTA of 83.3%. Additionally, on the WikiDiverse dataset, our model achieves a top-1 accuracy of 91.25%, surpassing the current SOTA of 86.3% by 4.95%. These improvements can be attributed to the model's enhanced ability to leverage multi-modal context and its superior capacity for distinguishing semantic differences between entity mentions and candidate entities.

4.6 Ablation Study

Through ablation studies, we evaluated the effectiveness of each component in our proposed framework. Experiments were conducted on three datasets:

Model		Accuracy		
		Richpedia-MEL	Wiki-MEL	Wikidiverse
W/O Check	W/O Enhance	89.394	92.165	86.36
	Only Text	90.753	93.368	87.74
	Only Visual	91.850	94.195	89.73
	Both	93.207	94.499	90.64
Checked	W/O Enhance	90.672(1.278↑)	92.342(0.177↑)	87.11(0.75↑)
	Only Text	92.308(1.555↑)	94.000(0.632↑)	89.46(1.72↑)
	Only Visual	93.475(1.600↑)	94.382(0.187↑)	90.87(1.14↑)
	Both	94.189(0.982↑)	94.829(0.330↑)	91.25(0.61↑)

Table 2. Experimental results of ablation studies on the impact of textual and visual enhancements, and self-validation across datasets.

Richpedia-MEL, Wiki-MEL, and Wikidiverse, with a focus on analyzing model accuracy after systematically removing each component. This allowed us to assess the impact of different semantic enhancement strategies and the self-validation process on overall performance.

The results, as shown in Table 2, demonstrate that models with either text or visual enhancement outperform those without any enhancements. Notably, models utilizing visual enhancement slightly surpass those with only textual enhancement, underscoring the importance of incorporating visual information. When both text and visual enhancements are combined, model performance improves further, achieving the best results, thus highlighting the complementary nature of the two modalities. Regarding the self-validation process, the accuracy of all configurations improved after applying self-validation. However, the extent of the improvement varied significantly across datasets, suggesting that dataset characteristics may influence the effectiveness of the self-validation process.

4.7 Conclusion

In this paper, we proposed a novel approach for multi-modal entity linking (MEL) that leverages the cross-modal attention and reasoning capabilities of large multi-modal models in an unsupervised setting. Our method effectively addresses key challenges in MEL by fully exploiting multi-modal contextual information and incorporating reasoning beyond direct textual and visual evidence. Experimental results on three benchmark datasets, Wiki-MEL, Richpedia-MEL, and WikiDiverse, demonstrate substantial improvements in top-1 accuracy, outperforming existing state-of-the-art methods by up to 4.95% or more. These results highlight the effectiveness of our model in enhancing entity disambiguation, particularly in scenarios where candidate entities are highly similar.

Our contributions advance the field of MEL by combining cross-modal information extraction, reasoning, and efficient model design. Future work may focus on further improving model efficiency and expanding the scope of multi-modal data sources, such as incorporating external knowledge bases or additional visual modalities, to further enhance entity linking performance.

References

1. Adjali, O., Besançon, R., Ferret, O., Borgne, H.L., Grau, B.: Multimodal entity linking for tweets. In: Advances in Information Retrieval - 42nd European Conference on IR Research, ECIR 2020, Lisbon, Portugal, April 14-17, 2020, Proceedings, Part I. Lecture Notes in Computer Science, vol. 12035, pp. 463–478. Springer (2020)
2. Devlin, J., Chang, M., Lee, K., Toutanova, K.: BERT: pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers). pp. 4171–4186. Association for Computational Linguistics (2019)
3. Eshel, Y., Cohen, N., Radinsky, K., Markovitch, S., Yamada, I., Levy, O.: Named entity disambiguation for noisy text. In: Proceedings of the 21st Conference on Computational Natural Language Learning (CoNLL 2017), Vancouver, Canada, August 3-4, 2017. pp. 58–68. Association for Computational Linguistics (2017)
4. Gan, J., Luo, J., Wang, H., Wang, S., He, W., Huang, Q.: Multimodal entity linking: A new dataset and A baseline. In: MM '21: ACM Multimedia Conference, Virtual Event, China, October 20 - 24, 2021. pp. 993–1001. ACM (2021)
5. Lu, J., Yang, J., Batra, D., Parikh, D.: Hierarchical question-image co-attention for visual question answering. In: Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5-10, 2016, Barcelona, Spain. pp. 289–297 (2016)
6. Luo, P., Xu, T., Wu, S., Zhu, C., Xu, L., Chen, E.: Multi-grained multimodal interaction network for entity linking. In: Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD 2023, Long Beach, CA, USA, August 6-10, 2023. pp. 1583–1594. ACM (2023)
7. Moon, S., Neves, L., Carvalho, V.: Multimodal named entity disambiguation for noisy social media posts. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, July 15-20, 2018, Volume 1: Long Papers. pp. 2000–2008. Association for Computational Linguistics (2018)
8. Mudgal, S., Lee, J., Ganapathy, H., Li, Y., Wang, T., Huang, Y., Chen, Z., Cheng, H., Collins, M., Strohman, T., Chen, J., Beutel, A., Beirami, A.: Controlled decoding from language models. In: Forty-first International Conference on Machine Learning, ICML 2024, Vienna, Austria, July 21-27, 2024. OpenReview.net (2024)
9. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., Sutskever, I.: Learning transferable visual models from natural language supervision. In: Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 18-24 July 2021, Virtual Event. Proceedings of Machine Learning Research, vol. 139, pp. 8748–8763. PMLR (2021)
10. Shi, S., Xu, Z., Hu, B., Zhang, M.: Generative multimodal entity linking. In: Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation, LREC/COLING 2024, 20-25 May, 2024, Torino, Italy. pp. 7654–7665. ELRA and ICCL (2024)
11. Song, S., Zhao, S., Wang, C., Yan, T., Li, S., Mao, X., Wang, M.: A dual-way enhanced framework from text matching point of view for multimodal entity linking. In: Thirty-Eighth AAAI Conference on Artificial Intelligence, AAAI 2024, Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence, IAAI 2024,

Fourteenth Symposium on Educational Advances in Artificial Intelligence, EAAI 2014, February 20-27, 2024, Vancouver, Canada. pp. 19008–19016. AAAI Press (2024)
12. Wang, P., Wu, J., Chen, X.: Multimodal entity linking with gated hierarchical fusion and contrastive training. In: SIGIR '22: The 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, Madrid, Spain, July 11 - 15, 2022. pp. 938–948. ACM (2022)
13. Wang, X., Tian, J., Gui, M., Li, Z., Wang, R., Yan, M., Chen, L., Xiao, Y.: Wikidiverse: A multimodal entity linking dataset with diversified contextual topics and entity types. In: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2022, Dublin, Ireland, May 22-27, 2022. pp. 4785–4797. Association for Computational Linguistics (2022)
14. Wu, L., Petroni, F., Josifoski, M., Riedel, S., Zettlemoyer, L.: Scalable zero-shot entity linking with dense entity retrieval. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP 2020, Online, November 16-20, 2020. pp. 6397–6407. Association for Computational Linguistics (2020)
15. Xing, S., Zhao, F., Wu, Z., Li, C., Zhang, J., Dai, X.: DRIN: dynamic relation interactive network for multimodal entity linking. In: Proceedings of the 31st ACM International Conference on Multimedia, MM 2023, Ottawa, ON, Canada, 29 October 2023- 3 November 2023. pp. 3599–3608. ACM (2023)
16. Yang, C., He, B., Wu, Y., Xing, C., He, L., Ma, C.: MMEL: A joint learning framework for multi-mention entity linking. In: Uncertainty in Artificial Intelligence, UAI 2023, July 31 - 4 August 2023, Pittsburgh, PA, USA. Proceedings of Machine Learning Research, vol. 216, pp. 2411–2421. PMLR (2023)
17. Zhang, D., Huang, L.: Multimodal knowledge learning for named entity disambiguation. In: Findings of the Association for Computational Linguistics: EMNLP 2022, Abu Dhabi, United Arab Emirates, December 7-11, 2022. pp. 3160–3169. Association for Computational Linguistics (2022)
18. Zhang, L., Li, Z., Yang, Q.: Attention-based multimodal entity linking with high-quality images. In: Database Systems for Advanced Applications - 26th International Conference, DASFAA 2021, Taipei, Taiwan, April 11-14, 2021, Proceedings, Part II. Lecture Notes in Computer Science, vol. 12682, pp. 533–548. Springer (2021)
19. Zhang, Z., Sheng, J., Zhang, C., Liangyunzhi, L., Zhang, W., Wang, S., Liu, T.: Optimal transport guided correlation assignment for multimodal entity linking. In: Findings of the Association for Computational Linguistics, ACL 2024, Bangkok, Thailand and virtual meeting, August 11-16, 2024. pp. 4103–4117. Association for Computational Linguistics (2024)
20. Zhou, X., Wang, P., Li, G., Xie, J., Wu, J.: Weibo-mel, wikidata-mel and richpedia-mel: Multimodal entity linking benchmark datasets. In: Knowledge Graph and Semantic Computing: Knowledge Graph Empowers New Infrastructure Construction - 6th China Conference, CCKS 2021, Guangzhou, China, November 4-7, 2021, Proceedings. Communications in Computer and Information Science, vol. 1466, pp. 315–320. Springer (2021)