计 算 机 研 究 与 发 展
Journal of Computer Research and Development
https://crad.ict.ac.cn
63(9): 2202−2214, 2026

多视图增强图注意力网络的多模态知识图谱补全

陈 伟 庄福振
(北京航空航天大学人工智能学院 北京 100191)
(chenwei23@buaa.edu.cn)

Multi-View Enhanced Graph Attention Network for Multimodal Knowledge Graph Completion

Chen Wei and Zhuang Fuzhen
(*School of Artificial Intelligence*, *Beihang University*, *Beijing* 100191)

Abstract In recent years, multimodal knowledge graphs have introduced auxiliary information such as entity textual descriptions and images, providing richer semantic sources for entity representation learning in knowledge graph completion. However, most existing multimodal knowledge graph completion methods focus on the alignment or fusion of features from different modalities, typically modeling such information as static external attributes of entities, while insufficiently exploiting the structural context formed by local topology. Since entity semantics are derived not only from their own multimodal attributes but also influenced by neighborhood relations, the lack of relation-constrained structural propagation may cause multimodal features to remain at the level of static semantics, making it difficult to form structured representations with reasoning capability. Particularly in scenarios where modal information contains strong noise, directly relying on initial multimodal features or prematurely fusing heterogeneous modalities may weaken the stability and discriminability of entity representations, thereby limiting the ability to predict complex relation patterns and potential facts. To address these issues, we propose a multimodal knowledge graph completion method based on a multi-view enhanced graph attention network, namely MEGAT. The proposed method models structural, visual, and textual information as three independent views under a shared graph topology, and constructs a relation-aware graph attention encoder for each view. In this way, relation-constrained neighborhood propagation and structural semantic modeling can be performed separately in different modal spaces. Subsequently, a gated late-fusion mechanism is employed to dynamically integrate multi-view representations, effectively alleviating the semantic interference caused by the direct fusion of heterogeneous modalities. Experimental results show that the proposed method enhances the ability of entity representations to capture graph structural context and complex relation patterns, achieving competitive completion performance on multiple public datasets.

Key words multimodal knowledge graph; knowledge enhancement; relation-aware graph attention network; multi-view representation learning; multimodal fusion

摘 要 近年来,多模态知识图谱通过引入实体文本描述、图像等辅助信息,为知识图谱补全中的实体表

收稿日期: 2026-06-05; **修回日期**: 2026-07-17
基金项目: 国家重点研发计划(2024YFF0729003); 国家自然科学基金项目(62176014, 62276015, 62206266); 中央高校基本科研业务费专项资金; 北京航空航天大学博士研究生卓越学术基金
This work was supported by the National Key Research and Development Program of China (2024YFF0729003), the National Natural Science Foundation of China (62176014, 62276015, 62206266), the Fundamental Research Funds for the Central Universities, and the Academic Excellence Foundation of BUAA for PhD Students.
通信作者: 庄福振(zhuangfuzhen@buaa.edu.cn)

示学习提供了更加丰富的语义来源。然而，现有多模态知识图谱补全方法大多侧重于不同模态特征的对齐或融合，通常将这些信息作为实体的静态外部属性进行建模，而对局部拓扑共同构成的结构上下文利用不足。由于实体语义不仅来源于自身的多模态属性，也受到其邻域关系的影响，缺乏关系约束下的结构传播容易使多模态特征停留在静态语义层面，难以形成具有推理能力的结构化表示。尤其在模态信息噪声较强的场景下，直接依赖初始多模态特征或过早融合异构模态，容易削弱实体表示的稳定性和判别性，从而限制对复杂关系模式和潜在事实的预测能力。为解决上述问题，提出一种基于多视图增强图注意力网络的多模态知识图谱补全方法 MEGAT（multi-view enhanced graph attention network）。该方法将结构、视觉和文本信息建模为共享图拓扑下的 3 个独立视图，并为每个视图构建关系感知图注意力编码器，以在不同模态空间中分别执行关系约束下的邻域传播和结构语义建模。随后，采用门控后融合机制动态整合多视图表示，有效缓解了直接融合异构模态带来的语义干扰。实验结果显示，所提出的方法能够增强实体表示对图结构上下文和复杂关系模式的感知能力，在多个公开数据集上实现竞争性补全性能。

关键词 多模态知识图谱；知识增强；关系感知图注意力网络；多视图表示学习；多模态融合

中图法分类号 TP311.13；TP309

DOI: 10.7544/issn1000-1239.202660438 **CSTR:** 32373.14.issn1000-1239.202660438

知识图谱以结构化三元组的形式刻画现实世界中实体及其关系，已广泛应用于智能问答、推荐系统、语义搜索、信息抽取和知识推理等任务[1]。尽管大规模知识图谱已经积累了海量事实，但由于开放世界知识持续演化、人工标注成本较高以及自动抽取过程存在噪声，现有知识图谱普遍面临事实缺失、关系长尾和实体连接稀疏等问题。因此，如何根据已有事实推断潜在缺失关系，即知识图谱补全，成为提升知识图谱完整性和下游任务性能的重要研究方向[2]。

早期知识图谱补全方法主要围绕结构化三元组进行表示学习，其核心思想是将实体和关系映射到低维连续向量空间，并通过特定评分函数衡量候选三元组的合理性。根据建模思路不同，这类方法通常可以分为基于几何变换的距离模型[3]和基于语义交互的匹配模型[4-6]。前者强调关系在实体表示空间中的变换作用，后者则更关注实体与关系之间的潜在关联模式。此类方法有效推动了知识补全技术的发展，但其主要依赖符号化三元组学习，实体表示易受限于图中既有连接关系，难以充分刻画实体丰富的语义属性。尤其在实体连接稀疏或语义相似实体较多的场景下，仅依赖结构三元组往往难以获得具有充分区分性的表示。

为缓解嵌入方法对局部结构上下文利用不足的问题，基于图神经网络的补全方法逐渐受到关注[7-8]。其通过邻域聚合和消息传递机制，将邻域结构与关系信息纳入表示学习，使实体表示不再仅由自身嵌入决定，而能够反映其在图结构中的上下文语义[9-10]。因此，图神经网络方法在建模结构依赖和关系传播方面具有优势。然而，这些方法主要围绕单一结构模态展开，对实体语义描述与视觉信息等外部模态利用不足，难以充分发挥多模态知识图谱中异构信息的互补作用。

随着多模态知识图谱的发展，研究者开始尝试将文本、图像等辅助模态引入知识图谱补全过程，以弥补结构三元组在语义表达方面的不足[11-13]。相比单纯依赖结构信息的补全方法，多模态方法能够从实体描述和视觉外观中获取补充语义，从而提升实体表示的丰富性和判别能力。已有研究表明，多模态信息对于长尾以及结构连接不足的实体尤其重要，能够在一定程度上缓解传统结构表示学习中的语义稀疏问题[14-15]。

尽管多模态信息为知识图谱补全提供了新的语义来源，现有方法在多模态融合与结构建模之间仍存在一定不足。一方面，它们主要关注不同模态特征的对齐或融合[16-17]，对多模态信息如何在关系约束下的图结构中传播考虑不足。实体语义不仅来源于自身的文本和图像信息，也受到其邻域关系与局部结构的影响。若缺乏结构传播建模，多模态特征容易停留在静态属性层面，难以形成具有推理能力的结构化表示。另一方面，不同关系在邻域信息传播中的作用并不相同。同一邻居实体在不同关系语义下可能传递差异化信息，简单的邻域聚合难以有效刻画这种关系条件下的语义影响。因此，多模态补全不仅需要融合多源模态信息，还需要在传播过程中显式建模关系对邻域消息的调制作用。此外，不同模态具有不同的特征分布与噪声特性，若过早进行粗粒度融合，容易造成异构语义混杂，削弱实体表

示的稳定性和判别性。

因此，本文提出一种基于多视图增强图注意力网络的多模态知识图谱补全方法 MEGAT(multi-view enhanced graph attention network)。该方法将结构、视觉和文本信息分别视为 3 个模态视图，并在共享知识图谱拓扑的基础上，为每个视图构建独立的关系感知图注意力编码器。不同于直接拼接多模态特征的早期融合方法，本文方法首先在各模态空间内部进行结构传播，使每个模态都能够学习到关系约束下的邻域语义；随后再采用门控后融合机制对不同视图表示进行动态整合。通过这种“模态分离编码—关系感知传播—门控后融合”的建模方式，模型既能够保留不同模态的特异性信息，又能够增强实体表示对图结构上下文和复杂关系模式的感知能力。

本文的主要贡献包括：1)提出一种面向多模态知识图谱补全的多视图结构编码框架。将结构、图像和文本模态分别建模为独立视图，在共享图拓扑上进行模态分离的关系感知传播，缓解了异构模态早期融合带来的语义干扰问题。2)设计关系感知图注意力编码器。通过引入平移差分、逐元素乘法、旋转组合和循环相关等多种关系组合算子，模型能够从不同交互角度构造关系感知消息；同时基于三元组的注意力机制能够自适应分配不同邻域关系的重要性，从而提升结构语义建模能力。3)构建门控式多模态后融合机制。根据实体在不同模态视图下的表示自适应学习融合权重，使模型在保留模态特异性的同时获得统一的多模态实体表示。(4)在多个公开数据集上取得了较优的补全性能，消融实验进一步验证了多视图结构传播和门控后融合的有效性。

1 相关工作

1.1 基于平移的知识图谱补全方法

基于平移的知识图谱补全方法是知识图谱表示学习中较具代表性的一类方法，其核心是将关系视为头实体到尾实体的几何变换，并通过距离函数判断三元组的合理性。典型模型 TransE[3] 将关系建模为向量平移，即对于真实三元组，头实体与关系的嵌入之和应尽可能接近尾实体嵌入。凭借结构简单、参数量少和训练效率高等优势，TransE 成为后续研究的重要基础。然而，该模型难以有效处理一对多、多对多等复杂映射关系及多样化的关系模式。为此，研究者围绕平移框架提出了多种改进方法[18-20]，比如：TransH 将实体投影到关系对应的超平面中，以增强对复杂关系的刻画能力；TransR 则为不同关系构建特定的表示空间，使实体在不同关系下具有差异化表示。

随着研究的不断深入，基于平移的方法开始从简单线性变换扩展到更复杂的几何建模形式。例如，RotatE[21] 将关系建模为复数空间中的旋转操作，能够较好地刻画对称、反对称、反演等多种关系模式；HAKE 通过极坐标表示将实体的层次结构与类别差异进行分离建模[22]；PairRE[23]、TripleRE[24] 等方法则进一步增强了模型对复杂关系模式和映射属性的表达能力。总体而言，基于平移的方法推动了知识图谱补全技术的发展，并在关系模式建模方面取得了显著进展。但此类方法主要依赖结构化三元组进行学习，对上下文语义及多模态信息的利用仍较为有限。

1.2 基于张量分解的知识图谱补全方法

基于张量分解的知识图谱补全方法又称语义匹配方法，其核心思想是将知识图谱表示为三阶张量，通过张量分解学习实体和关系的低维表示，从而挖掘二者之间潜在的语义关联[4-6]。这类方法通常采用双线性或多线性评分函数衡量头实体、关系与尾实体的匹配程度，在建模三者的语义交互方面具有较强优势。RESCAL[25] 是该方向的经典模型，它为每种关系学习独立的关系矩阵，能够刻画实体之间复杂的交互关系，但也带来了较大的参数规模和计算开销。为提升建模效率，DistMult[4] 将关系矩阵简化为对角矩阵，在减少参数量的同时提升了训练速度，但受评分函数对称性的限制，难以有效建模非对称关系。为克服这一问题，ComplEx[5] 将实体和关系扩展至复数空间，利用实部与虚部增强模型对方向性关系的表达能力。HolE[6] 则引入循环相关运算，在模型表达能力与计算效率之间取得了较好的平衡。

在上述工作的基础上，后续研究进一步从表示空间和分解形式等方面进行了拓展。ANALOGY[26] 通过类比推理思想增强语义建模能力，SimplE[27] 和 TuckER[28] 分别从分解结构简化和张量分解泛化角度提升模型性能与可解释性。同时，CrossE[29] 则着重建模实体与关系之间的交叉交互特征。此外，QuatE、DualE 和 HopfE[30-32] 等方法进一步拓展表示空间，以增强对复杂关系及高阶语义交互的刻画能力。因此，基于张量分解的方法在语义匹配和多元交互建模方面具有明显优势。然而，此类方法通常侧重三元组内部实体与关系的匹配，对图结构中局部拓扑模式利用不足，因此在复杂结构感知方面仍存在一定局限。

1.3 基于神经网络的知识图谱补全方法

随着深度学习技术的发展，基于神经网络的知识

图谱补全方法逐渐成为重要研究方向。相较于传统的平移模型和张量分解模型，神经网络具有更强的非线性表示能力，能够深入挖掘实体与关系之间复杂的隐含交互特征[33-35]。根据建模方式的不同，基于神经网络的知识图谱补全方法又可以分为传统神经网络方法、卷积神经网络方法和图神经网络方法。

传统神经网络方法较早将非线性结构引入知识图谱表示学习。SME[33]通过神经网络将实体与关系映射至潜在空间，以联合编码结构和语义信息。NTN[34]则引入双线性张量层，建模头尾实体之间复杂的非线性交互。NAM[35]通过结合关系调制神经网络与普通神经网络，实现了实体与关系表示的联合学习。此类方法突破了线性模型的表达限制，但对知识图谱拓扑结构的直接利用仍不充分。

随后，卷积神经网络被引入知识图谱补全任务中，以进一步提取实体和关系表示中的高阶交互特征。ConvE[36]通过二维卷积建模头实体与关系之间的交互，是卷积类知识图谱补全方法中的代表模型。此后，ConvKB、HypER、ConvR 和 InteractE[37-40]从特征交互方式和卷积核生成机制等方面进行改进，进一步增强了模型对复杂语义特征的提取能力。然而，卷积神经网络主要关注嵌入特征之间的局部交互，对图结构依赖关系的建模仍存在不足。

图神经网络能够直接利用知识图谱的拓扑结构，通过邻域聚合实现实体与关系信息的传播，因而在结构建模方面展现出较大优势。其中，RGCN[7]针对不同关系类型设计关系特定的变换矩阵，在关系约束下聚合邻域实体信息，是图神经网络应用于知识图谱补全的代表性模型。KBGAT[8]进一步将注意力机制引入图结构编码，根据邻居实体及其关联关系的重要性自适应分配聚合权重，从而增强对关键邻域信息的捕获能力。SACN[9]和 CompGCN[10]分别通过融合图卷积编码与卷积解码、联合建模实体和关系表示，进一步提升了多关系图中的结构信息利用能力。然而，现有图神经网络方法大多仍以结构化三元组为主要信息来源，对文本、图像等多模态信息与图结构信息之间的深层交互和协同传播关注不足，因而其多模态语义与结构信息的联合建模能力仍有待进一步提升。

2 模型与方法

2.1 问题定义

多模态知识图谱补全的目标是：在给定不完整三元组$(\boldsymbol{h},\boldsymbol{r},?)$或$(?,\boldsymbol{r},\boldsymbol{t})$的情况下，综合利用知识图谱中的结构信息以及实体对应的文本、图像等辅助模态信息学习实体和关系的嵌入表示，并据此预测缺失实体。为衡量候选三元组的合理性，通常模型需要定义一个评分函数$f_r(\boldsymbol{h},\boldsymbol{t})$用于衡量三元组$(\boldsymbol{h},\boldsymbol{r},\boldsymbol{t})$成立的可能性，评分越高则表示该三元组越有可能为真实事实。

为了统一实现头实体预测和尾实体预测，对每个关系$\boldsymbol{r}\in\mathcal{R}$构建其逆关系$\boldsymbol{r}^{-1}$，使得原三元组$(\boldsymbol{h},\boldsymbol{r},\boldsymbol{t})$对应逆三元组$\left(\boldsymbol{t},\boldsymbol{r}^{-1},\boldsymbol{h}\right)$。扩展后的关系集合$\mathcal{R}'=\mathcal{R}\cup\mathcal{R}^{-1}$，$|\mathcal{R}'|=2N_r$。通过引入逆关系，尾实体预测任务$(\boldsymbol{h},\boldsymbol{r},?)$与头实体预测任务$(?,\boldsymbol{r},\boldsymbol{t})\equiv\left(\boldsymbol{t},\boldsymbol{r}^{-1},?\right)$可统一为相同的形式，从而简化模型设计与训练流程。

2.2 模型总体框架

本文提出的基于多视图增强图注意力网络的多模态知识图谱补全方法采用编码—融合—解码的三阶段架构，整体框架如图 1 所示。

模型的输入包含实体的结构信息、文本信息和图像信息。其中，结构信息来自多模态知识图谱中实体的初始化可学习嵌入，文本信息和图像信息则分别由预训练的文本编码器和图像编码器提取。模型整体分为 3 个阶段：首先，采用多视图图注意力编码器学习不同模态下的实体表示；其次，通过多模态后融合策略整合结构、文本和图像特征；最后，利用 ComplEx 双线性解码器计算三元组得分，完成知识图谱补全。下文将依次介绍关系感知图注意力编码器、多视图编码策略、多模态后融合策略以及解码器的具体设计。

2.3 关系感知图注意力编码器

关系感知图注意力编码器是本文方法的核心，该模块的构建参考了近年来在知识图谱表示学习任务中表现出色的关系注意力网络[8]。关系感知图注意力编码器通过结合关系感知的消息构造机制和基于三元组的图注意力机制，能够在图结构上有效地聚合邻域信息，学习关系感知的多模态实体表示。

在多模态知识图谱中，实体之间的语义关联由关系类型决定。为在消息传递过程中充分建模关系语义，关系感知图注意力编码器在源节点特征传播到目标节点之前，先利用关系嵌入对源节点表示进行关系感知变换，并将多种变换结果融合为最终的传递消息。

具体地，给定源实体表示$\boldsymbol{e}_j\in\mathbb{R}^d$和关系表示$\boldsymbol{r}\in\mathbb{R}^d$，我们采用以下 4 种组合算子从不同角度刻画实体与关系之间的交互模式：

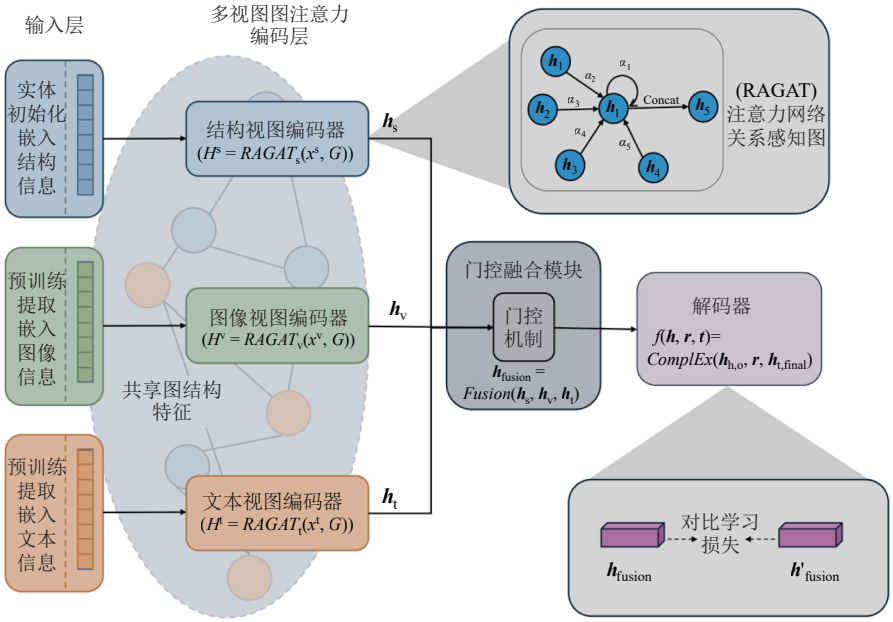


Fig. 1 The proposed multimodal knowledge graph completion framework

图 1 提出的多模态知识图谱补全框架

$$\begin{aligned}\varphi_{\text{sub}}\left(\boldsymbol{e}_j,\boldsymbol{r}\right)&=\boldsymbol{e}_j-\boldsymbol{r},\\ \varphi_{\text{mult}}\left(\boldsymbol{e}_j,\boldsymbol{r}\right)&=\boldsymbol{e}_j\odot\boldsymbol{r},\\ \varphi_{\text{rotate}}\left(\boldsymbol{e}_j,\boldsymbol{r}\right)&=\boldsymbol{e}_j\circ\boldsymbol{r},\\ \varphi_{\text{corr}}\left(\boldsymbol{e}_j,\boldsymbol{r}\right)&=\boldsymbol{e}_j\star\boldsymbol{r},\end{aligned}\tag{1}$$

其中，$\odot$表示逐元素乘法，$\circ$表示旋转组合操作，$\star$表示循环相关操作。上述 4 种算子分别源自 TransE[3] 的平移假设、DistMult[4] 的双线性交互、RotatE[21] 旋转建模以及 HolE[6] 全息嵌入。本文同时使用 4 种组合算子并将各自的变换结果分别经过独立的线性投影后进行拼接融合，充分整合不同表示空间中的语义信息。

关系感知图注意力编码器采用基于完整三元组的注意力机制来估计不同邻域信息的重要性。相比于仅基于节点对计算注意力的方法，引入关系信息参与注意力计算能够更精细地估计不同关系类型下信息传播的差异化权重。对于头实体i和尾实体j之间类型为r的边，注意力系数的计算方式如下。首先，将头实体、关系和尾实体在第$l-1$层的隐藏表示进行拼接，通过注意力投影矩阵和 LeakyReLU 激活计算原始注意力分数：

$$\alpha_{ijr}^{(l)}=\boldsymbol{a}^{(l)}L\left(\boldsymbol{W}_{\text{att}}^{(l)}\cdot\left[\boldsymbol{h}_i^{(l-1)}\left\|\boldsymbol{r}^{(l-1)}\right\|\boldsymbol{h}_j^{(l-1)}\right]\right),\tag{2}$$

其中，L表示 LeakyReLU 变换，$\boldsymbol{h}_i^{(l-1)}$，$\boldsymbol{h}_j^{(l-1)}\in\mathbb{R}^d$分别为头实体和尾实体在第$l-1$层的隐藏表示，$\boldsymbol{r}^{(l-1)}\in\mathbb{R}^d$为关系在第$l-1$层的嵌入表示，$\boldsymbol{W}_{\text{att}}^{(l)}\in\mathbb{R}^{d_{\text{out}}\times 3d}$为可学习的注意力投影矩阵，$\boldsymbol{a}^{(l)}\in\mathbb{R}^{1\times d_{\text{out}}}$为第$l$层的可学习注意力向量。随后对头实体$i$的所有邻居节点及其对应的所有关系类型进行 Softmax 归一化，这种双重求和归一化能够在同一对实体之间存在多种关系的情况下，合理分配各三元组的注意力权重。得到注意力权重和传递消息后，实体表示的聚合更新方程定义为

$$\boldsymbol{h}_i^{(l)}=\sigma'\left(\sum_{(j,\boldsymbol{r})\in N_i}\alpha_{ijr}^{(l)}\cdot m\left(\boldsymbol{e}_j,\boldsymbol{r}\right)+\boldsymbol{W}_{\text{res}}^{(l)}\cdot\boldsymbol{h}_i^{(l-1)}\right),\tag{3}$$

其中，σ'为 Tanh 激活函数，$\boldsymbol{W}_{\text{res}}^{(l)}\in\mathbb{R}^{d_{\text{out}}\times d}$为残差映射矩阵。该更新方程将注意力加权的邻域消息聚合与残差连接统一在同一框架中，残差项$\boldsymbol{W}_{\text{res}}^{(l)}\cdot\boldsymbol{h}_i^{(l-1)}$保留了节点自身在上一层的特征信息，用于缓解图神经网络中常见的过平滑问题。

关系感知图注意力编码器还引入了层间注意力传递机制，用以提升注意力学习的稳定性。该机制通过对当前层计算得到的注意力权重和上一层的注意力权重进行线性插值，使注意力信息在层间平滑传播：

$$\boldsymbol{\alpha}^{(l)}=(1-\beta)\cdot\boldsymbol{\alpha}_{\text{cur}}^{(l)}+\beta\cdot\boldsymbol{\alpha}^{(l-1)},\tag{4}$$

其中，$\beta\in[0,1]$为注意力传递系数，控制当前层注意力与上一层注意力的混合比例。当$\beta=0$时退化为仅使用当前层注意力。该机制使得深层网络中的注意力权重也能够参考浅层的分配模式，避免注意力分布在层间出现剧烈波动。同时，关系嵌入在每一层也通过线性变换进行同步更新，以保持关系表示与实体表示在语义空间上的一致性：

$$\boldsymbol{r}^{(l)}=\boldsymbol{W}_{\text{rel}}^{(l)}\cdot\boldsymbol{r}^{(l-1)},\tag{5}$$

其中，$\boldsymbol{W}_{\text{rel}}^{(l)}\in\mathbb{R}^{d_{\text{out}}\times d}$为第$l$层关系嵌入的可训练变换矩阵。

2.4 模态分离的多视图多模态编码

在多模态知识图谱中，不同模态的信息存在显著差异：结构模态反映实体间的关系连接模式，图像和文本模态则分别提供由预训练模型提取的视觉特征与语义描述。若在图注意力编码前直接采用早期融合，不同模态间的语义分布差异可能导致融合表示与图结构传播机制不匹配，从而削弱邻域信息的编码效果。此外，单一图注意力编码器也难以同时学习适配不同模态特征的最优传播方式。因此，本文提出模态分离的多视图多模态编码策略：为每种模态构建独立的关系感知图注意力编码器，形成 3 个并行的“视图”，在各自的特征空间中学习最适合该模态的图结构嵌入表示，最后再通过后融合策略进行整合。

结构视图以可学习的实体嵌入 $\boldsymbol{E}_{\mathrm{s}} \in \mathbb{R}^{N_e \times 2d}$ 作为初始节点特征，配合独立的关系嵌入 $\boldsymbol{R}_{\mathrm{s}} \in \mathbb{R}^{2N_r \times 2d}$，通过 L 层关系感知图注意力编码器进行编码：

$$\boldsymbol{h}_{\mathrm{s}}^{(0)} = \boldsymbol{E}_{\mathrm{s}}, \tag{6}$$

$$\begin{aligned} \boldsymbol{h}_{\mathrm{s}}^{(l)} &= RAGAT_{\mathrm{s}}^{(l)}\left(\boldsymbol{h}_{\mathrm{s}}^{(l-1)}\right), \\ \boldsymbol{r}_{\mathrm{s}}^{(l)} &= RAGAT_{\mathrm{s}}^{(l)}\left(\boldsymbol{r}_{\mathrm{s}}^{(l-1)}\right), \\ \boldsymbol{\alpha}_{\mathrm{s}}^{(l)} &= RAGAT_{\mathrm{s}}^{(l)}\left(\boldsymbol{\alpha}_{\mathrm{s}}^{(l-1)}\right) \end{aligned} \tag{7}$$

其中 $l = 1,2,\cdots,L$。图像视图以预训练的图像模态嵌入经线性投影后的表示作为初始节点特征，配合独立的关系嵌入 $\boldsymbol{R}_{\mathrm{v}} \in \mathbb{R}^{2N_r \times 2d}$：

$$\boldsymbol{h}_{\mathrm{v}}^{(0)} = \boldsymbol{W}_{\mathrm{v}} \cdot \boldsymbol{E}_{\mathrm{v}} + \boldsymbol{b}_{\mathrm{v}}, \tag{8}$$

$$\begin{aligned} \boldsymbol{h}_{\mathrm{v}}^{(l)} &= RAGAT_{\mathrm{v}}^{(l)}\left(\boldsymbol{h}_{\mathrm{v}}^{(l-1)}\right), \\ \boldsymbol{r}_{\mathrm{v}}^{(l)} &= RAGAT_{\mathrm{v}}^{(l)}\left(\boldsymbol{r}_{\mathrm{v}}^{(l-1)}\right), \\ \boldsymbol{\alpha}_{\mathrm{v}}^{(l)} &= RAGAT_{\mathrm{v}}^{(l)}\left(\boldsymbol{\alpha}_{\mathrm{v}}^{(l-1)}\right) \end{aligned} \tag{9}$$

其中，$\boldsymbol{W}_{\mathrm{v}} \in \mathbb{R}^{2d \times d_{\mathrm{v}}}$ 为图像投影矩阵。

文本视图与图像视图类似，以预训练的文本特征经线性投影后的表示作为初始节点特征，配合独立的关系嵌入 $\boldsymbol{R}_{\mathrm{t}} \in \mathbb{R}^{2N_r \times 2d}$：

$$\boldsymbol{h}_{\mathrm{t}}^{(0)} = \boldsymbol{W}_{\mathrm{t}} \cdot \boldsymbol{E}_{\mathrm{t}} + \boldsymbol{b}_{\mathrm{t}}, \tag{10}$$

$$\begin{aligned} \boldsymbol{h}_{\mathrm{t}}^{(l)} &= RAGAT_{\mathrm{t}}^{(l)}\left(\boldsymbol{h}_{\mathrm{t}}^{(l-1)}\right), \\ \boldsymbol{r}_{\mathrm{t}}^{(l)} &= RAGAT_{\mathrm{t}}^{(l)}\left(\boldsymbol{r}_{\mathrm{t}}^{(l-1)}\right), \\ \boldsymbol{\alpha}_{\mathrm{t}}^{(l)} &= RAGAT_{\mathrm{t}}^{(l)}\left(\boldsymbol{\alpha}_{\mathrm{t}}^{(l-1)}\right) \end{aligned} \tag{11}$$

其中，$\boldsymbol{W}_{\mathrm{t}} \in \mathbb{R}^{2d \times d_{\mathrm{t}}}$ 为文本投影矩阵。

3 个视图共享图结构信息，在相同的图拓扑结构上进行消息传播操作，但同时每个视图有独立的关系感知图注意力编码器网络参数。每个视图还维护独立的关系嵌入（$\boldsymbol{R}_{\mathrm{s}}$、$\boldsymbol{R}_{\mathrm{v}}$、$\boldsymbol{R}_{\mathrm{t}}$），以确保不同模态下的关系语义有不同的表达。

2.5 后融合策略

经过多视图图注意力编码器编码后，需要将结构视图、图像视图和文本视图的实体表示 $\boldsymbol{h}_{\mathrm{s}}$、$\boldsymbol{h}_{\mathrm{v}}$、$\boldsymbol{h}_{\mathrm{t}}$ 融合为统一的多模态实体表示 $\boldsymbol{h}_{\mathrm{final}}$。结合不同融合方式的实验比较结果，本文最终采用门控机制实现多视图表示融合。首先门控机制会为每个视图的实体嵌入学习一个门控向量，其计算方式为：

$$\boldsymbol{g}_{\mathrm{s}} = \sigma(\boldsymbol{W}_{\mathrm{s}}\boldsymbol{h}_{\mathrm{s}}), \boldsymbol{g}_{\mathrm{v}} = \sigma(\boldsymbol{W}_{\mathrm{v}}\boldsymbol{h}_{\mathrm{v}}), \boldsymbol{g}_{\mathrm{t}} = \sigma(\boldsymbol{W}_{\mathrm{t}}\boldsymbol{h}_{\mathrm{t}}), \tag{12}$$

其中，$\boldsymbol{W}_{\mathrm{s}}$、$\boldsymbol{W}_{\mathrm{v}}$ 和 $\boldsymbol{W}_{\mathrm{t}}$ 表示对应模态的可学习变换矩阵，$\sigma(\cdot)$ 为 Sigmoid 激活函数，用于将门控向量值约束到 (0,1) 区间，其定义为：

$$\sigma(x) = \frac{1}{1+e^{-x}}。 \tag{13}$$

为使不同模态在同一特征维度上的权重具有可比性，还需要对门控向量进行归一化处理：

$$\widetilde{\boldsymbol{g}}_m = \frac{\boldsymbol{g}_m}{\boldsymbol{g}_{\mathrm{s}} + \boldsymbol{g}_{\mathrm{v}} + \boldsymbol{g}_{\mathrm{t}} + \varepsilon}, \ m \in \{\mathrm{s},\mathrm{v},\mathrm{t}\}, \tag{14}$$

其中，ε 为防止分母为零的数值稳定项。得到归一化的门控权重后，3 个模态的表示按照门控权重进行加权融合，得到最终的多模态实体表示：

$$\boldsymbol{h}_{\mathrm{final}} = \widetilde{\boldsymbol{g}}_{\mathrm{s}} \odot \boldsymbol{h}_{\mathrm{s}} + \widetilde{\boldsymbol{g}}_{\mathrm{v}} \odot \boldsymbol{h}_{\mathrm{v}} + \widetilde{\boldsymbol{g}}_{\mathrm{t}} \odot \boldsymbol{h}_{\mathrm{t}}, \tag{15}$$

其中，$\odot$ 表示逐元素乘法。

2.6 解码器与评分函数

我们采用 ComplEx 作为解码器，ComplEx 通过在复数空间中建模实体和关系，能够有效处理对称关系、反对称关系和一般关系等多种关系模式。将实体嵌入和关系嵌入视为复数向量，其中 ComplEx 的评分函数定义为

$$\varphi(\boldsymbol{h},\boldsymbol{r},\boldsymbol{t}) \triangleq Re\left(\left\langle \boldsymbol{h},\boldsymbol{r},\bar{\boldsymbol{t}} \right\rangle\right), \tag{16}$$

其中，$\bar{\boldsymbol{t}}$ 表示尾实体表示的复数共轭，$\langle\cdot,\cdot,\cdot\rangle$ 表示三线性点积。展开后可写为

$$\begin{aligned} \varphi(\boldsymbol{h},\boldsymbol{r},\boldsymbol{t}) = &\left(\boldsymbol{h}^{\mathrm{re}} \odot \boldsymbol{r}^{\mathrm{re}} - \boldsymbol{h}^{\mathrm{im}} \odot \boldsymbol{r}^{\mathrm{im}}\right) \cdot \boldsymbol{t}^{\mathrm{re}} + \\ &\left(\boldsymbol{h}^{\mathrm{re}} \odot \boldsymbol{r}^{\mathrm{im}} + \boldsymbol{h}^{\mathrm{im}} \odot \boldsymbol{r}^{\mathrm{re}}\right) \cdot \boldsymbol{t}^{\mathrm{im}}。 \end{aligned} \tag{17}$$

对于链接预测，需要计算查询 $(\boldsymbol{h},\boldsymbol{r},?)$ 对所有候选实体的得分：

$$\boldsymbol{S} = \left(\boldsymbol{h}^{\mathrm{re}} \odot \boldsymbol{r}^{\mathrm{re}} - \boldsymbol{h}^{\mathrm{im}} \odot \boldsymbol{r}^{\mathrm{im}}\right) \cdot \boldsymbol{E}_{\mathrm{all}}^{\mathrm{re,T}} + \left(\boldsymbol{h}^{\mathrm{re}} \odot \boldsymbol{r}^{\mathrm{im}} + \boldsymbol{h}^{\mathrm{im}} \odot \boldsymbol{r}^{\mathrm{re}}\right) \cdot \boldsymbol{E}_{\mathrm{all}}^{\mathrm{im,T}}, \tag{18}$$

其中，$\boldsymbol{E}_{\mathrm{all}}$ 为融合后的所有实体嵌入矩阵。

2.7 训练与优化策略

本方法采用一对多的训练策略。对于每个训练三元组 $(\boldsymbol{h},\boldsymbol{r},\boldsymbol{t})$，将模型输出视为对所有候选实体的分类问题，使用交叉熵损失：

$$\mathcal{L}_{\mathrm{CE}} = -\sum_{(\boldsymbol{h},\boldsymbol{r},\boldsymbol{t}) \in T} \log \frac{\exp(\varphi(\boldsymbol{h},\boldsymbol{r},\boldsymbol{t}))}{\sum_{\boldsymbol{t}' \in E} \exp(\varphi(\boldsymbol{h},\boldsymbol{r},\boldsymbol{t}'))}。 \tag{19}$$

为防止过拟合，本文采用核范数正则化（N3 正则

化)[41]约束模型参数：

$$\mathcal{L}_{\mathrm{N3}}=\lambda\cdot\frac{1}{|B|}\cdot\sum_{(\boldsymbol{h},\boldsymbol{r},\boldsymbol{t})\in B}\left(\|\boldsymbol{h}\|_3^3+\|\boldsymbol{r}\|_3^3+\|\boldsymbol{t}\|_3^3\right),\tag{20}$$

其中，$\|\cdot\|_3^3=\sum_i|x_i|^3$，$\lambda$为正则化系数，$B$为当前批次的训练样本集合。

为增强多模态实体表示的区分性，借鉴监督对比学习的思想，本文引入对比学习损失[42]，利用三元组中的尾实体标签作为监督信号，构建对比学习损失：

$$\mathcal{L}_{\mathrm{SCL}}=\frac{-\tau}{\tau_0}\cdot\frac{1}{|P(i)|}\cdot\sum_{p\in P(i)}\log\frac{\exp\left(z_i\cdot z_p/\tau\right)}{\sum_{a\in A(i)}\exp\left(z_i\cdot z_a/\tau\right)},\tag{21}$$

其中，z_i为归一化后的实体嵌入，$P(i)$为与样本i具有相同尾实体标签的正样本集合，$A(i)$为除i以外的所有样本集合，τ为温度参数，τ_0为基础温度参数。该损失鼓励具有相同尾实体的三元组在嵌入空间中更加接近，而不同尾实体的三元组相互远离。

综合上述3项损失，模型的总优化目标可表示为：

$$\mathcal{L}=\mathcal{L}_{\mathrm{CE}}+\mathcal{L}_{\mathrm{N3}}+\lambda_{\mathrm{CL}}\cdot\mathcal{L}_{\mathrm{SCL}},\tag{22}$$

其中λ_{CL}为对比学习损失的权重系数。

3 实验设置

为验证提出模型的有效性，本节从数据集、评价指标、对比方法以及实现细节4个方面介绍实验设置。

3.1 数据集

我们在3个公开可用的多模态知识图谱补全基准数据集上对所提出的方法进行评估，分别是MKG-W、MKG-Y[12]和DB15K[13]。上述数据集均包含知识图谱结构信息以及与实体对应的文本描述和图像信息，能够较为全面地检验模型在多模态场景下的知识表示与补全能力。其中，DB15K数据集来源于DBpedia[43]，是多模态知识图谱补全研究中广泛使用的基准数据集之一；MKG-W和MKG-Y则由WikiData子集构建而成[44]，包含较为丰富的实体图像和文本描述信息。为保证实验的公平性，本文使用各数据集官方发布版本中提供的文本与图像模态特征，不额外引入其他外部知识或附加训练数据。3个数据集的详细统计信息如表1所示。

在多模态特征来源方面，图像特征由预训练图像模型VGG[45]提取的图像嵌入获得，文本特征由预训练语言模型BERT[46]提取的文本嵌入获得。

3.2 评价指标

我们采用多模态知识图谱补全任务中最常见的链接预测实验来评估模型性能。对于测试集中每一个待预测三元组$(\boldsymbol{h},\boldsymbol{r},?)$或$(?,\boldsymbol{r},\boldsymbol{t})$，模型根据评分函数对所有候选实体进行打分并排序，以预测缺失的头实体或尾实体。若正确实体在排序结果中位置越靠前，则说明模型的补全能力越强。本文采用以下2类基于排序的评价指标来全面评价模型性能：1)平均倒数排名(MRR)：通过对所有测试三元组的正确候选实体倒数排名取平均，反映模型将正确候选实体排在前列的总体能力；2)前k个命中率(Hits@k)：用于衡量模型是否能够在前k个候选结果中命中正确实体。对于每个测试三元组，若正确候选实体的排名不超过k，则认为该样本预测成功；否则预测失败。

Table 1 Statistical Information of the Experimental Datasets

表1 实验数据集统计信息

数据集	实体数	关系数	训练集	验证集	测试集
DB15K	12 842	279	79 222	9 902	9 904
MKG-W	15 000	169	34 196	4 276	4 274
MKG-Y	15 000	28	21 310	2 665	2 663

在测试集评估过程中，我们遵循多模态知识图谱补全任务的标准过滤评价协议。在对候选实体进行排序时，过滤掉除当前目标实体外能够与头实体和关系组成数据集中已知真实三元组的其他候选实体，再计算目标实体的排名。该方式可以避免模型因将其他正确答案排在目标实体之前而受到不合理惩罚。

3.3 对比方法

为充分验证本文方法的有效性，实验选取了多种具有代表性的知识图谱补全模型作为对比方法，具体包括单模态结构和多模态补全模型两大类。

1)单模态结构模型。单模态模型仅利用知识图谱中的结构三元组信息进行表示学习，主要包括：

① TransE[3]。将关系建模为实体间的平移操作，是最经典的知识图谱补全模型之一。

② DistMult[4]。采用双线性对角矩阵形式建模实体与关系之间的交互。

③ ComplEx[5]。在复数空间中对实体和关系进行表示，可有效建模对称与反对称关系。

④ RotatE[21]。将关系表示为复数空间中的旋转操作，在链式关系和反对称关系建模中具有较强能力。

2)多模态知识图谱补全模型。多模态模型综合利用结构、文本和图像等多源信息进行知识表示与补全，本文选取如下代表性方法作为对比：

① IKRL[14]。较早将实体图像模态信息引入知识图谱表示学习的方法。

② TransAE[15]。通过在平移式知识表示框架中引入自编码机制实现模态特征融合。

③ MMKRL[11]。联合建模结构、文本与图像信息等多模态信息以提升方法补全性能。

④ RSME[12]。针对不同关系对图像信息依赖程度不同的问题，通过关系感知的图像表示建模机制增强多模态知识图谱补全能力。

⑤ VBKGC[47]。利用图像模态信息辅助多模态知识图谱补全。

⑥ OTKGE[16]。最优传输视角对不同模态表示进行分布对齐，解决多模态信息直接融合带来的补全性能不足问题。

⑦ AdaMF[17]。通过自适应融合机制动态整合不同模态信息。

3.4 实现细节

本文方法基于 PyTorch 深度学习框架实现，所有实验均在 NVIDIA RTX 3090 GPU 上完成。为保证实验设置的统一性与可复现性，模型超参数均通过网格搜索方式进行选择。其中嵌入维度 d 遵从对比方法的一致性，设置 $d=256$；关系感知图注意力网络层数 L 从集合 $\{1,2,3,4\}$ 中选取；注意力传递系数 β 从 $[0,1]$ 区间中选取；Dropout 比例从集合 $\{0.0,0.1,0.2,0.3\}$ 中选取；学习率从集合 $\{0.1,0.05,0.01,0.001\}$ 中选取；正则化系数从集合 $\{0,0.001,0.01,0.1\}$ 中选取。

在模型训练过程中，本文采用 Adagrad[48] 优化器对参数进行更新。同时，为缓解模型过拟合问题并约束参数规模，我们引入 N3 正则化项对模型参数进行约束。本文最终实验结果均为 5 次随机重复实验结果的平均值，以减小偶然因素对实验结论的影响。此外，为保证不同对比模型之间的公平性，我们尽可能遵循各基线模型原始论文中的推荐设置。

4 实验结果与分析

4.1 总体实验结果分析

表 2 给出了本文提出的 MEGAT 模型与其余对比模型在 DB15K、MKG-W 和 MKG-Y 三个数据集上的链接预测结果。总体来看，MEGAT 在三个数据集上均取得了较为稳定且具有竞争力的结果，尤其在 MRR 和 Hits@1 等更关注排序前列准确性的指标上表现突出，说明 MEGAT 能够有效提升多模态知识图谱补全任务中的实体排序质量。

具体而言，在 DB15K 数据集上，MEGAT 在 4 项评价指标上均取得最优结果，分别达到 38.64%、30.60%、42.74% 和 53.77%，相较于其他对比模型表现出明显优势。这表明本文 MEGAT 能够更加充分地挖掘多模态语义信息与知识图谱结构信息之间的互补关系，从而学习到更具判别性的实体表示。在 MKG-W 数据集上，MEGAT 在 MRR 和 Hits@1 两项指标上取得最佳结果，虽然在 Hits@3 和 Hits@10 上略逊于 AdaMF，但整体结果仍保持在较高水平，表明 MEGAT 在该数据集上具有较好的稳定性和竞争力。在 MKG-Y 数

Table 2 Link Prediction Results of Different Models on Three Datasets

表 2 不同模型在 3 个数据集上的链接预测结果 %

类型	模型	DB15K				MKG-W				MKG-Y			
		MRR	Hits@1	Hits@3	Hits@10	MRR	Hits@1	Hits@3	Hits@10	MRR	Hits@1	Hits@3	Hits@10
单模态	TransE	24.86	12.78	31.48	47.07	29.19	21.06	33.2	44.23	30.73	23.45	35.8	43.37
	DistMult	23.03	14.78	26.28	39.59	20.99	15.93	22.28	30.86	28.71	22.26	27.8	35.95
	ComplEx	27.48	18.37	31.57	45.37	24.93	19.09	26.69	36.73	25.04	19.33	32.12	40.93
	RotatE	29.28	17.87	36.12	49.66	33.67	26.8	36.68	46.73	34.95	29.1	38.35	45.3
多模态	IKRL	26.82	14.09	34.93	49.09	32.36	26.11	34.75	44.07	33.22	30.37	34.28	38.26
	TransAE	28.09	21.25	31.17	41.17	30	21.23	34.91	44.72	28.1	25.31	29.1	33.03
	MMKRL	26.81	13.85	35.07	49.39	30.1	22.16	34.9	44.69	36.81	31.66	39.79	45.31
	RSME	29.76	24.15	32.12	40.29	29.23	23.36	31.97	40.43	34.44	31.78	36.07	39.09
	VBKGC	30.61	19.75	37.18	49.44	30.61	24.91	33.01	40.88	37.04	33.76	38.75	42.3
	OTKGE	23.86	14.85	25.89	34.83	34.36	28.85	36.25	44.88	35.51	31.97	37.18	41.38
	AdaMF	32.51	21.31	39.67	51.68	34.27	27.21	**37.86**	**47.21**	**38.06**	33.49	**40.44**	**45.48**
	MEGAT（本文）	**38.64**	**30.60**	**42.74**	**53.77**	**34.61**	**29.08**	34.65	44.62	37.63	**34.57**	39.26	43.03

注：黑体数值表示最优结果。

据集上，MEGAT 在 Hits@1 指标上取得最优值，同时在 MRR 上取得次优结果，说明该模型在不同数据集上均具备较好的泛化能力。

MEGAT 取得较优性能主要得益于以下 2 个方面：首先，采用模态分离的多视图编码方式，使不同模态信息能够在相对独立的表示空间中进行充分建模，从而更好地保留各模态的特异性信息。其次，通过后融合策略整合不同模态表示，避免了早期融合过程中可能出现的信息混杂和噪声干扰问题，使模型能够更加有效地利用多模态知识图谱中的互补信息。

4.2 消融实验分析

为验证本文所提出模块的有效性，本文从模态信息、融合策略和编码器机制 3 个角度设计了消融实验，并对 MEGAT 的性能表现进行深入分析。

1)模态信息消融。为验证不同模态信息对模型性能的具体贡献，本文在 MKG-Y 数据集上进行了模态信息消融实验，实验结果如表 3 所示。

Table 3 Results of Modal Ablation on MKG-Y Dataset

表 3 MKG-Y 数据集上模态消融结果 %

设置	MRR	Hits@1	Hits@3	Hits@10
完整模型	**37.63**	**34.57**	**39.26**	**43.03**
去掉图像	36.09	32.58	37.43	41.91
去掉文本	36.75	33.27	38.66	41.81
仅结构模态	33.42	29.53	35.54	40.03

注：黑体数值表示最优结果。

由表 3 可以看出，完整模型在 4 项评价指标上均取得最优的结果，这证明了联合建模能够有效提升多模态知识图谱补全任务的表示学习效果。

具体地，在移除图像模态信息后，模型的 MRR 由 37.63% 下降至 36.09%，Hits@1 由 34.57% 下降至 32.58%，Hits@3 和 Hits@10 也分别下降至 37.43% 和 41.91%。这一结果表明，图像信息能够为实体表示提供重要的补充语义，尤其有助于提升高置信度预测的准确性。移除文本模态后，模型性能同样下降，但降幅略小于移除图像模态的情况，说明文本描述对实体语义建模也具有积极作用。仅保留结构模态时，模型性能下降最为明显，表明单一结构信息难以全面刻画实体的复杂语义，而图像和文本信息能够从不同角度形成有效补充。

2)融合策略消融。为进一步分析不同融合机制对模型性能的影响，本文在 MKG-W 数据集上比较了多种常见融合策略，以选取最优融合策略作为多视图信息融合机制，实验结果如表 4 所示。

Table 4 Results of Ablation with Different Fusion Strategies on MKG-W Dataset

表 4 MKG-W 数据集上不同融合策略消融结果 %

融合策略	MRR	Hits@1	Hits@3	Hits@10
加权求和	28.38	23.36	30.22	37.28
拼接 MLP 融合	14.18	11.32	14.66	18.92
注意力融合	32.45	27.16	34.59	42.28
门控融合	**34.61**	**29.08**	**34.65**	**44.62**

注：黑体数值表示最优结果。

不同融合策略对模型性能具有显著影响。其中，门控融合策略在 4 项评价指标上均取得最优结果，说明该策略能够更有效地整合不同模态的表示信息，从而提升模型整体性能。针对其他融合策略的具体分析，拼接 MLP 融合策略的实验结果最差，MRR 仅为 14.18%。结果说明直接对多模态特征进行拼接并通过 MLP 映射，难以有效学习各模态之间的互补关系，反而可能由于信息冗余、噪声等问题对最终表示造成干扰，从而显著削弱模型性能。加权求和策略的表现相对一般，MRR 仅为 28.38%。结果显示了简单的线性加权方式虽然能够实现不同模态信息的初步整合，但由于其缺乏对模态差异性的细粒度建模能力，在复杂多模态场景下融合效果有限。同时注意力融合策略取得了较好的效果，明显优于加权求和和拼接 MLP 融合 2 个策略。这说明注意力机制能够在一定程度上根据输入特征的重要性为不同模态分配权重，从而提升融合表示的质量。然而，与门控融合相比注意力融合仍存在一定差距。

3)编码器机制消融。为进一步验证关系感知图注意力编码器中各组成机制的有效性，本文在 MKG-Y 数据集上分别移除关系感知消息构造、基于三元组的注意力机制和层间注意力传递机制。

由表 5 可知，完整模型在 MRR 和 Hits@1 指标上分别达到 37.63% 和 34.57%，整体表现优于各消融设置，说明编码器中的不同机制能够相互配合，共同提升模型的知识图谱补全性能。具体而言，移除关系感知消息构造后，模型性能下降最为明显，MRR 和 Hits@1 分别降至 36.25% 和 32.86%，Hits@3 与 Hits@10 也分别下降至 38.00% 和 42.56%。这表明，利用多种实体-关系组合算子构造邻域消息，能够从不同交互角度刻画关系语义，增强模型对复杂关系模式的表示能力。移除基于三元组的注意力机制后，MRR 和 Hits@1 分别下降至 37.58% 和 34.36%，但 Hits@3 略有提升，Hits@10 保持不变。总体变化幅度较小，说明三元组注意力机制主要有助于优化正确实体在排序

Table 5 Results of Encoder Mechanism Ablation

表 5 编码器机制消融结果 %

设置	MRR	Hits@1	Hits@3	Hits@10
完整模型	**37.63**	**34.57**	39.26	**43.03**
去掉关系感知消息构造	36.25	32.86	38.00	42.56
去掉三元组注意力机制	37.58	34.36	**39.39**	43.03
去掉层间注意力	37.37	34.30	38.72	42.64

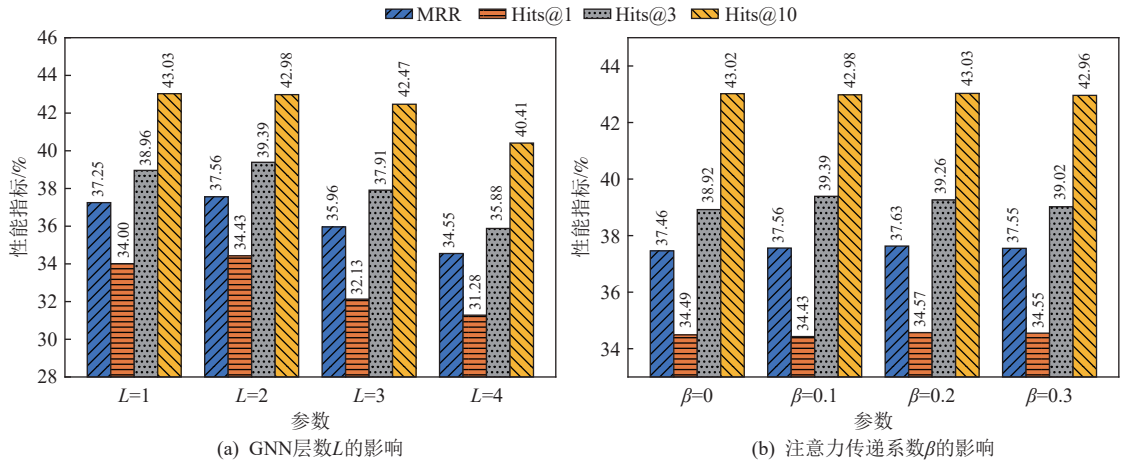
注：黑体数值表示最优结果。

前列的位置，而其对较大候选范围内命中能力的影响相对有限。移除层间注意力传递机制后，4 项指标均出现不同程度下降，其中 MRR 降至 37.37%，Hits@3 降至 38.72%。这一结果表明，在相邻图注意力层之间传递注意力信息，有助于维持注意力分布的稳定性，并提升多层邻域传播的效果。

4.3 参数敏感性分析

本节首先分析关系感知图注意力编码器层数L和层间注意力传递系数β对模型补全性能的影响，结果如图 2 所示；其次，通过向图像和文本特征中加入不同强度的高斯噪声，进一步评估模型在模态噪声干扰下的鲁棒性，结果如图 3 所示。

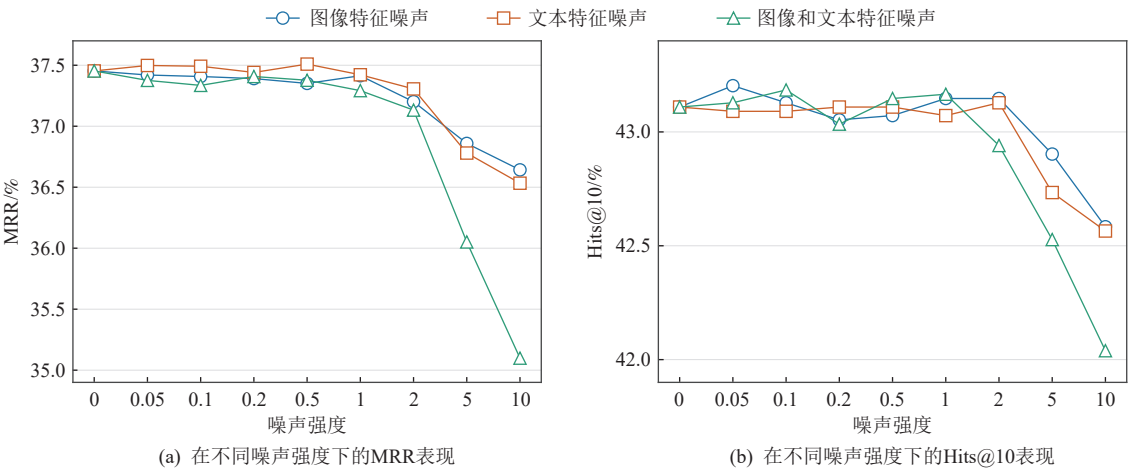
1）GNN 层数 L 的影响。如图 2(a)所示，在固定$\beta=0.2$的条件下，随着 RAGAT 层数L的增加模型性能呈现先升后降的趋势。$L=1$时模型仅聚合直接邻居信息，此时的指标中 MRR 为 37.25%，Hits@1 为 34.00%。而将层数增加至$L=2$后各项指标均达到最优或接近最优水平，此时除了直接邻居信息模型还能够捕获二阶邻域信息，使得指标中 MRR 提升至 37.56%，Hits@1 提升至 34.43%，Hits@3 提升至 39.39%。而当层数进一步增加至L=3 和$L=4$后模型性能出现明显下降，相较于最优设置分别下降了 3.01 个百分点和 3.15 个百分点。这一现象显示随着层数增加不同实体的表示在反复聚合过程中趋于一致，削弱了实体间的区分度。因此，本文在后续实验中均采用$L=2$的设置。



(a) GNN层数L的影响 (b) 注意力传递系数β的影响

Fig. 2 Sensitivity analysis of parameter on MKG-Y dataset

图 2 在 MKG-Y 数据集上的参数敏感性分析



(a) 在不同噪声强度下的MRR表现 (b) 在不同噪声强度下的Hits@10表现

Fig. 3 Sensitivity analysis of noise intensity on MKG-Y dataset

图 3 在 MKG-Y 数据集上的噪声强度敏感性分析

2)注意力传递系数β的影响。如图 2(b)所示,在固定$L=2$的条件下模型对注意力传递系数β的变化表现出较好的鲁棒性。图中不同β设置下各项指标的波动幅度较小,当$\beta=0$,不使用层间注意力传递,MRR 为 37.46%, Hits@1 为 34.49%;当$\beta>0$使用层间注意力传递后,模型在多数指标上有所提升,其中$\beta=0.2$时 MRR 达到最优值 37.63%, Hits@1 达到最优值 34.57%,而$\beta=0.1$时 Hits@3 达到最优值。说明相比于不使用层间注意力传递,适度的层间注意力传递有助于在多层网络中维持注意力分布的稳定性。而当β继续增大至 0.3 时,各项指标略有回落。结果说明过大的传递系数并不能取得好的效果,β在[0.1,0.2]区间内均能取得较好的效果。

3)噪声强度影响。为考察模型在模态特征受到噪声干扰时的鲁棒性,本文在 MKG-Y 数据集上分别对图像特征、文本特征以及 2 种模态特征同时加入不同强度的高斯噪声。具体而言,对于模态特征x,按照$x_{\text{noisy}}=x+\sigma\cdot std(x)\cdot\varepsilon,\varepsilon\sim N(0,I)$生成含噪特征,其中$std(x)$表示特征的标准差,$\sigma$表示噪声强度。实验中噪声强度取值为 0, 0.05, 0.1, 0.2, 0.5, 1, 2, 5, 10 时,结果如图 3 所示。当$\sigma\leqslant 2$时,模型的 MRR 和 Hits@10 仅小幅波动,说明其对轻度和中等噪声具有较好的鲁棒性。这主要得益于模态分离的多视图编码和门控后融合机制:前者能够避免单一模态的噪声直接干扰其他模态,后者可根据不同视图的表示质量动态调整融合权重。当噪声强度进一步增大至$\sigma=5$和$\sigma=10$时,模型性能开始明显下降。其中,仅对图像或文本特征加入噪声时, MRR 和 Hits@10 的下降幅度相对有限;而同时对 2 种模态特征加入噪声时,模型性能下降更为显著,在$\sigma=10$时, MRR 和 Hits@10 分别降至约 35.1% 和 42.0%,表明强噪声会削弱外部模态的有效语义,并在融合过程中产生累积影响。

5 结 论

在本文中,我们提出了一种基于多视图增强图注意力网络的多模态知识图谱补全方法 MEGAT。该方法将结构、文本和图像信息分别建模为独立视图,在共享知识图谱拓扑上构建关系感知图注意力编码器,通过关系约束下的邻域传播学习具有结构语义的多模态实体表示;随后利用门控后融合机制动态整合不同模态特征,并结合解码器完成链接预测任务。实验结果表明,所提出的方法在 3 个公开知识图谱数据集上取得了具有竞争力的性能,验证了模型的有效性。

尽管 MEGAT 在图结构信息与多模态信息的协同建模方面取得了一定进展,但仍存在进一步优化的空间。一方面,模型针对不同模态分别设置图注意力编码器,在保留模态特异性的同时,也增加了参数量与计算开销,其在大规模知识图谱上的应用效率仍有待提升。另一方面,模型采用固定的共享知识图谱拓扑进行多视图传播,虽然能够为不同模态提供统一的关系语境,但尚未充分刻画不同模态潜在的结构差异。在存在噪声边或缺失边的情况下,共享拓扑还可能导致错误信息跨模态传播,或限制相关实体之间的信息交互。未来可探索轻量化的参数共享策略,并引入模态特定的动态图构建、边可靠性估计与结构补全机制,以进一步提升模型的效率、鲁棒性和动态建模能力。

作者贡献声明:陈伟负责研究方案的总体设计、实验设计与实施,并完成论文撰写;庄福振负责研究指导,提出改进建议并参与论文修改。

参 考 文 献

[1] Hogan A, Blomqvist E, Cochez M, et al. Knowledge graphs[J]. ACM Computing Surveys, 2021, 54(4): 1−37

[2] Wang Q, Mao Z, Wang B, et al. Knowledge graph embedding: A survey of approaches and applications[J]. IEEE Transactions on Knowledge and Data Engineering, 2017, 29(12): 2724−2743

[3] Bordes A, Usunier N, Garcia-Duran A, et al. Translating embeddings for modeling multi-relational data[C]//Proc of Advances in Neural Information Processing Systems 26. Red Hook, NY: Curran Associates, 2013: 2787−2795

[4] Yang B, Yih W T, He X, et al. Embedding entities and relations for learning and inference in knowledge bases[C/OL]//Proc of the 3rd Int Conf on Learning Representations. San Diego: OpenReview, 2015[2026-07-18]. https://arxiv.org/abs/1412.6575

[5] Trouillon T, Welbl J, Riedel S, et al. Complex embeddings for simple link prediction[C]//Proc of the 33rd Int Conf on Machine Learning. New York: PMLR, 2016: 2071−2080

[6] Nickel M, Rosasco L, Poggio T. Holographic embeddings of knowledge graphs[C]//Proc of the 30th AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2016: 1955−1961

[7] Schlichtkrull M, Kipf T N, Bloem P, et al. Modeling relational data with graph convolutional networks[C]//Proc of the 15th European Semantic Web Conf. Cham: Springer, 2018: 593−607

[8] Nathani D, Chauhan J, Sharma C, et al. Learning attention-based embeddings for relation prediction in knowledge graphs[C]//Proc of the 57th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: ACL, 2019: 4710−4723

[9] Shang C, Tang Y, Huang J, et al. End-to-end structure-aware convolutional networks for knowledge base completion[C]//Proc of the 33rd AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2019: 3060−3067

[10] Vashishth S, Sanyal S, Nitin V, et al. Composition-based multi-relational graph convolutional networks[C/OL]//Proc of the 8th Int Conf on Learning Representations. Addis Ababa: OpenReview, 2020[2026-07-18]. https://openreview.net/forum?id=BylA_C4tPr

[11] Lu X, Wang L, Jiang Z, et al. MMKRL: A robust embedding approach for multi-modal knowledge graph representation learning[J]. Applied Intelligence, 2022, 52(7): 7480−7497

[12] Xu D, Xu T, Wu S, et al. Relation-enhanced negative sampling for multimodal knowledge graph completion[C]//Proc of the 30th ACM Int Conf on Multimedia. New York: ACM, 2022: 3857−3866

[13] Liu Y, Li H, Garcia-Duran A, et al. MMKG: Multi-modal knowledge graphs[C]//Proc of the 16th European Semantic Web Conf. Cham: Springer, 2019: 459−474

[14] Xie R, Liu Z, Luan H, et al. Image-embodied knowledge representation learning[C]//Proc of the 26th Int Joint Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2017: 3140−3146

[15] Wang Z, Li L, Li Q, et al. Multimodal data enhanced representation learning for knowledge graphs[C]//Proc of the Int Joint Conf on Neural Networks. Piscataway, NJ: IEEE, 2019: 1−8

[16] Cao Z, Xu Q, Yang Z, et al. OTKGE: Multi-modal knowledge graph embeddings via optimal transport[C]//Proc of Advances in Neural Information Processing Systems 35. Red Hook, NY: Curran Associates, 2022: 39090−39102

[17] Zhang Y, Chen Z, Liang L, et al. Unleashing the power of imbalanced modality information for multi-modal knowledge graph completion [C]//Proc of the 2024 Joint Int Conf on Computational Linguistics, Language Resources and Evaluation. Torino, Italy: ELRA and ICCL, 2024: 17120−17130

[18] Wang Z, Zhang J, Feng J, et al. Knowledge graph embedding by translating on hyperplanes[C]//Proc of the 28th AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2014: 1112−1119

[19] Lin Y, Liu Z, Sun M, et al. Learning entity and relation embeddings for knowledge graph completion[C]//Proc of the 29th AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2015: 2181−2187

[20] Ji G, He S, Xu L, et al. Knowledge graph embedding via dynamic mapping matrix[C]//Proc of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th Int Joint Conf on Natural Language Processing. Stroudsburg, PA: ACL, 2015: 687−696

[21] Sun Z, Deng Z H, Nie J Y, et al. RotatE: Knowledge graph embedding by relational rotation in complex space[C/OL]//Proc of the 7th Int Conf on Learning Representations. New Orleans: OpenReview, 2019[2026-07-18]. https://openreview.net/pdf?id=HkgEQnRqYQ

[22] Zhang Z, Cai J, Zhang Y, et al. Learning hierarchy-aware knowledge graph embeddings for link prediction[C]//Proc of the 34th AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2020: 3065−3072

[23] Chao L, He J, Wang T, et al. PairRE: Knowledge graph embeddings via paired relation vectors[C]//Proc of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th Int Joint Conf on Natural Language Processing. Stroudsburg, PA: ACL, 2021: 4360−4369

[24] Yu L, Luo Z, Liu H, et al. TripleRE: Knowledge graph embeddings via tripled relation vectors[J]. arXiv preprint, arXiv: 2209.08271, 2022

[25] Nickel M, Tresp V, Kriegel H P. A three-way model for collective learning on multi-relational data[C]//Proc of the 28th Int Conf on Machine Learning. Madison: Omnipress, 2011: 809−816

[26] Liu H, Wu Y, Yang Y. Analogical inference for multi-relational embeddings[C]//Proc of the 34th Int Conf on Machine Learning. New York: PMLR, 2017: 2168−2178

[27] Kazemi S M, Poole D. SimplE embedding for link prediction in knowledge graphs[C]//Proc of Advances in Neural Information Processing Systems 31. Red Hook, NY: Curran Associates, 2018: 4284−4295

[28] Balazevic I, Allen C, Hospedales T. TuckER: Tensor factorization for knowledge graph completion[C]//Proc of the 2019 Conf on Empirical Methods in Natural Language Processing and the 9th Int Joint Conf on Natural Language Processing. Stroudsburg, PA: ACL, 2019: 5185−5194

[29] Zhang W, Paudel B, Zhang W, et al. Interaction embeddings for prediction and explanation in knowledge graphs[C]//Proc of the 12th ACM Int Conf on Web Search and Data Mining. New York: ACM, 2019: 96−104

[30] Zhang S, Tay Y, Yao L, et al. Quaternion knowledge graph embeddings[C]//Proc of Advances in Neural Information Processing Systems 32. Red Hook, NY: Curran Associates, 2019: 2735−2745

[31] Cao Z, Xu Q, Yang Z, et al. Dual quaternion knowledge graph embeddings[C]//Proc of the 35th AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2021: 6894−6902

[32] Bastos A, Singh K, Nadgeri A, et al. HopfE: Knowledge graph representation learning using inverse Hopf fibrations[C]//Proc of the 30th ACM Int Conf on Information and Knowledge Management. New York: ACM, 2021: 89−99

[33] Bordes A, Glorot X, Weston J, et al. A semantic matching energy function for learning with multi-relational data[J]. Machine Learning, 2014, 94(2): 233−259

[34] Socher R, Chen D, Manning C D, et al. Reasoning with neural tensor networks for knowledge base completion[C]//Proc of Advances in Neural Information Processing Systems 26. Red Hook, NY: Curran Associates, 2013: 926−934

[35] Liu Q, Jiang H, Ling Z H, et al. Probabilistic reasoning via deep learning: Neural association models[J]. arXiv preprint, arXiv: 1603.07704, 2016

[36] Dettmers T, Minervini P, Stenetorp P, et al. Convolutional 2D knowledge graph embeddings[C]//Proc of the 32nd AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2018: 1811−1818

[37] Nguyen D Q, Nguyen T D, Nguyen D Q, et al. A novel embedding model for knowledge base completion based on convolutional neural network[C]//Proc of the 2018 Conf of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg, PA: ACL, 2018: 327−333

[38] Balazevic I, Allen C, Hospedales T. Hypernetwork knowledge graph embeddings[C]//Proc of Artificial Neural Networks and Machine Learning. Cham: Springer, 2019: 553−565

[39] Jiang X, Wang Q, Wang B. Adaptive convolution for multi-relational learning[C]//Proc of the 2019 Conf of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg, PA: ACL, 2019: 978−987

[40] Vashishth S, Sanyal S, Nitin V, et al. InteractE: Improving convolution-based knowledge graph embeddings by increasing feature interactions[C]//Proc of the 34th AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2020: 3009−3016

[41] Lacroix T, Usunier N, Obozinski G. Canonical tensor decomposition for knowledge base completion[C]//Proc of the 35th Int Conf on Machine Learning. New York: PMLR, 2018: 2863−2872

[42] Khosla P, Teterwak P, Wang C, et al. Supervised contrastive learning[C]//Proc of Advances in Neural Information Processing Systems 33. Red Hook, NY: Curran Associates, 2020: 18661−18673

[43] Lehmann J, Isele R, Jakob M, et al. DBpedia: A large-scale, multilingual knowledge base extracted from Wikipedia[J]. Semantic Web, 2015, 6(2): 167−195

[44] Vrandecic D, Krotzsch M. Wikidata: A free collaborative knowledgebase[J]. Communications of the ACM, 2014, 57(10): 78−85

[45] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[C]//Proc of the 3rd Int Conf on Learning Representations. San Diego: OpenReview, 2015[2026-07-18]. https://arxiv.org/abs/1409.1556

[46] Devlin J, Chang M W, Lee K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding[C]//Proc of the 2019 Conf of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg, PA: ACL, 2019: 4171−4186

[47] Zhang Y, Zhang W. Knowledge graph completion with pre-trained multimodal transformer and twins negative sampling[J]. arXiv preprint, arXiv: 2209.07084, 2022

[48] Duchi J, Hazan E, Singer Y. Adaptive subgradient methods for online learning and stochastic optimization[J]. Journal of Machine Learning Research, 2011, 12(7): 2121−2159



Chen Wei, born in 1998. PhD candidate. His main research interests include machine learning and recommendation systems.

陈　伟, 1998 年生。博士研究生。主要研究方向为机器学习、推荐系统。



Zhuang Fuzhen, born in 1983. PhD, professor, PhD supervisor. His main research interests include knowledge graph and transfer learning.

庄福振, 1983 年生。博士，教授，博士生导师。主要研究方向为知识图谱、迁移学习。