

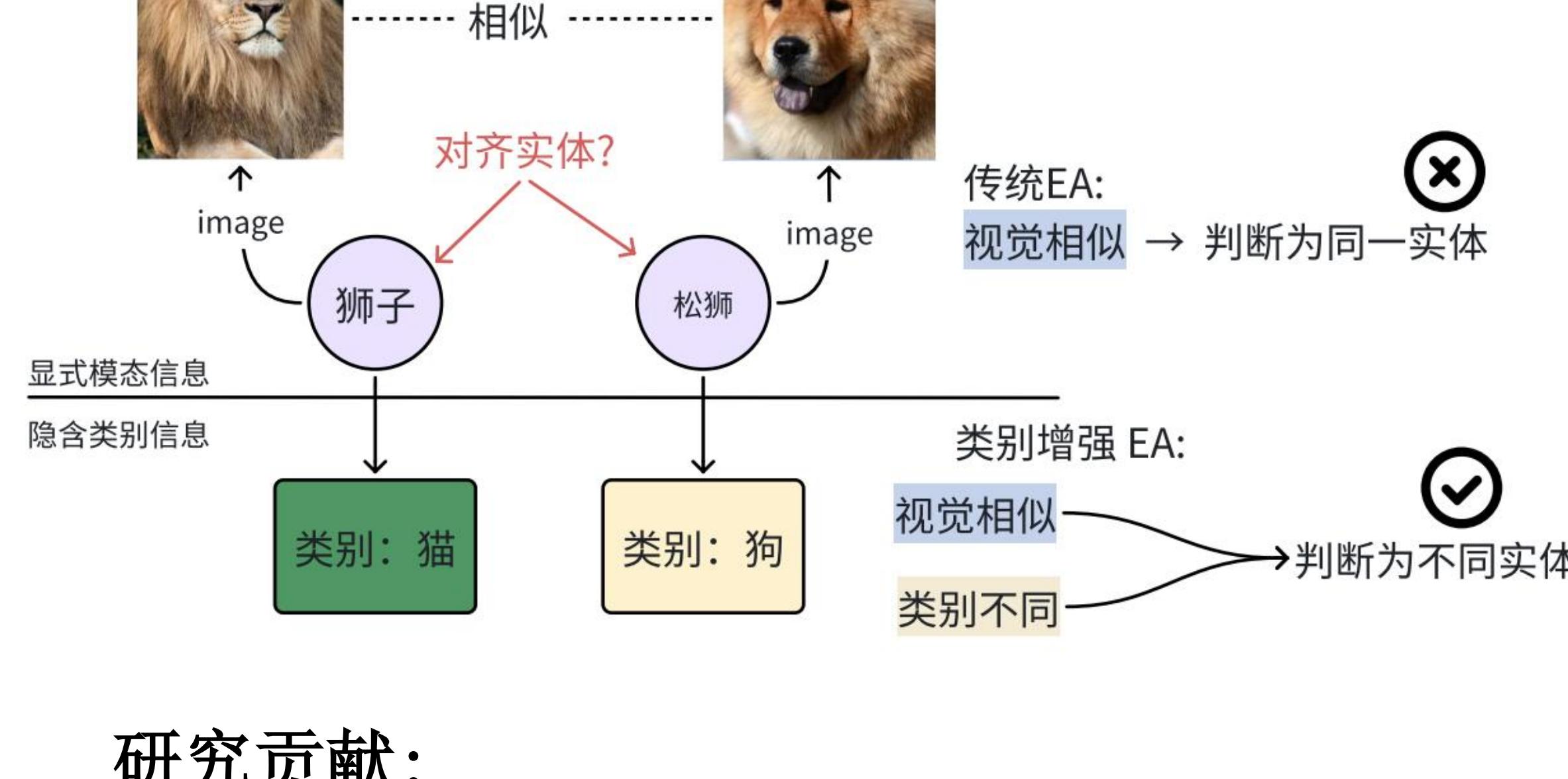
# CateEA: Enhancing Entity Alignment via Implicit Category Supervision

冯冠栋, 任涛\*, 胡军, 王丹丹\*

COLING 2025, Proceedings of the 31st International Conference on Computational Linguistics. 2025: 5975-5986.

联系人: 王丹丹 (dandan@iscas.ac.cn)

## 研究背景



知识图谱在众多人工智能应用中发挥着关键作用，其中实体对齐 (Entity Alignment) 对于跨异构知识图谱整合重叠实体至关重要。

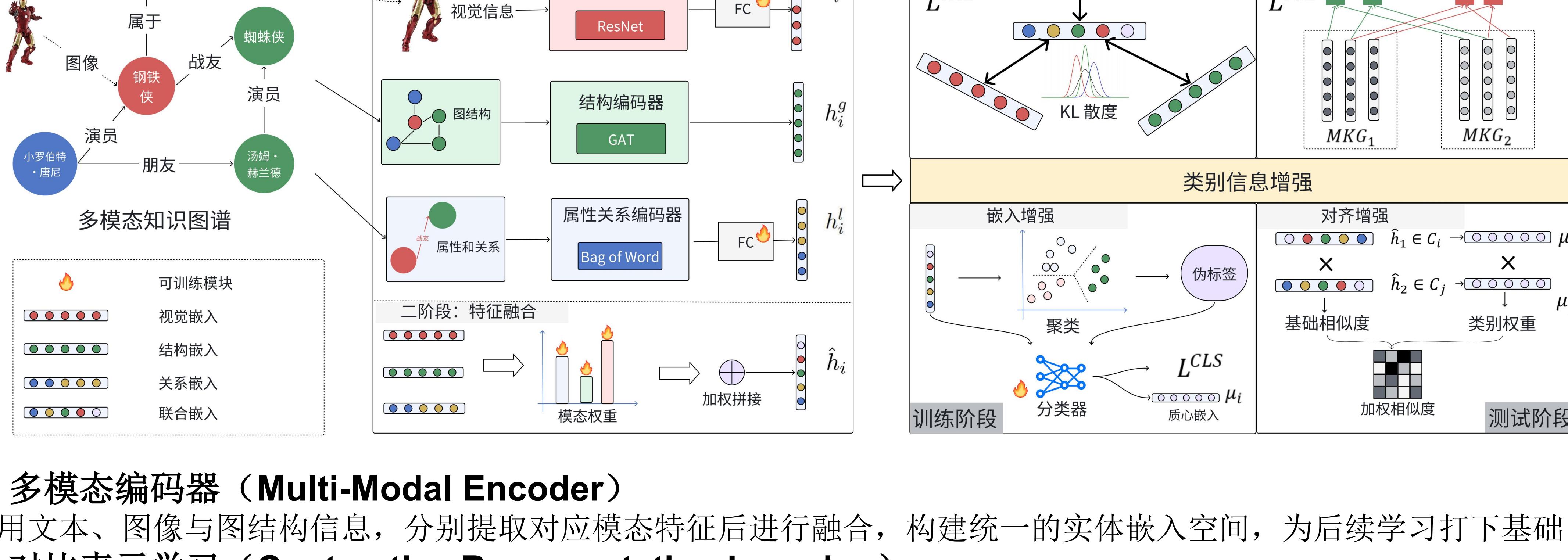
传统的EA方法通常依赖结构化数据和文本描述，而多模态实体对齐 (Multi-Modal Entity Alignment) 则利用多样化的模态以提供更丰富的语义信息。

然而，现有的MMEA方法主要关注显式的多模态特征相似性，忽视了隐含的类别关系，从而可能导致语义上不同但外观相似的实体被错误对齐。

### 研究贡献:

- ◆ 提出CateEA框架，引入实体的隐含类别信息以增强多模态实体表示的语义区分能力。
- ◆ 利用聚类生成伪类别标签，引入辅助分类任务，提升实体嵌入的判别性。
- ◆ 在测试阶段引入类别感知的相似度调整策略，使对齐结果更符合语义类别一致性。
- ◆ 在公开基准数据集MMKG和DBP15K上开展实证与消融实验，验证了方法的有效性与鲁棒性。

## 研究内容



### • 多模态编码器 (Multi-Modal Encoder)

利用文本、图像与图结构信息，分别提取对应模态特征后进行融合，构建统一的实体嵌入空间，为后续学习打下基础。

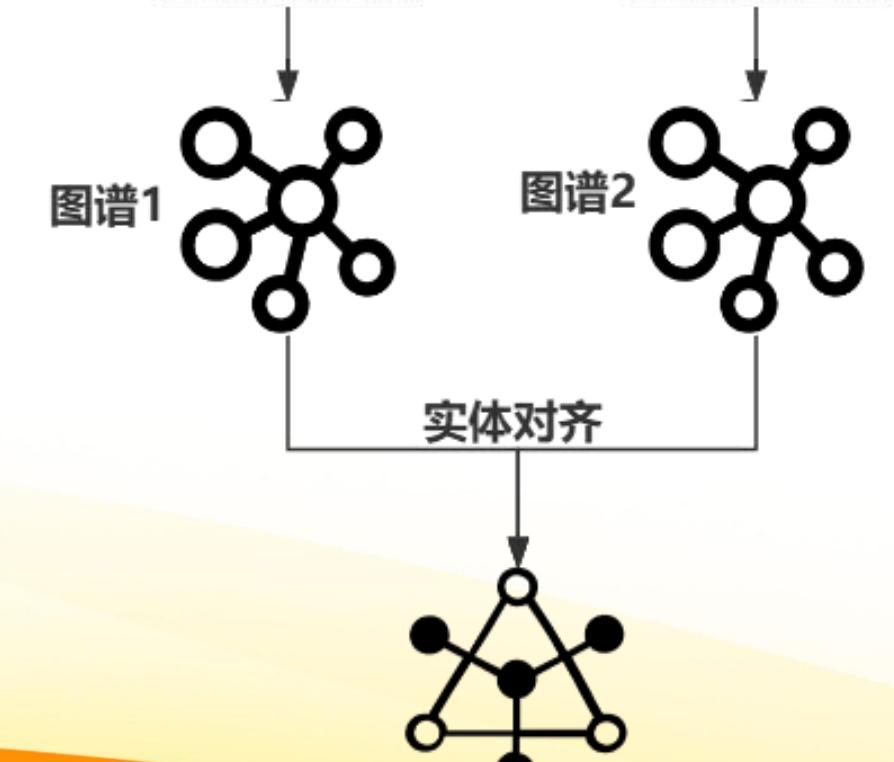
### • 对比表示学习 (Contrastive Representation Learning) :

采用多模态对比学习策略，最小化对齐实体对的距离、最大化非对齐实体对的距离，显著提升实体表示的对齐能力。

### • 类别信息增强 (Category Information Enhancement) :

- > 类别嵌入增强：基于实体嵌入进行聚类生成伪类别标签，引入辅助分类任务，引导模型学习具有类别判别性的表示。
- > 类别对齐增强：测试阶段使用训练好的分类器标注实体类别，并基于类别相似度调整实体相似度评分，提升语义一致实体的对齐准确率。

## 应用场景



随着知识图谱的发展，为了统一数据格式，消除知识冗余和冲突，提升知识图谱的质量和准确性，需要进行知识融合

实体对齐 (Entity Alignment) 是知识融合的主要手段。它指的是在不同的知识图谱中识别和匹配表示相同实体的节点。

## 实验结果

### 对比实验

种子集比例 %	方法	FB15K-DB15K		FB15K-YAGO15K		
		H@1	H@10	H@1	H@10	MRR
20%	MMEA	.265	.541	.357	.234	.317
	EVA	.199	.448	.283	.153	.224
	MSNEA	.114	.296	.175	.103	.249
	MCLEA	.295	.582	.393	.254	.484
	MEAformer	.417	.715	.518	.327	.595
	CateEA	.493	.759	.584	.402	.497
50%	提升	+7.6%	+4.4%	+6.6%	+7.5%	+8.0%
	MMEA	.417	.703	.512	.403	.465
	EVA	.334	.589	.422	.311	.534
	MSNEA	.288	.590	.388	.320	.589
	MCLEA	.555	.784	.637	.501	.705
	MEAformer	.619	.843	.694	.560	.778
80%	CateEA	.674	.874	.745	.608	.829
	提升	+5.5%	+3.1%	+4.7%	+4.8%	+5.1%
	MMEA	.590	.869	.685	.598	.839
	EVA	.484	.696	.563	.491	.662
	MSNEA	.518	.779	.613	.531	.778
	MCLEA	.735	.890	.790	.667	.824
提升	MEAformer	.765	.916	.820	.703	.873
	CateEA	.799	.933	.849	.742	.912
	提升	+3.4%	+1.7%	+2.9%	+3.9%	+3.9%

### 消融实验

