doi:10.12068/j.issn.1005－3026.2026.20260031

基于知识图谱增强与检索重排的归纳多模态知识图谱补全

艾笑天[1,2]，李星岐[1]，任 涛[1]，温晓玲[2]
（1.东北大学 软件学院，辽宁 沈阳 110819；2.航空工业沈阳飞机设计研究所，辽宁 沈阳 110034）

摘 要：针对归纳多模态知识图谱补全中测试阶段出现新兴实体且尾实体结构邻域缺失的问题，提出了1种基于知识图谱增强与检索重排的归纳多模态知识图谱补全方法．首先，利用视觉语言模型联合尾实体文本描述与图像信息，生成实体类型、描述提及和以尾实体为头实体的伪三元组，实现尾实体知识增强；其次，设计查询编码器与实体编码器，并通过统一跨模态对比学习实现文本-视觉语义对齐；再次，建立知识存储与实体存储，结合全局实体分布和语义邻居分布完成双路检索推理；最后，引入大语言模型对候选实体相关三元组进行过滤与重排．所构建框架在未见实体场景下兼顾了多模态表示对齐、先验知识记忆和证据驱动重排，为多模态知识图谱补全提供了一条可行的归纳推理路径．

关 键 词：多模态知识图谱补全；归纳知识图谱补全；跨模态对比学习；知识增强；检索重排
中图分类号：TP 182 **文献标志码**：A **文章编号**：1005-3026(2026)07-0132-07

Inductive Multimodal Knowledge Graph Completion Based on Knowledge Graph Enhancement and Retrieval Re-ranking

AI Xiao-tian[1,2], *LI Xing-qi*[1], *REN Tao*[1], *WEN Xiao-ling*[2]
(1. School of Software, Northeastern University, Shenyang 110819, China; 2. AVIC Shenyang Aircraft Design and Research Institute, Shenyang 110034, China. Corresponding author: LI Xing-qi, E-mail: chinaxingqili@163.com)

Abstract: For the problem that emerging entities appear in the testing phase, and tail entity structural neighborhoods are missing in inductive multimodal knowledge graph completion, an inductive multimodal knowledge graph completion method based on knowledge graph enhancement and retrieval re-ranking was proposed. First, a vision-language model was utilized to combine tail entity text descriptions and image information to generate entity types, descriptive mentions, and pseudo triples with the tail entity as the head entity, achieving tail entity knowledge enhancement. Second, a query encoder and an entity encoder were designed, and text-visual semantic alignment was achieved through unified cross-modal contrastive learning. Third, a knowledge store and an entity store were established, and dual-path retrieval reasoning was completed by combining global entity distribution and semantic neighbor distribution. Finally, a large language model was introduced to filter candidate entity-related triples and make re-ranking decisions. The constructed framework balances multimodal representation alignment, prior knowledge memory, and evidence-driven re-ranking under unseen entity scenarios and provides a feasible inductive reasoning path for multimodal knowledge graph completion.

Key words: multimodal knowledge graph completion; inductive knowledge graph completion; cross-modal contrastive learning; knowledge enhancement; retrieval re-ranking

收稿日期：2026-03-14
基金项目：国家自然科学基金资助项目(62276058)；中央高校基本科研业务费专项资金资助项目(N25GFZ011)．
作者简介：艾笑天(1994—)，男，辽宁沈阳人，东北大学博士研究生．
通信作者：李星岐，E-mail:chinaxingqili@163.com.

知识图谱(knowledge graph,KG)以三元组(头实体、关系、尾实体)刻画实体及其关系,是语义搜索、智能问答和推荐系统等应用的重要基础[1].然而,现实KG受构建成本和知识动态变化影响,普遍存在稀疏与不完整问题.知识图谱补全(knowledge graph completion,KGC)旨在基于已有三元组预测缺失链接,是提升KG完备性的关键技术[2].

早期KGC主要采用基于嵌入(embedding-based)的方法,如TransE,RotatE等[3-4].这类方法通过低维向量表示实体和关系,并利用评分函数判断三元组合理性,但依赖固定实体集合,难以泛化至未见实体[5].图神经网络(graph neural network,GNN)方法通过聚合邻域信息支持归纳式推理[6-7],但在稀疏和长尾知识图谱中,仍存在邻域不足、路径不稳定和新实体表示不充分等问题.

多模态知识图谱补全(multimodal knowledge graph completion,MKGC)融合文本、图像和属性等信息,以增强实体语义表示[8-9].多模态信息可在结构稀疏或邻居不足时提供补充线索,但现有方法仍存在语义对齐不足、模态融合不充分、噪声干扰和未见实体表示困难等问题.因此,如何学习并利用通用知识模式,而非简单记忆特定实体,是提升归纳式MKGC能力的关键.

近年来,大型语言模型(large language model,LLM)和视觉语言模型(vision-language model,VLM)展现出较强的语义理解能力、世界知识掌握能力和生成推理能力,为KG构建与补全提供了新思路[10-11].相关方法可通过提示工程(Prompting)或微调来生成候选答案或三元组[12].同时,对比学习能够提高表示空间的判别性[13];检索增强生成(retrieval-augmented generation,RAG)和开放域问答研究表明,外部知识检索有助于缓解幻觉并提升推理质量[14-15].

在此基础上,KGC研究逐渐由“直接生成答案”转向“结构增强–上下文增强–检索推理–候选重排”的协同范式.CATS利用类型约束、子图路径和邻域事实辅助LLM判断候选三元组合理性[16];KGR3通过检索、推理和重排提升长尾实体与复杂关系预测效果[17];ReranKGC采用粗粒度检索与细粒度重排,挖掘文本和视觉属性中的多模态语义[18].此外,FtG通过“先过滤、再生成”降低候选搜索难度和幻觉风险[19];SAT缓解自然语言空间与图结构空间的不一致[20];MusKGC利用多源知识增强和事实一致性验证提升开放世界KGC的可靠性[21].MgHiSal和Make your choice for MKGC则分别关注层次化语义对齐与模态选择融合问题[22-23].这些研究表明,归纳、多模态和开放世界KGC正逐步走向“知识增强–候选检索–语义推理–精细重排”的协同框架.

此外,StAR通过结构增强文本表示提升KGC效率[24];kNN-KGE利用近邻嵌入增强记忆化推理能力[25];StATIK联合结构与文本信息增强未见实体表示[26];CMR通过“对比学习–记忆增强”引入语义邻居检索,为归纳MKGC提供支持[27].但现有方法仍存在候选召回不足、通用知识模式利用不充分和检索结果语义判别不精细等问题.

针对上述问题,本文提出1种基于知识图谱增强与检索重排的归纳多模态知识图谱补全框架.首先,通过跨模态对比学习与知识图谱增强,学习查询与实体的统一表示空间,提升多模态语义对齐能力.其次,构建知识存储和实体存储机制,分别记忆通用知识模式与实体语义表示,并通过K-SN(K-surrounding neighbor)相似度检索和实体相似度检索实现候选实体粗召回.最后,引入基于大模型的精细化重排器,对候选列表进行过滤和重排序,以提高预测准确性与鲁棒性.

本文主要贡献如下:

1)提出1个融合知识增强、记忆、检索与重排的归纳推理框架,提升未见实体场景下的泛化能力;

2)设计1种尾实体多模态知识增强方法和1种双路检索机制,缓解新实体表示不足和候选召回困难问题;

3)引入大模型作为精细化重排器,对候选答案进行判别与排序,提高预测准确性、鲁棒性和可解释性;

4)在多个公开归纳多模态知识图谱基准数据集上进行实验验证,结果表明,本文方法优于现有基线模型,验证了所提框架的有效性.

1　网络模型

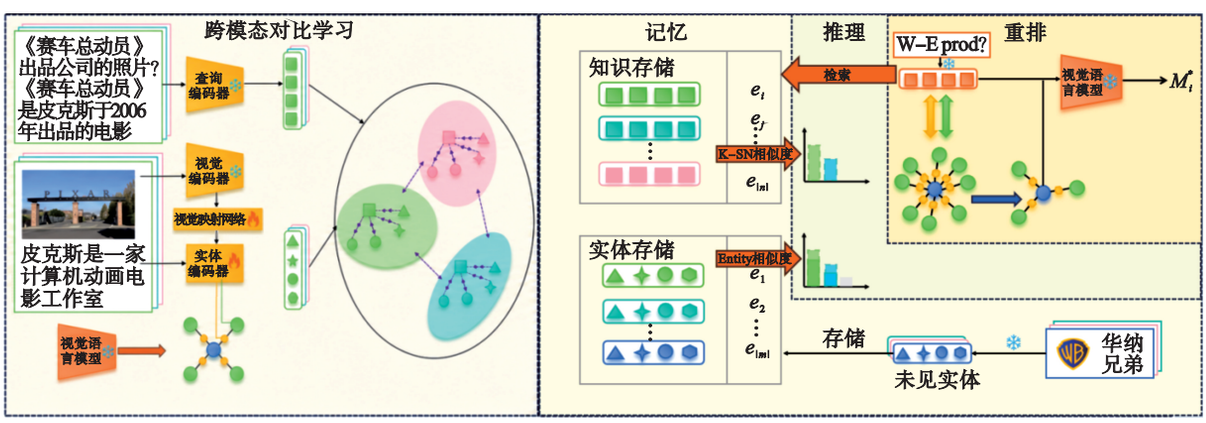
1.1　问题定义

在归纳多模态知识图谱补全(Inductive MKGC)任务中,给定1个知识图谱$\mathcal{G}=\{\mathcal{E},\mathcal{R},\mathcal{T}\}$,其中$\mathcal{E}$和$\mathcal{R}$分别表示实体和关系的集合,$\mathcal{T}=\{(h,r,t)\subseteq\mathcal{E}\times\mathcal{R}\times\mathcal{E}\}$表示三元组集合.对于每个实体$e\in\mathcal{E}$,伴有其文本描述$x_e^{\mathrm{t}}$与视觉图像$x_e^{\mathrm{v}}$.给定

一个查询 $q=(h,r,?)$,模型需要根据已知的头实体 h 和关系 r 推断缺失的尾实体 t. 与传统知识图谱补全任务不同,在归纳场景下,测试集包含训练阶段未见过的新兴实体,这对模型的跨模态表示学习与泛化能力提出了更高要求.

1.2 整体网络架构

本文提出的基于知识图谱增强与检索重排的归纳多模态知识图谱补全方法整体包括4个核心模块:跨模态对比学习与知识图谱增强、知识表示记忆、检索增强归纳推理和检索重排,整体网络结构如图1所示. 首先,模型对查询与实体进行统一编码,并结合尾实体的文本与图像信息生成伪三元组增强知识,提升尾实体表示能力;随后,构建知识存储与实体存储两类记忆结构,分别保存查询-实体对应关系和实体多模态表示;在推理阶段,基于增强后的查询表示进行双路检索,通过融合语义邻居信息与全局实体分布得到候选实体预测结果;最后,结合候选实体对应的筛选三元组信息,利用大语言模型进行重排,输出最终的尾实体预测结果.



注:W-E prod为WALL-E Production Company;M_t^*为最后的结果,即得到的尾部实体.

图1 整体网络结构

Fig. 1 Overall network architecture

1.3 VLM辅助的尾实体知识增强

由于尾实体本身缺乏已知的关联三元组,模型需根据尾实体 t 的多模态信息,主动推导潜在尾实体的关键属性. 这一过程旨在生成关于目标实体的预测性描述,给定尾实体 t 的文本描述 x_t^{t} 和视觉特征 x_t^{v},将其输入多模态大模型中,并通过预先设定的提示模板(Prompt)指导VLM执行以下任务:

1) 识别实体类型;

2) 生成描述提及;

3) 生成以该实体为头实体的伪三元组.

此过程可形式化为

$$e_{\mathrm{d}}=\mathcal{T}_{\mathrm{gen}}^{t}=\mathrm{VLM}\left(\mathrm{Prompt},x_t^{\mathrm{t}},x_t^{\mathrm{v}}\right). \tag{1}$$

其中:x_t^{t} 表示尾实体 t 的文本描述;x_t^{v} 表示尾实体 t 的视觉特征;$\mathcal{T}_{\mathrm{gen}}^{t}$ 表示由视觉语言模型针对尾实体 t 生成的结构化知识;e_{d} 表示融合生成结构化知识后的增强实体描述.

1.4 跨模态对比学习与实体表示对齐

为了在缺乏显式结构信息的条件下对新兴实体进行精准表征,模型必须构建1个统一的跨模态语义空间. 该空间旨在打破文本与视觉模态间的壁垒,将实体描述与视觉特征进行深度对齐. 本阶段沿用经典的双编码器结构,通过解耦查询端与实体端的建模,实现高效的检索与对齐.

1.4.1 查询编码器

对于三元组查询 $\boldsymbol{q}=(h,r,?)$,模型利用预训练语言模型(pre-trained language model,PLM)作为查询编码器 f_q. 为了使模型能够理解头实体 h 与关系 r 之间的潜在逻辑,查询输入 T_q 通过融合头实体名称、描述 h_{d} 以及关系定义 r_{d} 的模板构建:

$$T_q(h,r)=\text{“A photo of }h\text{'s }r_{\mathrm{d}}\text{? }h_{\mathrm{d}}\text{”}. \tag{2}$$

查询编码器提取模板输入在最后一层的隐藏状态,经均值池化并进行L2归一化后,得到查询嵌入向量 $\boldsymbol{q}$:

$$\boldsymbol{q}=\mathrm{L2Norm}\left(\mathrm{Mean}\left(f_q\left(T_q(h,r)\right)\right)\right). \tag{3}$$

其中:f_q 表示查询编码器;Mean表示均值池化操作,用于将多个token的向量表示聚合为1个整体向量;L2Norm表示L2归一化操作,用于将向量归一化到单位长度.

1.4.2 实体编码器

针对具有图像模态的实体,本文引入视觉映

射网络（visual mapping network，VMN）将视觉特征 v 投影至文本语义空间．若 v 为预训练视觉模型（ViT）提取的原始特征，则 VMN 通过多层感知机生成可训练的视觉前缀 v'：

$$v' = \mathrm{MLP}(v). \tag{4}$$

其中：MLP(·) 表示多层感知机，用于完成特征变换和维度映射．

实体输入序列 T_e 通过将视觉前缀与实体描述 e_{d} 拼接而成：

$$T_e(e_{\mathrm{d}}, v) = v' + e_{\mathrm{d}}. \tag{5}$$

该公式表示将视觉模态特征与文本描述信息进行组合，从而形成统一的多模态实体输入．

随后，利用实体编码器 f_e 对上述序列进行特征提取，生成归一化后的实体嵌入 e：

$$e = \mathrm{L2Norm}\Big(\mathrm{Mean}\big(f_e(T_e(e_{\mathrm{d}}, v))\big)\Big). \tag{6}$$

1.4.3　统一跨模态对比损失

为缓解多模态矛盾（如语义相关但视觉差异巨大的头尾实体）带来的干扰，并增强表示的判别性，模型采用统一跨模态对比损失进行训练．该损失在联合向量空间中拉近真实的查询–实体对，并推远批次内的负样本对．

$$L_{\mathrm{cl}} = -\sum_{i=1}^{B} \log \left[\frac{\exp(\mathrm{sim}(\boldsymbol{q}_i, e_i)/\tau)}{\sum_{j=1}^{B} \exp(\mathrm{sim}(\boldsymbol{q}_i, e_j)/\tau)} \right]. \tag{7}$$

其中：L_{cl} 表示对比学习损失函数；B 为批次大小；sim(·) 表示余弦相似度；τ 为温度超参数；$\boldsymbol{q}_i$ 表示第 i 个查询向量；e_i 表示与 $\boldsymbol{q}_i$ 正确匹配的实体标签，即正样本；e_j 表示批次中的候选实体标签，当 $j \neq i$ 时可视为负样本．该机制确保了查询编码器与实体编码器能够在同一度量空间内实现精准的语义对齐．

1.5　知识表示记忆与双路检索机制

1.5.1　知识表示记忆

本文构建了两类持久化存储结构以保留训练阶段的先验知识：知识存储（knowledge store，KS）与实体存储（entity store，ES）．

知识存储（KS）：记录训练集中的查询及其对应的目标实体，定义为

$$\mathcal{M}_{\mathrm{KS}} = \left\{ \left(\boldsymbol{h}_{q_i}, e_i \right) \right\}_{i=1}^{N_{\mathrm{train}}}. \tag{8}$$

其中：$\boldsymbol{h}_{q_i}$ 为训练查询 $\boldsymbol{q}_i$ 的嵌入向量；$\mathcal{M}_{\mathrm{KS}}$ 为 KS 集合；N_{train} 为训练集样本总数．

实体存储（ES）：记录所有已知实体的表征，定义为

$$\mathcal{M}_{\mathrm{ES}} = \left\{ \left(\boldsymbol{h}_{e_j}, e_j \right) \right\}_{j=1}^{|\mathcal{E}|}. \tag{9}$$

其中：$\boldsymbol{h}_{e_j}$ 表示实体的多模态融合特征；$\mathcal{M}_{\mathrm{ES}}$ 为 ES 集合．

1.5.2　双路检索机制

在推理阶段，模型基于增强后的查询表示 $\boldsymbol{h}_{\mathrm{q_{aug}}}$ 执行双路检索，通过结合全局信息与局部语义邻居实现归纳推理．

1）第一路：全局实体分布预测．计算查询与 ES 中所有实体的相似度，并通过 Softmax 函数将其转化为全局概率分布：

$$p_{\mathrm{ES}}\left(e|\boldsymbol{h}_{\mathrm{q_{aug}}}\right) = \frac{\exp(\mathrm{sim}(\boldsymbol{h}_{\mathrm{q_{aug}}}, \boldsymbol{h}_e)/\tau)}{\sum_{e' \in \mathcal{E}} \exp(\mathrm{sim}(\boldsymbol{h}_{\mathrm{q_{aug}}}, \boldsymbol{h}_{e'})/\tau)}. \tag{10}$$

其中：$p_{\mathrm{ES}}\left(e|\boldsymbol{h}_{\mathrm{q_{aug}}}\right)$ 表示实体 e 在全实体空间中的预测概率；$\boldsymbol{h}_{\mathrm{q_{aug}}}$ 表示增强后的查询向量；$\boldsymbol{h}_e$ 表示候选实体 e 的向量表示；$\mathcal{E}$ 表示全部候选实体集合．

2）第二路：语义邻居分布预测．在 KS 中检索与当前查询最接近的 k 个邻居集合 $\mathcal{N}_k(\boldsymbol{q})$，并根据邻居的相似度权重构造分布：

$$p_{\mathrm{KS}}\left(e|\boldsymbol{h}_{\mathrm{q_{aug}}}\right) = \sum_{\left(\boldsymbol{h}_{q_i}, e_i\right) \in \mathcal{N}_k(\boldsymbol{q})} \mathbb{I}(e = e_i) \cdot \frac{\exp(\mathrm{sim}(\boldsymbol{h}_{\mathrm{q_{aug}}}, \boldsymbol{h}_{q_i})/\tau)}{\sum_{\left(\boldsymbol{h}_{q_j}, e_j\right) \in \mathcal{N}_k(\boldsymbol{q})} \exp(\mathrm{sim}(\boldsymbol{h}_{\mathrm{q_{aug}}}, \boldsymbol{h}_{q_j})/\tau)}. \tag{11}$$

其中：$p_{\mathrm{KS}}\left(e|\boldsymbol{h}_{\mathrm{q_{aug}}}\right)$ 表示基于语义邻居得到的预测概率；$\mathcal{N}_k(\boldsymbol{q})$ 表示与当前查询 $\boldsymbol{q}$ 最相近的 k 个邻居集合；$\mathbb{I}(\cdot)$ 表示指示函数，当条件成立时取 1，否则取 0．

3）双路融合．引入线性插值系数 $\lambda \in [0, 1]$，将上述 2 个分布进行融合，得到基础预测概率 p_{base}：

$$p_{\mathrm{base}}(e|\boldsymbol{q}) = (1-\lambda)\, p_{\mathrm{ES}}(e|\boldsymbol{q}) + \lambda p_{\mathrm{KS}}(e|\boldsymbol{q}). \tag{12}$$

1.6　检索重排阶段

利用 LLM 强大的语境理解与逻辑推理能力，从候选集 $\mathcal{C}$ 中辨析最佳匹配实体．本文设计了 1 种逐步提示（step-by-step prompting）策略，将查询提及的文本信息、三元组背景以及每个候选实体的详细三元组信息输入模型．

重排过程遵循以下 2 个逻辑阶段：

1）候选知识过滤．知识库中每个候选实体 $e_m \in \mathcal{C}$ 可能关联数以万计的三元组，直接输入会引入冗余噪声．为确保 LLM 能够专注于最具判别性的信息并降低计算开销，本文为每个候选实体 e_m 提供了过滤后的三元组集 $\mathcal{T}_{e_m}^{*}$．这种通过知识过

滤与结构化知识锚定的方式,使得预测过程不仅依赖于文本的浅层相似性,更建立在深层的结构化逻辑基础之上.

通过计算查询向量$\boldsymbol{h}_{q_{aug}}$与实体e_m相关三元组集合$\mathcal{T}_{e_m}$的余弦相似度,仅保留Top-L个最相关的核心三元组,构建精简知识集$\mathcal{T}_{e_m}^*$:

$$\mathcal{T}_{e_m}^*=\left\{\tau\in\mathcal{T}_{e_m}\middle|\operatorname{rank}\left(\operatorname{sim}\left(\boldsymbol{h}_{q_{aug}},\boldsymbol{h}_\tau\right)\right)\leqslant L\right\}. \quad (13)$$

其中:$\boldsymbol{h}_\tau$表示三元组τ的语义表示向量,用于衡量该三元组与增强查询表示$\boldsymbol{h}_{q_{aug}}$之间的相关性.

2) LLM决策.基于选定的支持三元组证据,评估各候选实体的逻辑符合度,并确定最合适的实体e^*.

形式上,最终的实体选择过程定义如下:

$$e^*=\mathrm{LLM}\left(P_{\mathrm{rerank}}\left(q,\left\{\left(e,\mathcal{T}_e^{(\mathrm{filt})}\right)\right\}_{e\in C}\right)\right). \quad (14)$$

其中:P_{rerank}为引导LLM根据文本和三元组信息确定最相关实体的提示词模板;q为原始查询信息;$\mathcal{T}_e^{(\mathrm{filt})}$为实体$e$过滤后的最终支撑三元组证据集合.

2 实验结果与分析

2.1 数据集与归纳划分

本文在2个权威的多模态知识图谱数据集FB15k-237-ind和WN18RR-ind上验证模型性能.为了模拟真实的归纳推理场景(inductive setting),本文遵循前人研究中的划分方法,确保测试集中的实体在训练阶段是不可见的.

FB15k-237(Inductive):源自Freebase,包含了丰富的现实世界实体关系.归纳版本通过移除训练集中出现的实体,构建了1个仅包含未见实体及其关联三元组的测试集,对模型捕捉多模态通用语义的能力要求极高.

WN18RR(Inductive):源自WordNet,主要包含层级和词义关系(如上义词、下义词等).相比于FB15k-237,其关系类型更少但语义更抽象.因此,在归纳场景下,模型依赖VLM生成的描述和三元组来弥补缺失的层级结构.

2.2 基线模型

为保证对比的公平性,本文选取3种基于预训练语言模型的知识图谱补全方法作为基线,分别为StAR[24],kNN-KGE[25]和StATIK[26].其中,StAR通过结构增强文本表示实现高效知识图谱补全;kNN-KGE通过最近邻记忆增强模型对新兴实体的推理能力;StATIK融合文本语义信息与图结构信息,以提升归纳场景下的补全效果.此外,本文进一步引入CMR(Base)[27]作为强对比方法,其中query encoder和entity encoder均采用BERT_base配置.

2.3 评价指标

本文采用链路预测任务中的标准评价指标:

1) MRR(mean reciprocal rank),正确实体的平均倒数排名;

2) Hits@N N=1,3,10,正确实体排在前N名的比例.

所有实验结果均在Filtered设置下进行,即在排序时剔除除目标三元组外已存在于知识库中的候选三元组,以确保评估的客观性.

2.4 实验结果与分析

为了全面评估所提方法的有效性与泛化能力,本文在FB15k-237-ind与WN18RR-ind两个具有代表性的基准数据集上进行了对比实验.该方法在所有评价指标上均刷新了现有纪录,特别是在极具挑战性的归纳场景下,其展现出了相较于基准模型CMR更为稳健的性能增益,充分验证了多模态知识注入与大模型重排机制在处理复杂知识图谱任务中的协同效应.

从表1可以看出,所提方法在所有评价指标上均达到了最优(state of the art,SOTA)性能.

表1 主实验结果

Table 1 Main experimental results

模型	FB15k-237-ind				WN18RR-ind			
	MRR	Hits@1	Hits@3	Hits@10	MRR	Hits@1	Hits@3	Hits@10
StAR	0.163	0.092	0.176	0.309	0.321	0.192	0.381	0.576
kNN-KGE	0.198	0.146	0.214	0.293	0.294	0.223	0.320	0.431
StATIK	0.224	0.143	0.248	0.381	0.516	0.425	0.558	0.690
CMR (Base)	0.371	0.269	0.418	0.567	0.631	0.577	0.644	0.788
Ours	**0.384**	**0.303**	**0.425**	**0.590**	**0.648**	**0.612**	**0.658**	**0.805**

注:黑体表示结果最佳.

在FB15k-237-ind数据集上,相比于强基准模型CMR,本文方法实现了全面的性能提升:MRR从0.371增长至0.384,Hits@10从0.567增长至0.590.这种显著的改进主要归功于以下两点.

多模态知识补全结构缺失:在归纳场景下,大量不可见的尾实体由于缺乏拓扑连接,传统模型容易产生误判.本文方法通过引入视觉语言模型(VLM)的多模态知识并构建伪三元组进行增强,有效弥补了结构缺失,将 Hits@3从0.418提升至0.425.

LLM精细化重排对抗语义噪声:针对初筛阶段易受复杂语义干扰的问题,本文方法利用LLM进行重排.这极大地提升了预测首位的精确度,使得Hits@1直接由0.269跃升至0.303,展现了极强的判别推理能力.

在WN18RR-ind数据集上,本文方法同样展现了稳健的泛化能力.MRR从0.631提升至0.648,Hits@1则由0.577显著增至0.612.这一结果进一步验证了本研究所提出的多模态增强与重排框架在不同领域知识图谱任务中的普适性与优越性.

2.5　消融实验

为了深入探究模型各核心组件的具体贡献,本文在FB15k-237-ind数据集上进行了详细的消融实验,实验结果如表2所示.

表2　消融实验结果

Table 2　Ablation experiment results

变体模型	FB15k-237-ind			
	MRR	Hits@1	Hits@3	Hits@10
Ours(完整模型)	0.384	0.303	0.425	0.590
w/o VLM knowledge(移除VLM知识增强)	0.372	0.278	0.402	0.569
w/o LLM reranking(移除LLM重排阶段)	0.375	0.272	0.408	0.584
w/o KS retrieval(仅使用ES全局检索)	0.378	0.291	0.415	0.576
w/o triple filtering(移除重排时的三元组过滤)	0.380	0.289	0.420	0.588

分析表2可以得出以下结论:

VLM增强的有效性(w/o VLM knowledge).当移除由VLM生成的尾实体结构化三元组和多模态描述后,MRR从0.384下降至0.372,Hits@10出现了最明显的跌落(降至0.569).这定量地证明了对于长尾和孤立实体,利用大视觉语言模型进行模态理解和知识补充是至关重要的.VLM为缺乏拓扑结构的实体提供了关键的语义背景,为后续的对比学习提供了更坚实的特征基础.

检索重排的有效性(w/o LLM reranking).移除大语言模型重排模块(即退化为纯KNN检索机制)导致Hits@1指标由0.303显著下降至0.272.这说明在候选集语义较为接近的情况下,单纯依靠向量空间距离难以实现极高精度的排序.而LLM凭借其强大的上下文推理能力,能够有效识别隐含的语义关联,纠正检索阶段带来的偏差,从而确保首位预测的准确性.

三元组过滤的有效性(w/o triple filtering).实验观察到,如果在送入LLM之前不对背景三元组进行相关性过滤,Hits@3性能会出现小幅下滑(由0.425降至0.420).这证明了过多的无关知识会构成“上下文噪声”,从而干扰LLM的判断逻辑.该结果印证了本研究采取“先过滤再重排”策略的必要性,即通过精炼知识输入来提升推理质量.

知识存储器的有效性(w/o KS retrieval).移除KS历史查询检索后,模型失去了对语义邻居经验的“记忆”能力,MRR降至0.378.性能的全面退化再次印证了局部经验检索(KS)与全局表示检索(ES)双路策略的正确性,证明了历史经验对处理复杂查询具有不可替代的支撑作用.

3　结　　论

本文提出了1种结合视觉语言模型(VLM)知识注入与大语言模型细粒度重排的归纳式多模态知识图谱补全框架.在FB15k-237-ind和WN18RR-ind数据集上的实验结果表明,该框架全面达到了SOTA性能.在FB15k-237-ind任务中,所提方法将MRR从0.371提升至0.384(提升3.5%),Hits@1更是由0.269跨越式增长至0.303,实现了12.6%的性能跃升,验证了其在精准推理方面的卓越能力.消融实验进一步证实,VLM知识增强通过补全实体缺失结构,有效支撑了广度检索(Hits@10提升3.7%至0.590);而带有前置过滤的LLM重排机制,则通过纠正KNN检索的

度量偏差,确保了预测精度的显著提高.本研究证明了“补齐-检索-重排”层级化策略的稳健性,为解决知识图谱中不可见实体的推理难题提供了高效、鲁棒的新范式.

参考文献:

[1] Ji S X, Pan S R, Cambria E, et al. A survey on knowledge graphs: representation, acquisition, and applications [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2022, 33(2): 494-514.

[2] Pan S R, Luo L H, Wang Y F, et al. Unifying large language models and knowledge graphs: a roadmap [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2024, 36(7): 3580-3599.

[3] Bordes A, Usunier N, Garcia-Durán A, et al. Translating embeddings for modeling multi-relational data [C]// Proceedings of the 27th International Conference on Neural Information Processing Systems (NIPS).South Lake Tahoe: Curran Associates Inc.,2013: 2787-2795.

[4] Sun Z Q, Deng Z H, Nie J Y, et al. RotatE: knowledge graph embedding by relational rotation in complex space [EB/OL]. (2019-02-19)[2026-01-05]. https://arxiv.org/abs/1902.10197.

[5] Teru K K, Denis E G, Hamilton W L. Inductive relation prediction by subgraph reasoning [C]//Proceedings of the 37th International Conference on Machine Learning. Online: PMLR,2020: 9448-9457.

[6] Schlichtkrull M, Kipf T N, Bloem P, et al. Modeling relational data with graph convolutional networks [C]//The Semantic Web. Cham: Springer, 2018: 593-607.

[7] Vashishth S, Sanyal S, Nitin V, et al. Composition-based multi-relational graph convolutional networks [EB/OL]. (2019-11-08)[2026-01-05]. https://arxiv.org/abs/1911.03082.

[8] Liu Y, Li H, Garcia-Durán A, et al. MMKG: multi-modal knowledge graphs [C]//The Semantic Web. Cham: Springer, 2019: 459-474.

[9] Zhu X R, Li Z X, Wang X D, et al. Multi-modal knowledge graph construction and application: a survey [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2024, 36(2): 715-735.

[10] Yao L, Mao C S, Luo Y. KG-BERT: BERT for knowledge graph completion [EB/OL]. (2019-09-11) [2026-01-05]. https://arxiv.org/abs/1909.03193.

[11] Wang X Z , Gao T Y, Zhu Z C, et al. KEPLER: a unified model for knowledge embedding and pre-trained language representation [J]. *Transactions of the Association for Computational Linguistics*, 2021,9: 176-194.

[12] Jiang Z B, Xu F F, Araki J, et al. How can we know what language models know? [J]. *Transactions of the Association for Computational Linguistics*, 2020, 8: 423-438.

[13] Gao T Y, Yao X C, Chen D Q. SimCSE: simple contrastive learning of sentence embeddings [C]// Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Online: Association for Computational Linguistics,2021: 6894-6910.

[14] Lewis P, Perez E, Piktus A, et al. Retrieval-augmented generation for knowledge-intensive NLP tasks [C]// Proceedings of the 34th International Conference on Neural Information Processing Systems. Online: Curran Associates Inc.,2020: 9459-9474.

[15] Izacard G, Grave E. Leveraging passage retrieval with generative models for open domain question answering [C]// Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics. Online: Association for Computational Linguistics, 2021: 874-880.

[16] Li M Z, Yang C H, Xu C J, et al. Context-aware inductive knowledge graph completion with latent type constraints and subgraph reasoning [J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2025, 39 (11): 12102-12111.

[17] Li M Z, Yang C H, Xu C J, et al. Retrieval, reasoning, re-ranking: a context-enriched framework for knowledge graph completion [C]//Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers). Albuquerque, 2025: 4349-4363.

[18] Gao M, Xie Y T, Chen W, et al. ReranKGC: a cooperative retrieve-and-rerank framework for multi-modal knowledge graph completion [J]. *Neural Networks*, 2025, 188: 107467.

[19] Liu B, Zhang J H, Lin F Q, et al. Filter-then-generate: large language models with structure-text adapter for knowledge graph completion [C]//Proceedings of the 31st International Conference on Computational Linguistics. Abu Dhabi,2025: 11181-11195.

[20] Liu Y, Cao Y N, Lin X X, et al. Enhancing large language model for knowledge graph completion via structure-aware alignment-tuning [C]//Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing. Suzhou, 2025: 20970-20984.

[21] Song X, Liu H Y, Wang H Y, et al. MusKGC: a flexible multi-source knowledge enhancement framework for open-world knowledge graph completion [C]//Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing. Suzhou, 2025: 10031-10049.

[22] Chen J, Zhang W Y, Zhao S, et al. MgHiSal: MLLM-guided hierarchical semantic alignment for multimodal knowledge graph completion [J]. *Knowledge-Based Systems*, 2025, 330: 114552.

[23] Xue Y F, Ren S Y, Cai W Q, et al. Make your choice for multimodal knowledge graph completion [J]. *Knowledge-Based Systems*, 2025, 323: 113750.

[24] Wang B, Shen T, Long G D, et al. Structure-augmented text representation learning for efficient knowledge graph completion [C]//Proceedings of the Web Conference 2021. Ljubljana, 2021: 1737-1748.

[25] Wang P, Xie X, Wang X H, et al. Reasoning through memorization: nearest neighbor knowledge graph embeddings [C]//Natural Language Processing and Chinese Computing. Cham: Springer, 2023: 111-122.

[26] Markowitz E, Balasubramanian K, Mirtaheri M, et al. StATIK: structure and text for inductive knowledge graph completion [C]//Findings of the Association for Computational Linguistics: NAACL 2022. Seattle, 2022: 604-615.

[27] Zhao Y, Zhang Y, Zhou B H, et al. Contrast then memorize: semantic neighbor retrieval-enhanced inductive multimodal knowledge graph completion [C]//Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval. Washington DC ,2024: 102-111.