



ISWC
2023
NOVEMBER 6-10
ATHENS-GREECE

Rethinking Uncertainly Missing and Ambiguous Visual Modality in Multi-Modal Entity Alignment

Zhuo Chen^{1,2}, Lingbing Guo¹, Yin Fang¹, Yichi Zhang¹, Jiaoyan Chen³,

Jeff Z. Pan⁴, Yangning Li⁵, Huajun Chen^{1,2}, Wen Zhang^{1‡}

<https://github.com/zjukg/UMAEA>

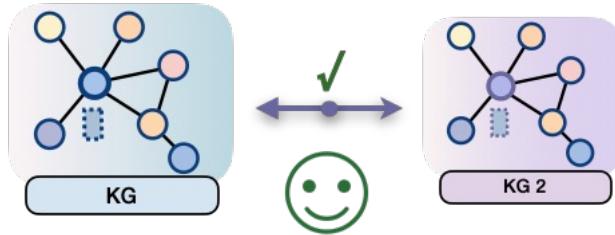
1 Zhejiang University *2* Donghai laboratory

3 The University of Manchester and The University of Oxford *4* The University of Edinburgh *5* Tsinghua University



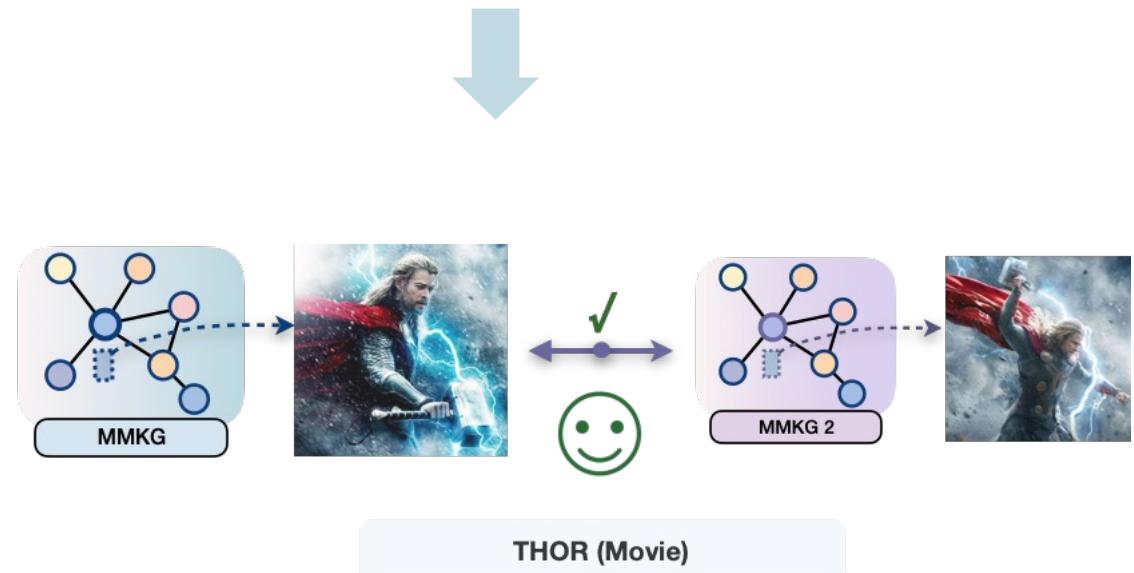
Background: MMEA

- **Entity Alignment (EA):**
 - *Identify equivalent entities across knowledge graphs (KGs):*
 - Addressing challenges such as disparate naming conventions, multilingualism, and heterogeneous graph structures



- **Multi-modal Entity Alignment (MMEA):**

- *Leverage visual modality:*
 - From the Internet as supplementary information for EA
 - Entity is associated with its name-related images
- *Current emphasis for MMEA approaches:*
 - Modality-specific attention weights
 - Visual-guided relational learning
 - Intra-modal enhancement via contrastive learning
 - Image filtering via ontology



Challenge: Missing and Ambiguous Visual Modality

- **Two Ideal Assumptions:**

- **One-to-one correspondence:**

- A single image sufficiently encapsulates and conveys all the information about an entity.

- **Images are always available:**

- An entity consistently possesses a corresponding image.

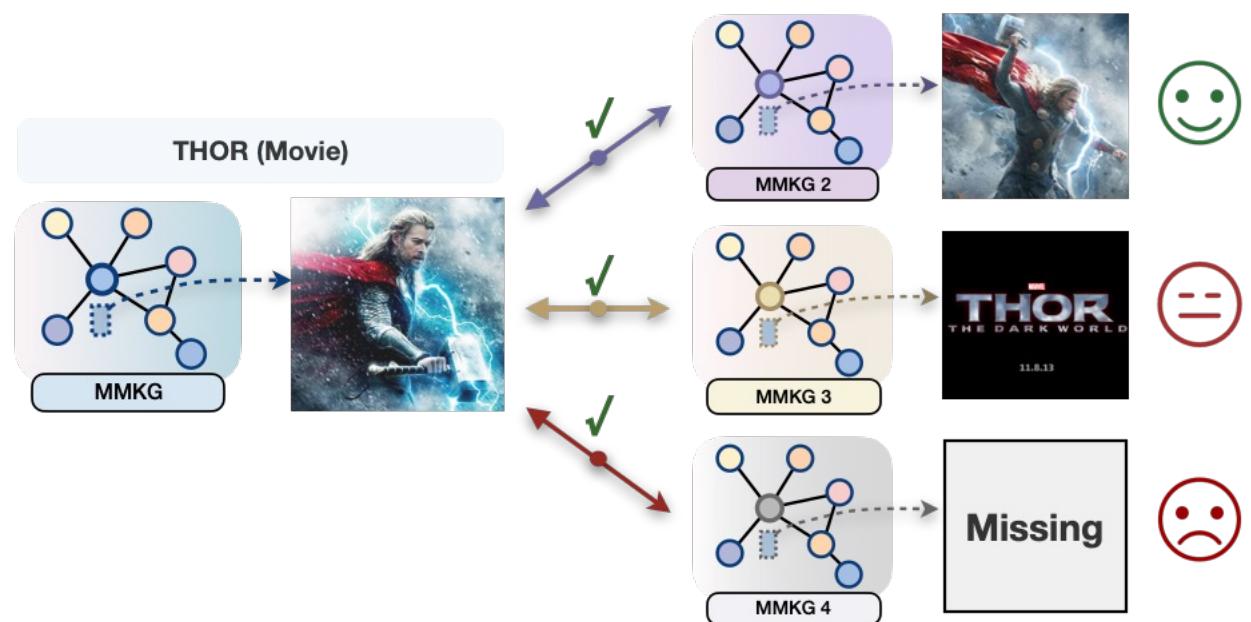
- **Practical Scenario:**

- **Uncertainly missing visual modality:**

- A varying degree of image absence

- **Uncertainly ambiguous visual modality:**

- A single entity might have heterogeneous visual representations



Contribution

- ***Propose the MMEA-UMVM dataset:***
 - Contain seven separate datasets with a total of **97 splits**
- ***Benchmark latest MMEA models:***
 - Standard (non-iterative) and iterative training paradigms
- ***Identify two critical phenomena:***
 - Models may succumb to **overfitting noise** during training
 - Models exhibit performance **oscillations or even declines** at high missing modality rates
- ***Propose our model UMAEA:***
 - multi-scale modality hybrid and circularly missing modality imagination
 - Consistently achieve SOTA results across **all** benchmark splits
 - limited parameters and runtime

Dataset: UMVM

- **UMVM (uncertainly missing visual modality) Datasets:**

- **Basic MMEA datasets:**

- **DBP15K:** multilingual versions of DBpedia
 - $DBP15K_{ZH-EN}$, $DBP15K_{JA-EN}$ and $DBP15K_{FR-EN}$
- **Multi-OpenEA:** multi-modal variants of the OpenEA
 - $EN-FR-15K$, $EN-DE-15K$, $D-W-15K-V1$, $D-W-15K-V2$



- **Random image dropping on MMEA datasets:**

- Varying degrees of visual modality missing
- Range from 0.05 to the maximum R_{img} of the raw datasets



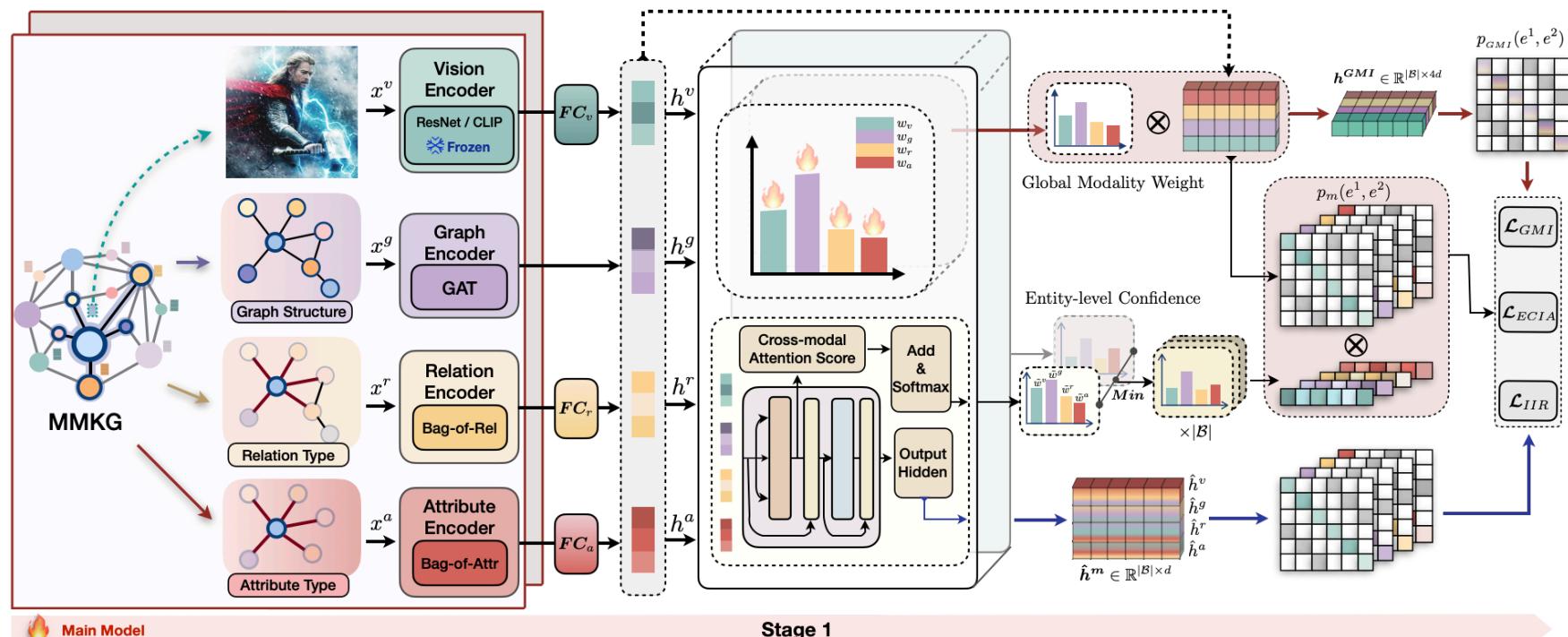
Dataset	KG	# Ent.	# Rel.	# Attr.	# Rel. Triples	# Attr. Triples	# Image	# EA pairs
DBP15K _{ZH-EN}	ZH (Chinese)	19,388	1,701	8,111	70,414	248,035	15,912	15,000
	EN (English)	19,572	1,323	7,173	95,142	343,218	14,125	
DBP15K _{JA-EN}	JA (Japanese)	19,814	1,299	5,882	77,214	248,991	12,739	15,000
	EN (English)	19,780	1,153	6,066	93,484	320,616	13,741	
DBP15K _{FR-EN}	FR (French)	19,661	903	4,547	105,998	273,825	14,174	15,000
	EN (English)	19,993	1,208	6,422	115,722	351,094	13,858	
OpenEA _{EN-FR}	EN (English)	15,000	267	308	47,334	73,121	15,000	15,000
	FR (French)	15,000	210	404	40,864	67,167	15,000	
OpenEA _{EN-DE}	EN (English)	15,000	215	286	47,676	83,755	15,000	15,000
	DE (German)	15,000	131	194	50,419	156,150	15,000	
OpenEA _{D-W-V1}	DBpedia	15,000	248	342	38,265	68,258	15,000	15,000
	Wikidata	15,000	169	649	42,746	138,246	15,000	
OpenEA _{D-W-V2}	DBpedia	15,000	167	175	73,983	66,813	15,000	15,000
	Wikidata	15,000	121	457	83,365	175,686	15,000	

Dataset	R_{img}
DBP15K _{ZH-EN}	0.05, 0.1, 0.15, 0.2, 0.3, 0.4, 0.45, 0.5, 0.55, 0.6, 0.7, 0.75, 0.7829 (STD)
DBP15K _{JA-EN}	0.05, 0.1, 0.15, 0.2, 0.3, 0.4, 0.45, 0.5, 0.55, 0.6, 0.7, 0.7032 (STD)
DBP15K _{FR-EN}	0.05, 0.1, 0.15, 0.2, 0.3, 0.4, 0.45, 0.5, 0.55, 0.6, 0.6758 (STD)
OpenEA _{EN-FR}	0.05, 0.1, 0.15, 0.2, 0.3, 0.4, 0.45, 0.5, 0.55, 0.6, 0.7, 0.8, 0.9, 0.95, 1.0 (STD)
OpenEA _{EN-DE}	0.05, 0.1, 0.15, 0.2, 0.3, 0.4, 0.45, 0.5, 0.55, 0.6, 0.7, 0.8, 0.9, 0.95, 1.0 (STD)
OpenEA _{D-W-V1}	0.05, 0.1, 0.15, 0.2, 0.3, 0.4, 0.45, 0.5, 0.55, 0.6, 0.7, 0.8, 0.9, 0.95, 1.0 (STD)
OpenEA _{D-W-V2}	0.05, 0.1, 0.15, 0.2, 0.3, 0.4, 0.45, 0.5, 0.55, 0.6, 0.7, 0.8, 0.9, 0.95, 1.0 (STD)

◎ The proportion R_{img} of entities containing images in our setting, with “STD” refers to the standard R_{img} in raw datasets

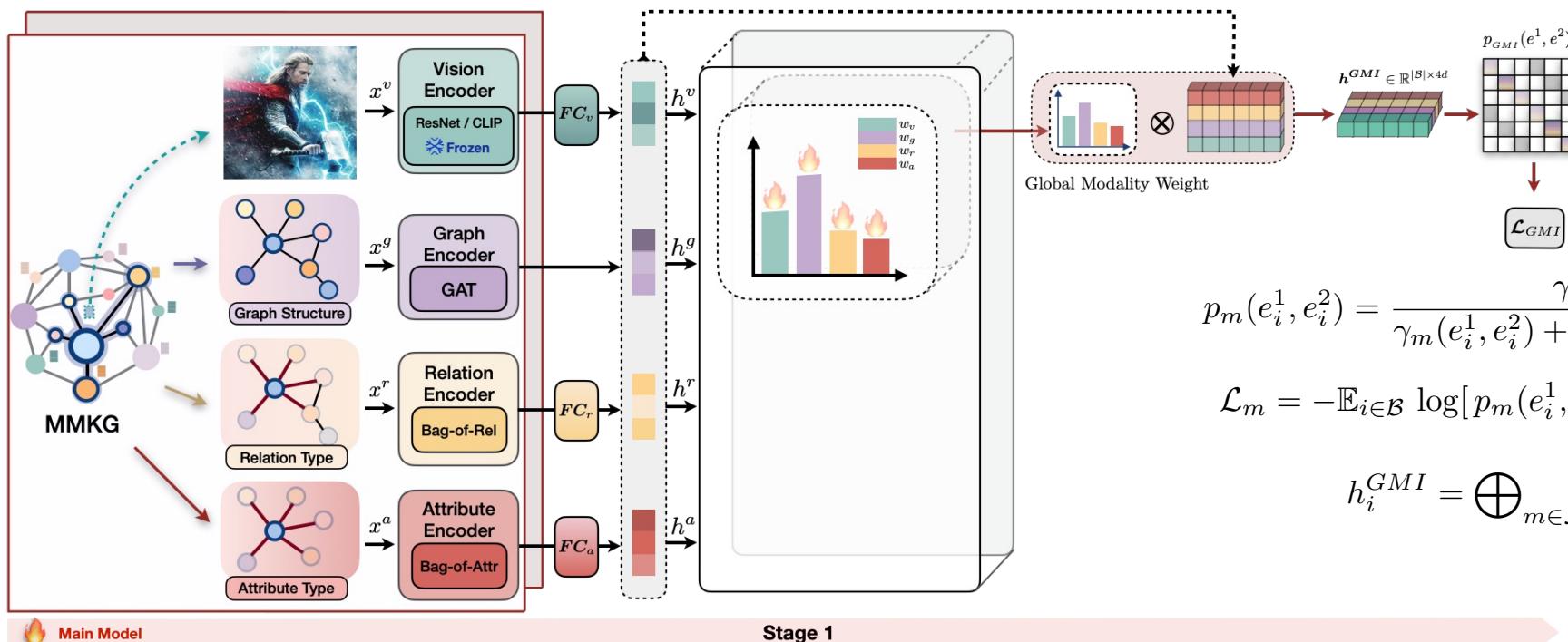
- **Multi-scale Modality Hybrid:**

- **Global Modality Integration (GMI)**
- **Entity-level Modality Alignment**
- **Late Modality Refinement**

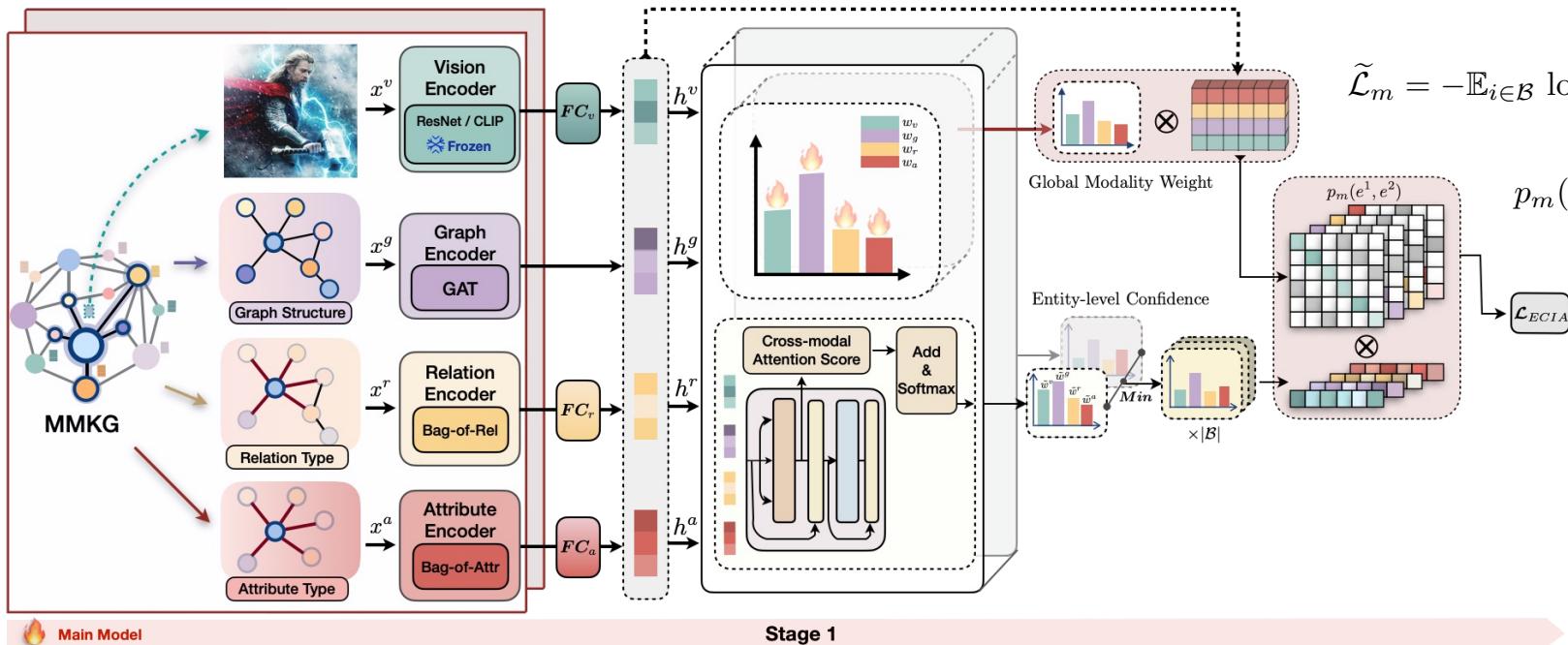


- **Multi-scale Modality Hybrid:**

- **Global Modality Integration (GMI)**
 - Entity-level Modality Alignment
 - Late Modality Refinement



- **Multi-scale Modality Hybrid:**
 - *Global Modality Integration (GMI)*
 - *Entity-level Modality Alignment*
 - *Explicit confidence-augmented intra-modal alignment (ECIA)*
 - *Implicit inter-modal refinement (IIR)*
 - *Late Modality Refinement*



$$\mathcal{L}_{ECIA} = \sum_{m \in \mathcal{M}} \tilde{\mathcal{L}}_m ,$$

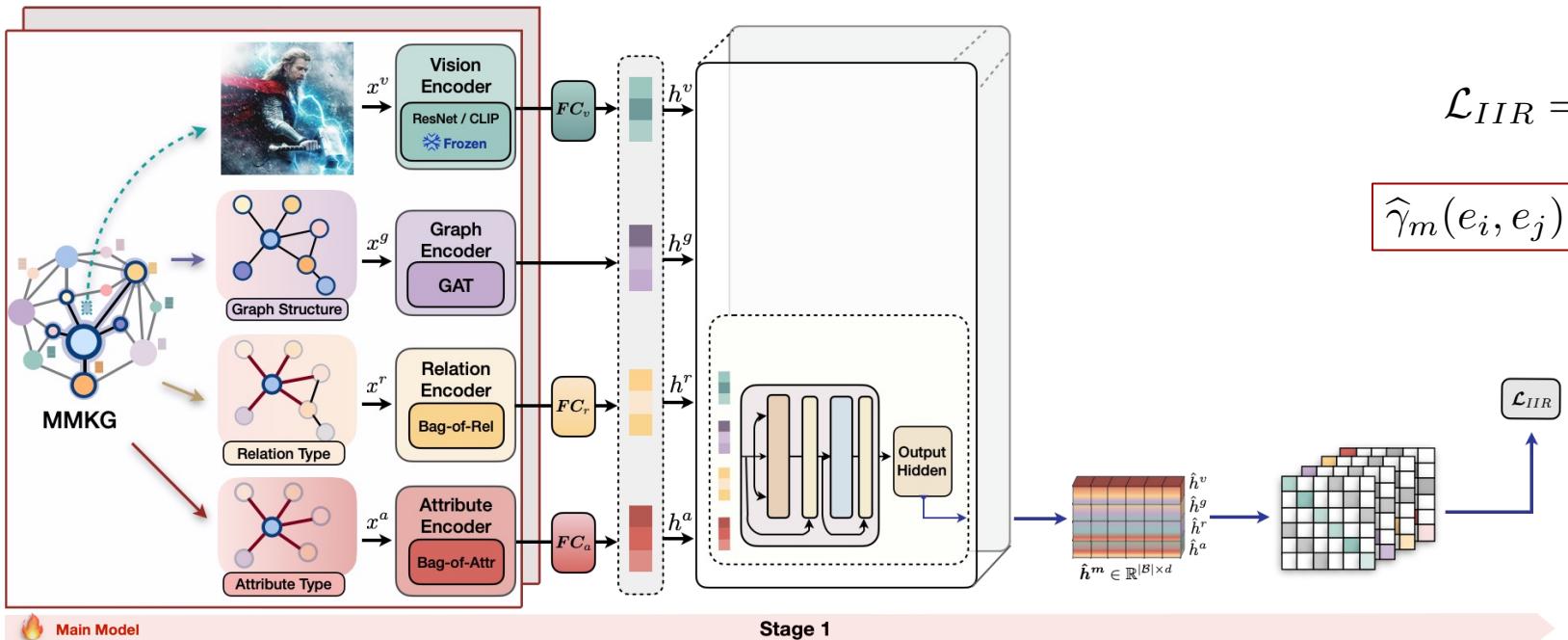
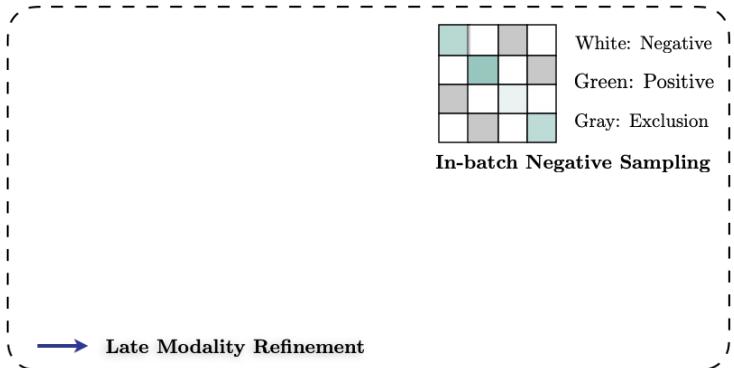
$$\tilde{\mathcal{L}}_m = -\mathbb{E}_{i \in \mathcal{B}} \log[\phi_m(e_i^1, e_i^2) * (p_m(e_i^1, e_i^2) + p_m(e_i^2, e_i^1))] / 2 .$$

$$p_m(e_i^1, e_i^2) = \frac{\gamma_m(e_i^1, e_i^2)}{\gamma_m(e_i^1, e_i^2) + \sum_{e_j \in \mathcal{N}_i^{ng}} \gamma_m(e_i^1, e_j)} ,$$

$$\phi_m(e_i, e_j) = \text{Min}(\tilde{w}_i^m, \tilde{w}_j^m)$$

$$\gamma_m(e_i, e_j) = \exp(h_i^{m\top} h_j^m / \tau)$$

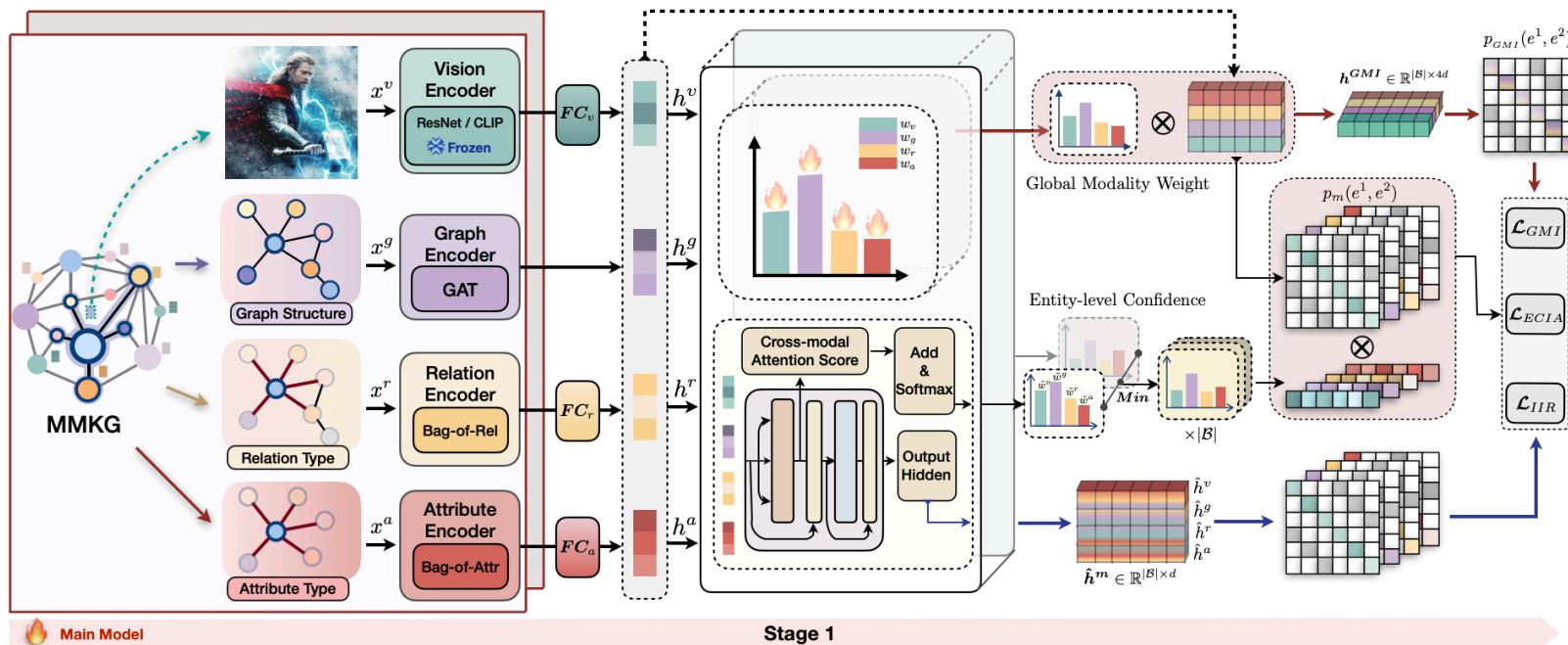
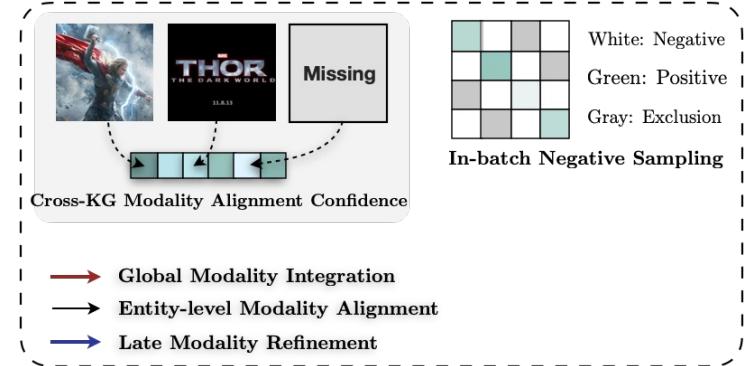
- **Multi-scale Modality Hybrid:**
 - *Global Modality Integration (GMI)*
 - *Entity-level Modality Alignment*
 - *Explicit confidence-augmented intra-modal alignment (ECIA)*
 - *Implicit inter-modal refinement (IIR)*
 - **Late Modality Refinement**



$$\mathcal{L}_{IIR} = \sum_{m \in \mathcal{M}} \hat{\mathcal{L}}_m ,$$

$$\hat{\gamma}_m(e_i, e_j) = \exp(\hat{h}_i^m \top \hat{h}_j^m / \tau).$$

- **Multi-scale Modality Hybrid:**
 - *Global Modality Integration (GMI)*
 - *Entity-level Modality Alignment*
 - *Late Modality Refinement*



- **Circularly Missing Modality Imagination:**

- **Global Modality Integration (GMI)**
- **Entity-level Modality Alignment**
- **Late Modality Refinement**

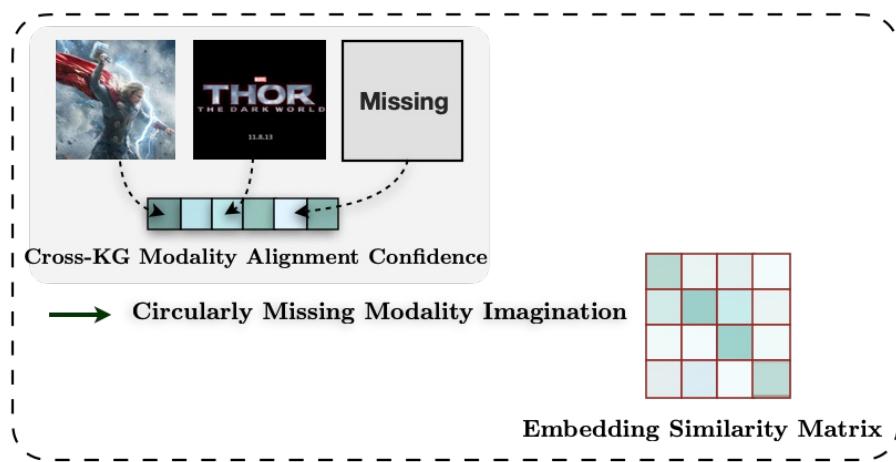
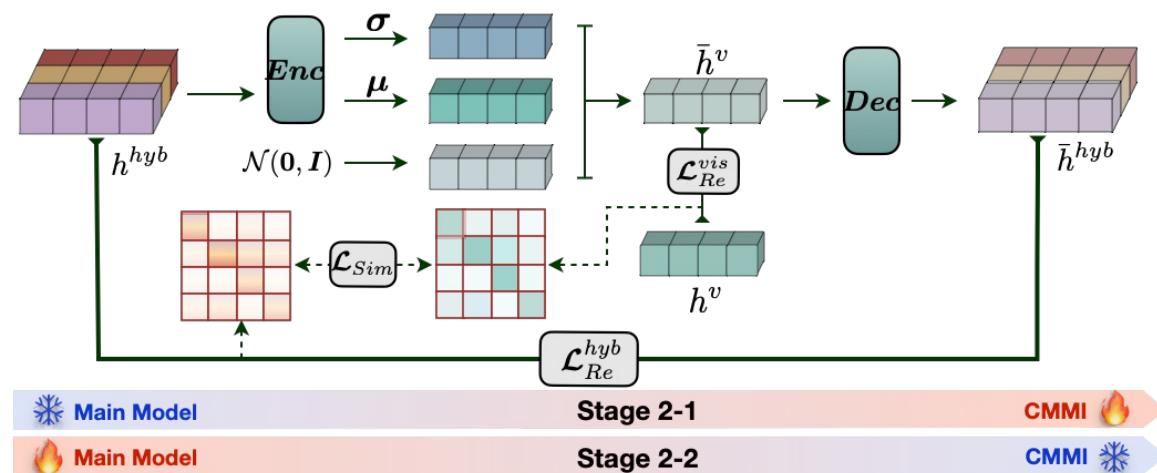
$$h_i^{hyb} = [h_i^r \oplus h_i^a \oplus h_i^g]$$

$$\begin{aligned} [\mu_i \oplus \log(\sigma_i)^2] &= MLP_{Enc}(h_i^{hyb}), \\ \bar{h}_i^v &= z \odot \sigma_i + \mu_i, \quad z \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \\ \bar{h}_i^{hyb} &= MLP_{Dec}(\bar{h}_i^v). \end{aligned}$$

$$\mathcal{L}_{KL} = \mathbb{E}_{i \in \bar{\mathcal{B}}} ((\mu_i)^2 + (\sigma_i)^2 - \log(\sigma_i)^2 - 1)/2,$$

$$\mathcal{L}_{Sim} = \mathbb{E}_{i \in \bar{\mathcal{B}}} D_{KL}(p_{hyb}(e_i^1, e_i^2) || \bar{p}_v(e_i^1, e_i^2)),$$

$$\mathcal{L}_2 = \mathcal{L}_{KL} + \mathcal{L}_{Re}^{vis} + \mathcal{L}_{Re}^{hyb} + \mathcal{L}_{Sim}.$$



Training Details

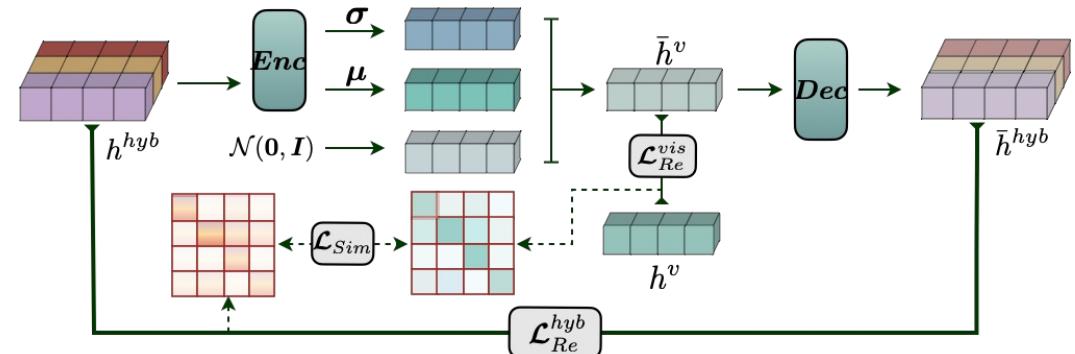
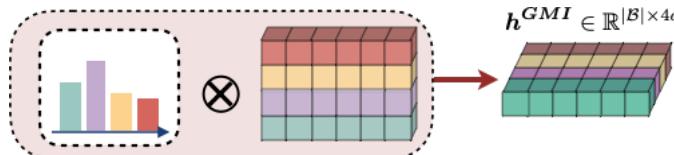
- **Pipeline:**

$$\begin{aligned} \text{Stage 1} : \mathcal{L} &\leftarrow \mathcal{L}_1, \\ \text{Stage 2-1/2-2} : \mathcal{L} &\leftarrow \mathcal{L}_1 + \mathcal{L}_2, \end{aligned}$$



- **Entity Representation:**

- *For those entities without images:*
 - *Evaluation:* replace the original random vectors with the generated μ_i
 - *Training:* pseudo-visual embedding \bar{h}_i^v
- *Final multi-modal entity representation:* h_i^{GMI}



Experiments

② Iterative Training

- *Concretely, every K_e (where $K_e = 5$) epochs, we propose cross-KG entity pairs that are **mutual nearest neighbors** in the vector space and add them to a **candidate list** N_{cd}*
- *An entity pair in N_{cd} will be added into the training set if it **remains a mutual nearest neighbor** for K_s ($= 10$) consecutive rounds.*

Main Experiments

- Non-iter. Results on Bilingual Datasets**

	Models	$R_{img} = 0.05$			$R_{img} = 0.2$			$R_{img} = 0.4$			$R_{img} = 0.6$		
		H@1	H@10	MRR	H@1	H@10	MRR	H@1	H@10	MRR	H@1	H@10	MRR
DBP15K_{ZH-EN}													
MSNEA [7]	.413	.722	.517	.411	.725	.518	.446	.743	.546	.520	.786	.611	
EVA [30]	.623	.878	.715	.624	.878	.716	.623	.875	.714	.625	.876	.717	
MCLEA [29]	.638	.905	.732	.588	.865	.686	.611	.874	.704	.661	.896	.744	
w/o CMMI	.703	.934	.787	.710	.937	.793	.721	.939	.801	.753	.949	.825	
UMAEA	.720	.938	.800	.727	.941	.806	.727	.941	.806	.758	.951	.829	
Improve ↑	8.2%	3.3%	.068	10.3%	6.3%	.090	10.4%	6.6%	.092	9.7%	5.5%	.085	
DBP15K_{JA-EN}													
MSNEA [7]	.313	.643	.425	.311	.644	.422	.369	.678	.472	.480	.744	.569	
EVA [30]	.615	.877	.708	.616	.877	.710	.616	.878	.711	.624	.881	.716	
MCLEA [29]	.599	.897	.706	.579	.846	.675	.613	.867	.703	.686	.898	.761	
w/o CMMI	.708	.943	.794	.712	.947	.798	.730	.950	.810	.772	.962	.843	
UMAEA	.725	.949	.807	.726	.949	.808	.732	.952	.813	.775	.963	.845	
Improve ↑	11.0%	5.2%	.099	11.0%	7.2%	.098	11.6%	7.4%	.102	8.9%	6.5%	.084	
DBP15K_{FR-EN}													
MSNEA [7]	.297	.690	.427	.304	.690	.428	.360	.710	.474	.478	.772	.574	
EVA [30]	.624	.895	.720	.624	.895	.720	.626	.898	.721	.634	.900	.728	
MCLEA [29]	.634	.930	.741	.582	.863	.682	.601	.879	.702	.675	.901	.757	
w/o CMMI	.727	.956	.813	.733	.960	.817	.746	.961	.828	.790	.968	.857	
UMAEA	.752	.970	.830	.755	.960	.832	.763	.962	.838	.792	.970	.859	
Improve ↑	11.8%	4.0%	.089	13.1%	6.7%	.112	13.7%	6.4%	.117	11.7%	6.9%	.102	

	Models	$R_{img} = 0.05$			$R_{img} = 0.2$			$R_{img} = 0.4$			$R_{img} = 0.6$		
		H@1	H@10	MRR	H@1	H@10	MRR	H@1	H@10	MRR	H@1	H@10	MRR
OpenEA_{EN-FR}													
MSNEA [7]	.200	.431	.278	.213	.439	.290	.260	.477	.334	.360	.560	.427	
EVA [30]	.528	.833	.634	.533	.835	.638	.539	.835	.642	.547	.830	.647	
MCLEA [29]	.545	.852	.653	.547	.852	.655	.531	.839	.637	.597	.852	.688	
w/o CMMI	.587	.893	.695	.590	.893	.697	.614	.900	.715	.664	.912	.753	
UMAEA	.605	.898	.708	.604	.896	.708	.618	.899	.718	.665	.914	.753	
Improve ↑	6.0%	4.6%	.055	5.7%	4.4%	.053	7.9%	6.1%	.076	6.8%	6.2%	.065	
OpenEA_{EN-DE}													
MSNEA [7]	.242	.486	.323	.253	.495	.333	.309	.542	.387	.412	.622	.484	
EVA [30]	.717	.917	.787	.718	.918	.788	.721	.920	.791	.734	.921	.800	
MCLEA [29]	.723	.918	.791	.721	.915	.789	.697	.907	.771	.745	.906	.803	
w/o CMMI	.752	.938	.818	.757	.941	.822	.771	.946	.833	.804	.954	.858	
UMAEA	.757	.942	.823	.759	.943	.824	.774	.947	.835	.804	.957	.860	
Improve ↑	3.4%	2.4%	.032	3.8%	2.5%	.035	5.3%	2.7%	.044	5.9%	3.6%	.057	
OpenEA_{D-W-V1}													
MSNEA [7]	.238	.452	.31	.254	.465	.326	.318	.514	.385	.432	.601	.490	
EVA [30]	.570	.801	.653	.575	.806	.658	.567	.797	.650	.595	.811	.673	
MCLEA [29]	.585	.834	.675	.574	.824	.663	.581	.813	.665	.848	.726		
w/o CMMI	.640	.879	.727	.644	.882	.730	.667	.891	.749	.722	.908	.790	
UMAEA	.647	.881	.733	.649	.882	.735	.669	.892	.750	.724	.908	.791	
Improve ↑	6.2%	4.7%	.058	7.4%	5.8%	.072	8.8%	7.9%	.085	6.9%	6.0%	.065	
OpenEA_{D-W-V2}													
MSNEA [7]	.397	.690	.497	.405	.695	.503	.454	.727	.546	.545	.781	.626	
EVA [30]	.775	.952	.839	.767	.947	.832	.773	.950	.837	.788	.954	.848	
MCLEA [29]	.771	.965	.842	.753	.957	.827	.757	.935	.822	.800	.948	.855	
w/o CMMI	.828	.983	.883	.829	.982	.885	.844	.984	.896	.857	.986	.905	
UMAEA	.840	.984	.890	.832	.982	.887	.844	.984	.896	.859	.987	.905	
Improve ↑	6.5%	1.9%	.048	6.5%	2.5%	.055	7.1%	3.4%	.059	5.9%	3.3%	.050	

- Non-iterative results of four models with “w/o CMMI” setting indicating the absence of the Stage-2. The best results within the baselines are marked with underline, and we highlight our results with **bold** when we achieve SOTA.

Main Experiments

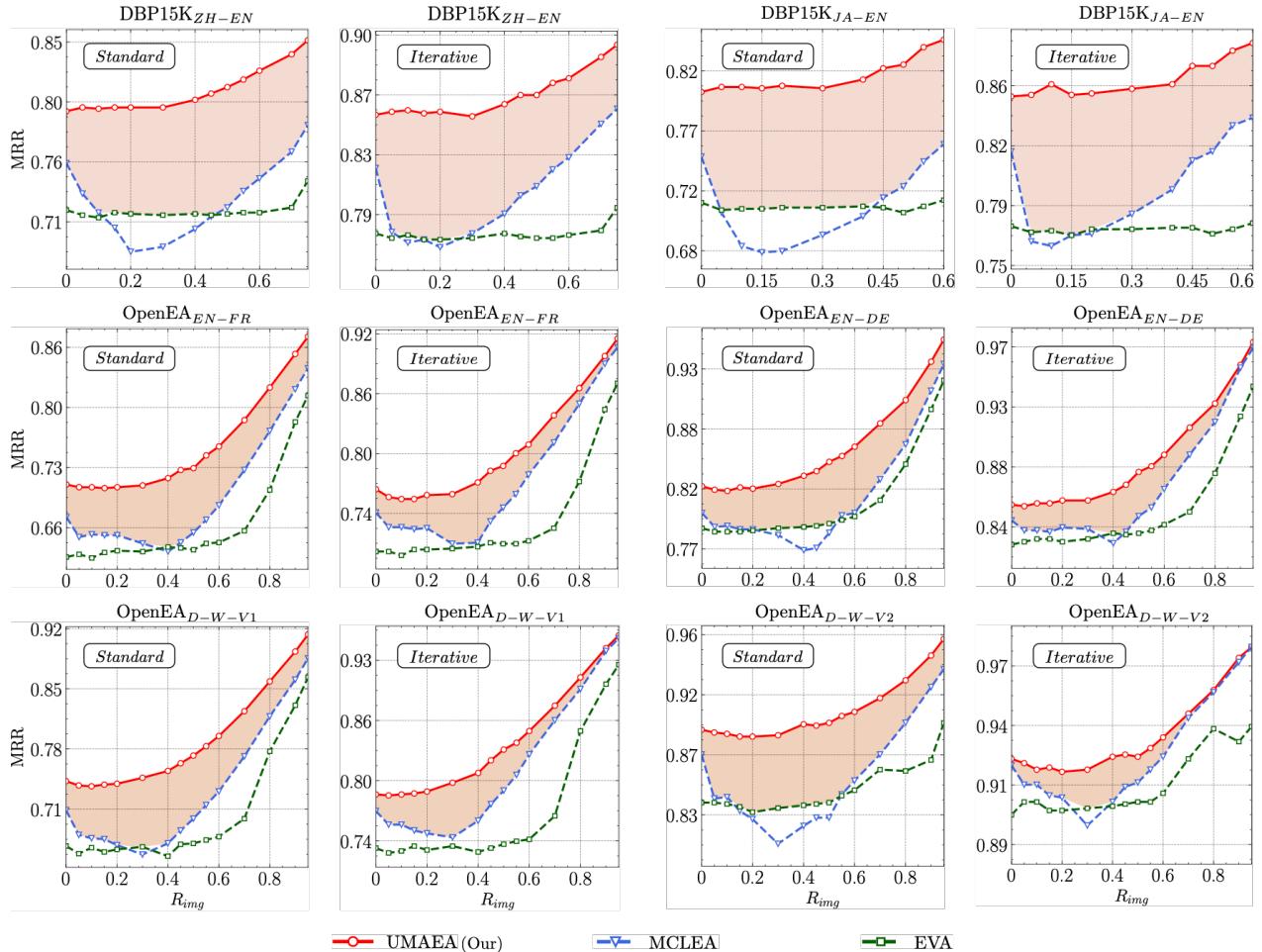
- **Model Overall Performance**

➤ Existing models exhibit performance oscillations (EVA) or declines (MCLEA) at higher modality missing

□ Introducing images for half of the entities means that the remaining half may become noise

➤ Our method can gain benefits with fewer visual modality data in entity.

□ Less oscillation and greater robustness than other methods



Main Experiments

- **Complete Visual Modality**
 - Prove the robustness of the models against ambiguous images

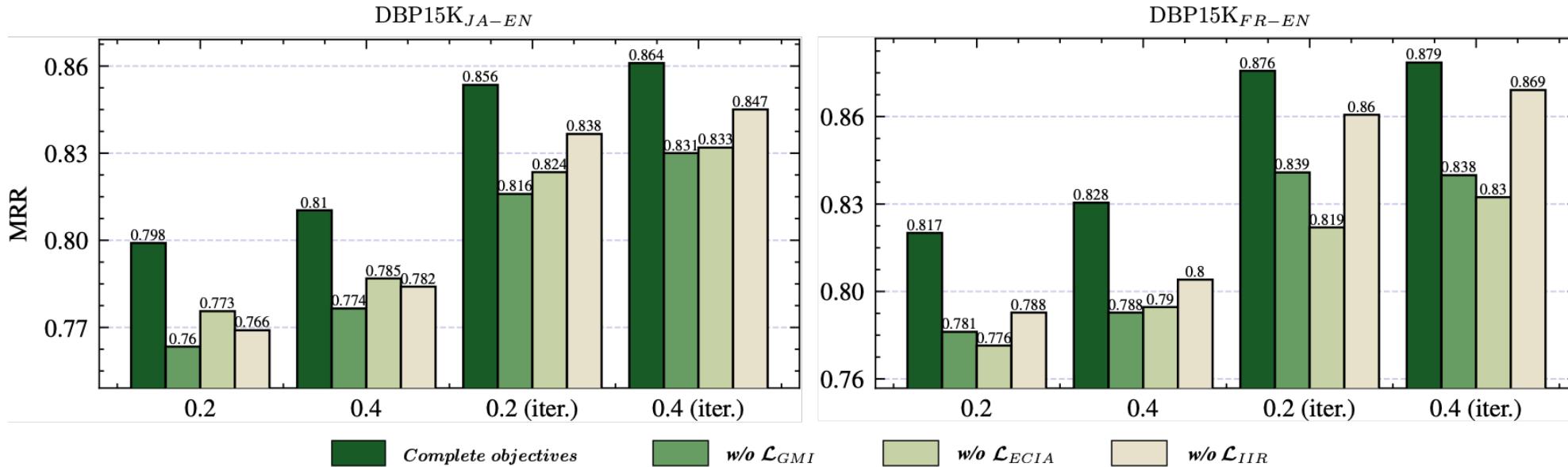
Models	DBP15K _{ZH-EN}			DBP15K _{JA-EN}			DBP15K _{FR-EN}			
	H@1	H@10	MRR	H@1	H@10	MRR	H@1	H@10	MRR	
Non-iter.	AlignEA [39]	.472	.792	.581	.448	.789	.563	.481	.824	.599
	KECG [27]	.478	.835	.598	.490	.844	.610	.486	.851	.610
	MUGNN [3]	.494	.844	.611	.501	.857	.621	.495	.870	.621
	AliNet [41]	.539	.826	.628	.549	.831	.645	.552	.852	.657
	MSNEA* [7]	.609	.831	.685	.541	.776	.620	.557	.820	.643
	EVA* [30]	.683	.906	.762	.669	.904	.752	.686	.928	.771
	MCLEA* [29]	.726	.922	.796	.719	.915	.789	.719	.918	.792
	UMAEAE*	.800	.962	.860	.801	.967	.862	.818	.973	.877
Iter.	w/o IMG	.718	.930	.797	.723	.941	.803	.748	.956	.826
	BootEA [39]	.629	.847	.703	.622	.854	.701	.653	.874	.731
	NAEA [62]	.650	.867	.720	.641	.873	.718	.673	.894	.752
	MSNEA* [7]	.648	.881	.728	.557	.804	.643	.583	.848	.672
	EVA* [30]	.750	.912	.810	.741	.921	.807	.765	.944	.831
	MCLEA* [29]	.811	.957	.865	.805	.958	.863	.808	.963	.867
	UMAEAE*	.856	.974	.900	.857	.980	.904	.873	.988	.917
	w/o IMG	.793	.952	.852	.794	.960	.857	.820	.976	.880

Models	OpenEA _{EN-FR}			OpenEA _{EN-DE}			OpenEA _{D-W-V1}			OpenEA _{D-W-V2}			
	H@1	H@10	MRR	H@1	H@10	MRR	H@1	H@10	MRR	H@1	H@10	MRR	
Non-iter.	MSNEA* [7]	.692	.813	.734	.753	.895	.804	.800	.874	.826	.838	.940	.873
	EVA* [30]	.785	.932	.836	.922	.983	.945	.858	.946	.891	.890	.981	.922
	MCLEA* [29]	.819	.943	.864	.939	.988	.957	.881	.955	.908	.928	.983	.949
	UMAEAE*	.848	.966	.891	.956	.994	.971	.904	.971	.930	.948	.996	.967
Iter.	MSNEA* [7]	.699	.823	.742	.788	.917	.835	.809	.885	.836	.862	.954	.894
	EVA* [30]	.849	.974	.896	.956	.985	.968	.915	.986	.942	.925	.996	.951
	MCLEA* [29]	.888	.979	.924	.969	.993	.979	.944	.989	.963	.969	.997	.982
	UMAEAE*	.895	.987	.931	.974	.998	.984	.945	.994	.965	.973	.999	.984

- Non-iterative (Non-iter.) and iterative (Iter.) results on seven MMEA datasets (complete version), where “*” refers to involving the visual information for EA. The DBP15K dataset (left table) only has part of the entities with images attached (e.g., 78.29% in DBP15K_{ZH-EN}, 70.32% in DBP15K_{FR-EN}, and 67.58% in DBP15K_{JA-EN})

Experiments: Details Analysis

- **Component Analysis**



- IIR serves as an enhancement for ECIA, and its influence is comparatively less significant than that of GMI and ECIA

Experiments: Details Analysis

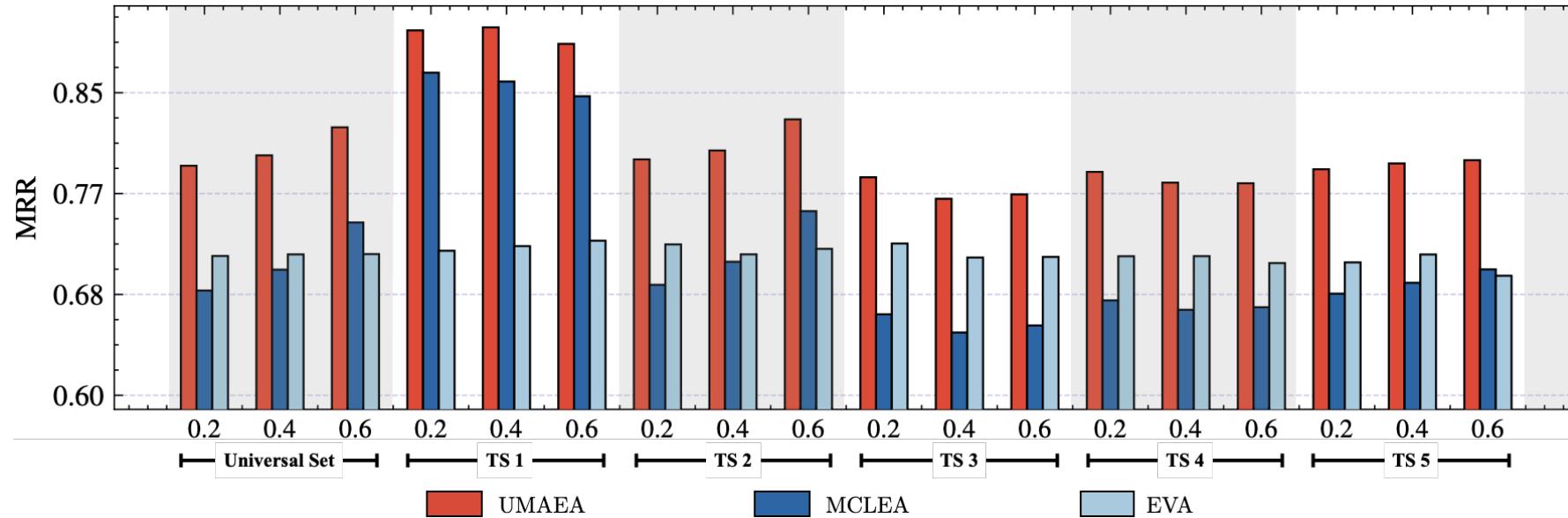
- **Efficiency Analysis**

Models	DBP15K _{JA-EN}			DBP15K _{FR-EN}			OpenEA _{EN-FR}		
	Para. (M)	Time (Min)	MRR	Para. (M)	Time (Min)	MRR	Para. (M)	Time (Min)	MRR
EVA* [30]	13.27	30.9	.711	13.29	30.8	.721	9.81	17.8	.642
MCLEA* [29]	13.22	15.3	.703	13.24	15.7	.702	9.75	19.5	.637
w/o CMMI	13.82	30.2	.810	13.83	28.8	.828	10.35	17.9	.715
UMAEA	14.72	33.4	.813	14.74	32.7	.838	11.26	23.1	.718

- *Relationship between model parameter size (Para.), training time (Time), and performance (MRR)*
- *UMAEA improves the performance with only a minor increase in parameters and time consumption*

Experiments: Details Analysis

- Entity Distribution Analysis



- Models (EVA) succumb to **overfitting noise** during training
- Models (MCLEA) exhibit performance **oscillations or even declines** at high missing modality rates

- EA prediction distribution analysis on DBP15K ZH-EN(non-iterative), with $R_{img} \in \{0.2, 0.4, 0.6\}$. "TS" denotes the testing set, where:
 - TS 1 (both entities in an alignment pair have images);
 - TS 2 (at least one entity in an alignment pair has images);
 - TS 3 (only one entity in an alignment pair has images);
 - TS 4 (at least one entity in an alignment pair loss images);
 - TS 5 (neither entity in an alignment pair has images).

Thank you!

<https://github.com/zjukg/UMAEA>

Rethinking Uncertainly Missing and Ambiguous Visual Modality in Multi-Modal Entity Alignment

