计算机科学
Computer Science
ISSN 1002-137X,CN 50-1075/TP

《计算机科学》网络首发论文

题目：　跨模态特征融合与对齐的虚假信息检测模型
作者：　杨晨光，卢记仓，郭嘉兴
网络首发日期：　2025-11-25
引用格式：　杨晨光，卢记仓，郭嘉兴．跨模态特征融合与对齐的虚假信息检测模型[J/OL]．计算机科学. https://link.cnki.net/urlid/50.1075.TP.20251125.0806.002

跨模态特征融合与对齐的虚假信息检测模型

杨晨光 卢记仓 郭嘉兴

信息工程大学 郑州 450001
（yangmg1996@163.com）

摘 要 虚假信息通常以夸大、歪曲或误导性的陈述进行传播，进而塑造负面社会舆论，严重危害公共安全。当前虚假信息通常是多模态的，现有检测方法往往在各模态特征分别提取后进行融合，忽略了模态间的相关性，难以全面捕获细节及相关性信息，导致检测性能不够理想。针对这些问题，本文提出了基于跨模态特征融合与对齐的检测模型（CMFFA）。CMFFA 优化了特征提取、融合和分类的模式，从宏观和微观两个角度提取模态特征，通过注意力机制进行特征增强，并通过计算模态间相似度评估模态间歧义性，以自适应的进行跨模态特征融合。首先，采用预训练模型编码文本和图像的单模态特征和跨模态特征；然后，在进行跨模态特征融合时，通过模态间歧义性分析，自适应地调整跨模态特征的使用比例，以更好地实现跨模态特征的融合，进而提升虚假信息检测性能。实验结果表明，在公开的中英文虚假信息数据集上，与已有虚假信息检测方法对比，本文所提模型在 F1 值、精确率、召回率上均有明显提升，验证了本文模型的有效性和优越性。

关键词： 虚假信息检测；跨模态特征融合；模态间歧义性分析；注意力机制

中图法分类号 TP391

Fake News detection model based on Cross-modal Feature Fusion and Alignment

Yang Chenguang，Lu Jicang and Guo Jiaxing
Information Engineering University Zhengzhou 450001, China

Abstract Fake news is usually spread by exaggerated, distorted or misleading statements, which can shape negative social opinion and seriously endanger public safety. The current fake news is usually multi-modal, and the existing detection methods often fuse the features of each modality after extracting them separately, ignoring the correlation between modalities, which makes it difficult to fully capture the details and correlation information, resulting in unsatisfactory detection performance. To solve these problems, this paper proposes a detection model based on Cross-modal feature fusion and alignment (CMFFA). CMFFA optimizes the modes of feature extraction, fusion and classification, extracts modal features from macro and micro perspectives, enhances features through the attention mechanism, and evaluates the ambiguity between modalities by calculating the similarity between modalities, so as to adaptively perform cross-modal feature fusion. Firstly, the pre-trained model was used to encode the single-modal features and cross-modal features of text and image. Then, in the cross-modal feature fusion, the proportion of cross-modal features is adaptively adjusted through the ambiguity analysis between modalities, so as to better realize the fusion of cross-modal features and improve the performance of fake news detection. The experimental results show that, compared with the existing fake news detection methods on the public Chinese and English fake news datasets, the proposed model has significantly improved the F1 value, precision and recall rate, which verifies the effectiveness and superiority of the proposed model.

Keywords Fake news detection; Cross-modal feature fusion; Inter-modal ambiguity analysis; Attention mechanism

通信作者：卢记仓（lujicang@sina.com）

1 引言

社交媒体凭借着开放性和匿名性成为了用户分享和交流信息的重要场所，同时，由于其提供和传播在线信息的门槛低、速度快，也为虚假消息的迅速扩散提供了温床。《中国新媒体发展报告 2024》[1]指出，网络空间作为媒体发展和融合转型的基础设施，其健康生态关系主流意识形态引导的质量和成败。社交媒体上的虚假信息不仅威胁到网络空间安全，也可能在重大事件中影响舆论导向，对民众认知产生严重的干扰，从而对现实世界的政治秩序、经济、社会造成严重负面影响。例如，2016年美国总统大选期间，出现著名的“披萨门”假新闻，这一毫无根据的理论直接影响了选民认知，甚至改变了双方支持人数的比例，对大选的公平性造成了严重影响；2020年全世界范围内的新冠肺炎大流行期间伴随的“信息瘟疫”，许多带有误导性内容的新闻报道，通过社交媒体传播，导致社会经济紊乱，全国流行病预防效果减弱；2022 年“俄乌冲突”爆发期间，虚假信息在网络空间肆意传播，导致仇恨情绪无限蔓延，深刻地干扰了公众对冲突态势的认知与判断。

对于虚假信息的定义，国内外并没有唯一确定的标准，一种比较主流的定义为：虚假信息是指制造者以故意误导读者为目的，并能够通过一些其它来源信息证实为假的信息，通常具有故意性和可证实性[2,3]。虚假信息检测属于内容可信度检测研究范围，旨在信息传播的早期阶段有效地识别出虚假信息。该任务本质上可以看作是一个分类问题，即在给定输入的情况下，输出真或假标签。

当前，基于信息内容的虚假信息检测方法，越来越多的使用跨模态融合特征，在特征融合时具有一定的盲目性，导致融合不充分、歧义性高等问题，进而影响虚假信息检测效果。现有的基于跨模态融合的检测方法，往往通过预训练语言模型提取多模态全局特征，并简单地进行拼接（特征相“加”），这不仅缺乏可解释性，也可能会导致语义信息的丢失。同时，大多数方法在解决图文内容语义不一致的问题上，仅依靠简单的模态间特征歧义性分析，例如计算余弦相似度，将模态间歧义性固化为一个静态参数来指导模态间特征融合程度，这也会影响模型的检测结果。上述问题是当前社交媒体虚假信息检测领域亟待解决的难题。为此，本文通过构建一个既能在宏观层面把握多模态信息的全局关联，又能在微观层面深入剖析每个模态内部细节的复杂虚假线索的多维度、多层次模型，探索如何更有效、更准确、更有目的地进行多模态特征融合。具体来说，本文主要贡献：

1. 提出多通道跨模态特征提取方法，将文本和图像特征映射到统一语义空间的同时，保留单模态特征检测通道以应对高歧义场景；

2. 提出双层注意力跨模态融合模块，通过构建文本—图像双向联合注意力层，捕捉图像和文本之间的复杂相关性，从而提取多模态深度确定性信息，为处理虚假信息检测特征融合的复杂性提供了解决方案。

3. 提出一个新的模态歧义性评估方法，通过计算生成模态歧义分数，根据模态歧义程度动态调整单模态特征与跨模态特征使用比例。

2 相关工作

2.1 单模态虚假信息检测方法

该类方法主要基于文本模态进行检测，根据对文本建模的方法差异，可分为基于文本序列和基于文本结构两类方法。

2.1.1 基于文本序列的检测方法

基于文本序列的方法通常以字、词、句及其他文

本特征为建模对象，然后借助于循环神经网络、预训练语言模型等方法对序列的处理优势，揭示虚假信息内容中潜在的逻辑冲突或叙事不连贯[4]。在单词级，Wang 等[5]利用 CNN 和双向长短期记忆网络(Bidirectional long short-term memory network, Bi-LSTM)整合词嵌入以检测虚假信息。Chawda 等[6]将递归卷积神经网络(Recurrent convolutional neural network, RCNN)和 LSTM 应用在虚假信息检测中，捕获相邻单词之间的上下文依赖关系。Nishant 等[7]首次将 BERT 和 LSTM 模型结合，利用 BERT 的上下文单词表示能力与 LSTM 的记忆模式相结合，以发挥它们强大的捕捉新闻标题中的语义和长距离关系的能力，从而根据新闻标题进行虚假信息分类，但该方法容易掉入标题陷阱，从而影响检测效果。在句子级，Ahn 等人[8]使用预训练语言模型 BERT[9]，通过设计的词片模型针对韩文进行虚假信息检测。在段落级，Yu 等[10]基于 CNN 模型学习段落嵌入表征以提取信息的高级文本特征。Nabbel 等[11]在 Nishant 等[7]基础上进行改进，采用了基于 BERT 模型并结合渐进式训练策略的改进方案，以提高分类准确率和模型的鲁棒性，使模型能够逐步学习并完善其对区分事实报道与伪造内容的语言细微差别的理解。

2.1.2 基于文本结构的检测方法

基于文本结构的方法通常将文本语义结构化，这种方法能够显著增强模型对社交媒体帖子内容进行细粒度语义分析的能力，有效克服传统序列模型在捕捉词汇间长距离语义依赖方面的局限性。例如，徐凡等人提出知识图谱驱动的图卷积网络模型（KDRD）[12]将文本转换为语义-实体图，其中节点不仅包含原贴文中的词语，而且利用世界知识和语言知识库扩展了语义词，从而丰富了原社交媒体文本的表示。梁宵等人[13]将文本转换成句法依存图，并通过设计子图注意力聚合和关键词去偏模块，有效提升了虚假信息的检测性能。

这些基于文本内容的检测方法虽在虚假信息早期检测方面取得了一定的进展，但随着虚假信息内容的不断变化，该方法局限性越来越显著，主要体现在两个方面：一是该类方法容易受到轻微的语义变化或虚假细节修改的影响；另一方面，由于社交媒体帖子的简短性质以及虚假信息的多样性和隐蔽性，仅依赖单一文本模态数据很难全面捕捉到有效识别虚假信息的关键线索。

2.2 多模态虚假信息检测方法

随着新媒体时代的到来，社交媒体信息中越来越多的包含图片、视频和音频等多媒体信息，也是当前新闻的一种主要传播渠道。基于多模态的虚假信息检测核心在于文本与图像间信息关联性的挖掘，这种关联性信息不仅包括图文一致性信息和互补性信息，同时还包括图文矛盾或冲突的歧义性信息。

2.2.1 基于多模态特征融合的检测方法

早期的多模态虚假信息检测方法侧重于结合不同模态的特征来提高检测精度。这些方法利用多模态数据的互补性来提取更多样化和有弹性的特征。例如，中科院曹娟团队等[14]提出一种基于神经元级别注意力机制的循环神经网络来融合图文信息，这种方法侧重于多模态内容的单向增强，即在文本引导下突出重要的图片区域，但该模型泛化性较差，对新事件的信息识别准确度较低。EANN 模型[15]将 Text-CNN[16]与 VGG-19[17]相结合，引入对抗神经网络，设计了事件判别器来学习各类事件之间相同特征，以提高模型泛化能力，但忽略了模态之间的相关性。MKEMN[18]通过构建多模态知识感知网络，学习社交媒体帖子的多模态表示，并从现实世界的知识图谱中检

索外部知识，以补充帖子短文本的语义表示，同时还构建了一个事件记忆网络，用来提取事件不变特征并将其存储到全局记忆中，从而提高新事件下模型泛化能力，增强虚假信息检测的准确性。MCAN[19]将多个 co-attention 层堆叠在一起以融合多模态特征并学习多模态之间的相互依赖关系，但该方法易受情绪化词汇影响，可能误判客观但情绪化的真实信息。HMCAN[20]利用多模态上下文信息和层次编码网络捕获文本信息的层次语义，从而增强信息的多模态表示，但该方法文本特征的层次化语义提取仍有改进空间，可能忽略局部语义细节模型复杂，训练和推理所需资源较大。CARMN[21]结合跨模态注意力残差网络[22]和多通道卷积神经网络[23]，保持不同模态的独特信息，减少模态差异带来的噪声，但该方法使用深度注意力堆叠和多通道卷积神经网络可能增加推理延迟，难以满足大规模流式检测需求。MPFN[24]提出了一种新的多模态渐进融合网络，通过 Swin - Transformer[25]同时考虑不同层次图像采样的深层和浅层信息，并设计了一个多层感知器（MLP）混频器来集成不同层次的视觉特征和文本特征，但其依赖大规模多模态预训练，对小语种或垂直领域（如医疗）适配性较差。

这些研究表明，融合多模态特征可以提高模型的性能。然而，仅仅融合视觉和文本特征并不能始终保证准确的信息提取。同时，文本和图像之间往往存在潜在的不一致性，直接融合特征而没有正确对齐会导致信息失真，导致特征冗余，融合性能差，给模型训练带来挑战。

2.2.2 基于多模态特征一致性的检测方法

图文不一致性增加了模型融合和对齐多模态数据特征的难度。因此，模型必须从误导或冲突的内容中联合筛选出有价值的细节信息。一部分方法侧重于度量模态之间的相似性[26-28]。Zhou 等[26]提出的 SAFE 模型采用余弦相似度衡量图文特征的一致程度，并将此度量融入损失函数以指导模型训练。Chen 等[27]受信息论启发，首次提出使用跨模态歧义性学习改进 SAFE 的相似性度量方法，他们设计的 CAFE 模型能比较简单地实现自适应地聚合单模态特征和多模态特征关联，从而提高假信息检测精度。上述 2 种方法通过对图像和文本进行压缩，只有有限的数据输入到网络中，多模态特征对齐无法保证。为此，MMFN[28]通过提取不同模态多粒度特征并计算单模态相似度，自适应地辅助多模态分类。近年来，学者们进一步探索了如何通过改进特征对齐技术来提高一致性检测的效果。Wang 等[29]提出 MSACA 模型，为每个多模态数据构建了分层的多尺度图像，增强了文本和图像在潜空间中的语义一致性，并利用注意力模块以端到端的方式选择确定性的嵌入。Kumar 等[30]引入了对抗性学习机制，以创建与文本密切一致的视觉特征，提高了模型在多模态一致性检测中的鲁棒性。He 等[31]提出了一种基于无证据多级融合的跨模态检测方法，该方法通过利用跨模态对齐处理来解决语义不一致问题，并利用注意力机制对文本和图像特征进行多级融合，无需其他证据特征的辅助，从而进一步增强特征的表达能力。Liu 等[32]提出模态交互混合专家网络，旨在通过交互门控机制来增强检测能力，明确地对模态交互进行建模，从而为不同的模态交互场景定制一种模态融合策略。

上述方法在识别模态间差异方面展现出了巨大潜力，但仍存在一个核心局限：跨模态特征交互的技术瓶颈。例如，模态间歧义性导致的特征融合缺陷，在图文虚假信息检测中，跨模态注意力机制对深层语义关联建模不足，易忽略细粒度实体矛盾；再如，跨模态特征的映射误差与信息

损失，多模态数据需统一映射至公共空间（如BEV视图或共享嵌入层），该过程易因投影偏差丢失关键信息。

上述单模态和多模态的方法都是基于信息内容特征的检测方法，除此之外，按照检测所使用的特征分，还有基于社交上下文和基于知识驱动等方法。例如，Hu 等[33]提出了一种结合信息文本内容的双向图卷积神经网络,它将消息内容与传播路径相结合，在信息传播过程中,双向图卷积神经网络随后学习了事件传播网络的特征表征，将这些表示与原始文本内容特征合并，实现了综合的信息检测增强了检测性能。Chen 等[34]利用语言模型生成外部知识，仅将其作为训练集的补充信息，并引入了一个基于图的语义感知特征对齐模块来解决知识矛盾，同时引入了一个基于信息瓶颈的知识蒸馏模块以确保在推理过程中隐式地生成这些特征。

3 跨模态特征融合与对齐的虚假信息检测模型

针对已有跨模态虚假信息检测中存在的模态融合不充分、模态间歧义性引起检测偏差等问题，本文从多通道特征编码、跨模态双层联合注意力感知、跨模态歧义分析等角度出发，提出CMFFA 模型，通过跨模态特征融合与对齐来解决多模态虚假信息检测中常见的模态特征利用不充分和相互干扰的歧义性问题。模型包含 4 个模块，包括跨模态特征提取、跨模态特征融合、跨模态歧义分析和虚假信息检测，所提模型整体框架如图 1 所示。

3.1 跨模态特征提取模块

多模态信息数据集表示为 $P=[T, V]$，其中 T 表示文本输入集，V 表示图像输入集。为了提取有助于分类任务的综合多模态信息，跨模态特征提取模块将 BERT 和 ResNet-50[22]分别作为文本模态和图像模态的单模态特征编码器，将模态一致的对比语言-图像预训练模型（Contrastive Language-Image Pre-training, CLIP）[35]编码器作为跨模态特征编码器。

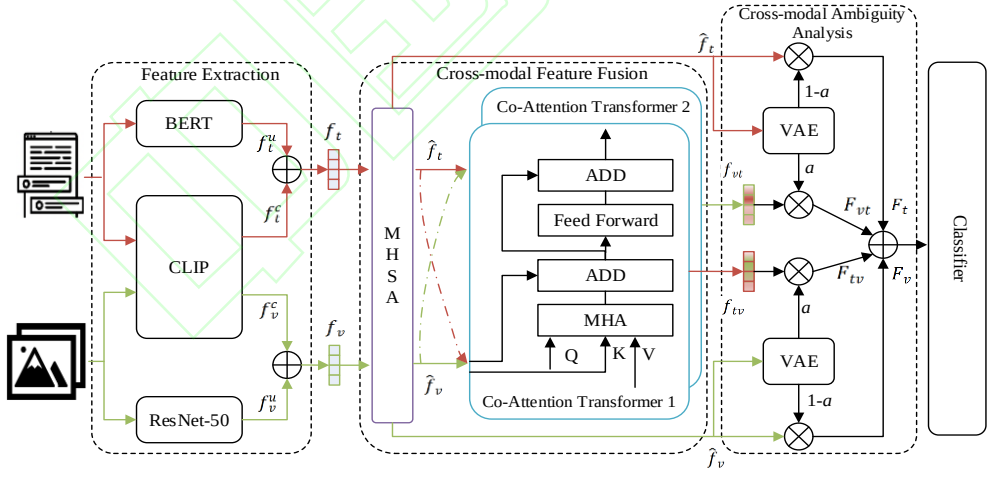

图 1 CMFFA 模型框架图

Fig.1 The framework of the CMFFA

在跨模态特征提取阶段，给定一篇信息内容$p = [t, v] \in P(t \in T, v \in V)$使用单模态特征编码器和跨模态特征编码器分别提取文本和图像的单模态特征和跨模态特征，f_t^u和f_t^c分别表示文本模态的单模态特征和跨模态特征，f_v^u和f_v^c分别表示图像模态的单模态特征和跨模态特征。

3.1.1 文本特征编码

BERT 是一种流行的基于 Transformer[36]的预训练语言模型，在大型语料库上使用无监督学

习进行训练，在许多 NLP 下游任务中取得了优异的结果。因此，我们使用 BERT 模型对来自信息内容 p 的文本特征进行编码。

p 的文本内容，是一个表示为 $t=\{t^1,t^2,\cdots,t^n\}$的单词序列列表，其中 n 是单词的数量。将 BERT 应用于 t 后，得到编码后的文本单模态特征$f_t^u \in \mathbb{R}^{d_b}$，如式（1）所示。

$$f_t^u = BERT(t) = \{t_b^1, t_b^2, \cdots, t_b^n\} \quad (1)$$

其中t_b^n表示文本嵌入的最后一个隐藏状态的输出，d_b是词嵌入的维度。

3.1.2 图像特征编码

信息内容 p 中的图像信息 v 使用 ResNet-50 预训练模型编码，获取图像模态特征表示。p 的图像内容，是一个表示为$v=\{v^1,v^2,\cdots,v^s\}$的图像序列，其中 s 是图片的数量。将 ResNet-50 应用于 v 后，得到编码后的图像单模态特征$f_v^{iu} \in \mathbb{R}^{d_r}$，如式（2）所示。

$$f_v^u = ResNet(v) = \{v_r^1, v_r^2, \cdots, v_r^s\} \quad (2)$$

其中，d_r 是视觉嵌入的维度，$\{v_r^i | i = 1,2\cdots s\}$表示使用 *ResNet* 编码后的图像特征 v(*r* 用来表示 *ResNet* 模型，与原特征 v 区分开来）。

3.1.3 基于 CLIP 的跨模态特征编码

CLIP 是一个在大量数据上训练的多模态预训练模型，其核心思想是将图像和文本嵌入映射到共享语义空间。

为了更有效地从不同模态中提取确定性信息，将 CLIP 预训练编码器和单模态编码器相结合，得到基于 CLIP 预训练的文本和图像的跨模态特征，表示为$f_t^c \in \mathbb{R}^{d_c}$和$f_v^c \in \mathbb{R}^{d_c}$，其中$d_c$是特征向量的长度。

$$f_t^c = CLIP(t) \quad (3)$$

$$f_v^c = CLIP(v) \quad (4)$$

3.1.4 单模态特征融合

根据 Chen[27]等对模态歧义性问题的定义，为了更好的处理跨模态融合在模态歧义度高时特征表示能力下降的问题，有针对性地使用单模态特征，仍然需要保留单模态检测通道，分别将文本和图像的跨模态和单模态特征进行融合，得到单模态融合特征$f_t \in \mathbb{R}^{d_f}$和$f_v \in \mathbb{R}^{d_f}$，其中，$d_f = d_c + d_l$是单模态融合特征的维度，d_l是文本模态和图像模态通过线性层映射到相同的维度。（经跨模态歧义分数加权后得到单模态检测特征F_t和F_v，3.3 节中将具体表述）。

具体来说，对于单文本模态，先在 token 级维度上将文本模态特征f_t^u池化成一个特征向量，并将其与f_t^c串连，得到单文本模态融合特征，如式（5）所示：

$$f_t = \text{concat}(f_t^u, f_t^c) \quad (5)$$

同理，对于图像模态，我们将池化的图像特征f_v^u与f_v^c串连起来，得到单图像模态融合特征，如式（6）所示：

$$f_v = concat(f_v^u, f_v^c) \quad (6)$$

3.2 跨模态特征融合

多模态信息往往无法直接根据内容判断真假，使得仅根据显性信息进行真假分类具有挑战性。在现实生活中，当人们面对难以理解的新闻帖子时，通常会反复阅读新闻帖子，并结合图片内容来帮助他们理解。基于此，本文设计了 2 个跨模态联合注意力融合模块，试图从多层次的跨模态特征融合的角度出发，更好捕获隐藏在文本和图像中的隐含信息。

3.2.1 单模态特征增强

在进行跨模态特征融合之前，使用多头自注意力（MHSA）[12]来增强每个单模态特征的表示。MHSA 将相同的嵌入作为多头注意力（MHA）模块的输入，使模型能够专注于输入序列中的复杂关系。

以单文本模态特征为例，首先被投影为第 i 个头的 *query*、*key* 和 value：

$$Q_i^t = f_t W_i^{sq}, K_i^t = f_t W_i^{sk}, V_i^t = f_t W_i^{sv} \quad (7)$$

其中Q_i^t、K_i^t和V_i^t分别是第 i 个头的查询、键和值嵌入。W_i^{sq}、W_i^{sk}和$W_i^{sv} \in \mathbb{R}^{d \times d_n}$是可学习的权重矩阵。$d$ 是输入特征长度，d_n是投影的输出维度。然后，每个头的自注意力计算如下：

$$h_i^t = softmax(\frac{Q_i^t K_i^t}{\sqrt{d_h}}) V_i^t \quad (8)$$

$$h_t = h_1^t \oplus h_2^t \oplus \cdots \oplus h_n^t \quad (9)$$

$$M_t = h_t W_i^o \quad (10)$$

$d_h = d_n/n$，n 为头的个数，$\oplus$表示拼接操作，$W_i^o \in \mathbb{R}^{d \times d_n}$为输出线性变换。

然后将自注意力模块输入由两个全连接（FC）线性层和 ReLU 激活函数[37]组成的前馈神经网络[38]（Feedforward Neural Network, FNN）后，得到增强的单文本模态特征如下：

$$\hat{f}_t = MHA(f_t, f_t, f_t) = (M_t + f_t) + FNN(M_t + f_t) \quad (11)$$

类似地，可以得到增强的单图像模态特征$\hat{f}_v$。

3.2.2 跨模态特征融合

由于我们在现实生活中阅读新闻等内容时，往往将文本和图像内容结合起来，反复观察对比，即文本和图像内容相互补充。因此，设计 2 个联合注意力层，即图像特征增强和文本特征增强的跨模态联合注意力层，以充分吸收跨模态的语义信息进行跨模态特征融合。

具体而言，对于单文本特征，利用增强后的图像特征$\hat{f}_v$计算查询矩阵，将文本特征$\hat{f}_t$投影为键矩阵和值矩阵。MHA 的第 i 个头的输入表示如下：

$$Q_i^{tv} = \hat{f}_v W_i^{cq}, K_i^{tv} = \hat{f}_t W_i^{ck}, V_i^{tv} = \hat{f}_t W_i^{cv} \quad (12)$$

其中Q_i^{tv}、K_i^{tv}和V_i^{tv}是图像特征增强的文本跨模态联合注意力层第 i 个头的 *query*、*key* 和 *value*，W_i^{cq}、W_i^{ck}和W_i^{cv}是可学习的投影矩阵。

通过 MHA 获得图像增强的文本特征：

$$f_{tv} = \text{MHA}(\hat{f}_v, \hat{f}_t, \hat{f}_t) \quad (13)$$

类似地，对于图像特征，利用增强后的文本特征$\hat{f}_t$计算查询矩阵，将图像特征$\hat{f}_v$投影为键矩阵和值矩阵，最终，文本增强的图像特征表示为：

$$f_{vt} = MHA(\hat{f}_t, \hat{f}_v, \hat{f}_v) \quad (14)$$

3.3 跨模态歧义分析

使用基于 CLIP 的 KL 散度[39]进行模态歧义分析，生成模态歧义分数，以自适应进行跨模态特征对齐。具体来说，模态歧义分数较大表明文本和视觉分布之间存在实质性差距，考虑到这种不一致性，仅依赖单一模态的预测变得不那么可靠。在这种情况下，应该更多地关注融合后的多模态特征，这可能包含了深层隐含的决定性因素。相反，当模态歧义分数较小时，单模态特征在虚假信息检测中应赋予更大的权重。

为了计算文本特征$\hat{f}_t$和图像特征$\hat{f}_v$的 KL 散度，首先利用变分自编码器（Variational autoencoder, VAE[40]）学习这些数据的分布；VAE 是一种生成模型，可以将数据编码为潜空间的均值 μ和方差 σ，并通过重新参数化生成样本。给定文本特征$\hat{f}_t$和图像特征$\hat{f}_v$，其隐变量的后验可以表示为：

$$q(z_t|\hat{f}_t) = \mathcal{N}(z_t|\mu(\hat{f}_t), \sigma(\hat{f}_t)) \quad (15)$$

$$q(z_v|\hat{f}_v) = \mathcal{N}(z_v|\mu(\hat{f}_v), \sigma(\hat{f}_v)) \quad (16)$$

式（15）、（16）中，$\mathcal{N}$为随机变量z的高斯分布，其均值为μ，方差为σ，z_t和z_v分别为f_t^{uni}和f_v^{uni}的潜在变量。$\mu(f_t^{uni})$、$\sigma(f_t^{uni})$、$\mu(f_v^{uni})$和$\mu(f_v^{uni})$可以从独立的 MLP 层中得到。计算两个分布的 KL 散度的平均值，并经过 sigmoid 函数映射成 0 到 1 之间的值，作为模态歧义分数 a：

$$a_1 = \mathcal{D}_{KL}((q(z_t|f_t^c))||(q(z_v|f_v^c))) \quad (17)$$

$$a_2 = \mathcal{D}_{KL}((q(z_v|f_v^c))||(q(z_t|f_t^c))) \quad (18)$$

$$a = sigmoid(\frac{a_1 + a_2}{2}) \quad (19)$$

其中$\mathcal{D}_{KL}$表示 KL 散度，sigmoid（·）是将模态间歧义分数映射到 0 到 1 之间值的激活函数。

然后，将模态间歧义分数 a 作为权重，动态感知单模态特征和跨模态融合特征在虚假信息检测过程中的重要性。

$$\begin{cases} F_t = (1-a)f_t^{uni} \\ F_v = (1-a)f_v^{uni} \\ F_{tv} = af_{tv} \\ F_{vt} = af_{vt} \end{cases} \quad (20)$$

3.4 虚假信息检测分类

检测模型使用交叉熵损失函数进行分类，并通过对抗训练增强模型的鲁棒性。经过模态歧义分数加权的单模态融合特征和跨模态特征被连接起来作为最终的信息表示：

$$F = concat(F_t, F_v, F_{tv}, F_{vt}) \quad (21)$$

然后将特征 F 输入到分类器中：

$$\hat{y} = softmax(FCs(F)) \quad (22)$$

其中，$FCs(\cdot)$是虚假信息分类器，它除最后一层全连接层外，由五个具有 Relu 激活函数的全连接层组成。

采用交叉熵损失函数作为目标函数：

$$\mathcal{L}_{\text{cls}} = ylog(\hat{y}) + (1-y)log(1-\hat{y}) \quad (23)$$

其中，y 代表真实标签，1 表示假，0 表示真。本文方法实现过程参考算法 1。

算法 1 CMFFA 虚假信息检测算法

输入：多模态新闻数据集 $P = [T, V]$，其中 T 表示文本输入，V 表示图像输入；新闻的真假标签 y，0 为真，1 为假；*epoch* 为迭代次数。

输出：分类结果$\hat{y}$。

1.**for** i = 1 to *epoch* **do**;

2.通过公式（1）、（2）生成文本和图像模态的单模态特征：f_t^u、f_v^u；

3.通过公式（3）、（4）生成文本和图像模态的 基于 CLIP 模型的跨模态特征：f_t^c、f_v^c；

4.通过公式（5）、（6）分别得到单文本模态和单图像模态融合特征f_t和f_v；

5.通过公式（7）—（11）生成得到增强的单文本模态和单图像模态特征：$\hat{f}_t$、$\hat{f}_v$；

6.通过公式（12）—（14）分别生成图像增强的文本特征f_{tv}和文本增强的图像特征f_{vt}；

7.通过公式（15）—（19）得到模态歧义分数 a；

8.通过公式（20）得到经模态歧义分数 a 加权的单模态融合特征F_t、F_v和跨模态增强特征F_{tv}、F_{vt}；

9.通过公式（21）将F_t、F_v、F_{tv}、F_{vt}拼接得到信息 P 最终的特征表示 F;

10.通过公式（22）预测分类结果；

11.通过最小化目标函数$\mathcal{L}_{cls}$（公式（23））以更新整个网络；

12.**end for**

13.return $\hat{y}$.

4 实验

4.1 实验数据集

本文使用的 3 个公开中英文数据集进行实验，分别是中文的 Weibo[14]、英文的 Pheme[41]和 Gossipcop[42],数据集内容如表 1 所列。

表 1 数据集统计

Table 1 Datasets statistics

	Weibo	Pheme	Gossipcop
Fake News	3630	590	4547
Real News	3479	1563	10126
image	6844	2018	7542

（1）Weibo 数据集是最早由中科院曹娟团队设计，且已广泛应用于虚假信息检测的中文数据集。数据来自中国的权威新闻媒体新华社。在实验中，过滤掉缺乏完整文本或图像的新闻文章，并去除重复和低质量的文本，以保证数据集的质量，按 7:1:2 的比例划分为训练集、验证集和测试集。

（2）Pheme 数据集由来自 Twitter 平台的推文组成，涵盖了 9 个不同事件的突发新闻，去除与用户相关的社交结构信息后，保留文本、图像

信息，按 7:1:2 的比例划分为训练集、验证集和测试集。

（3）Gossipcop 数据集来源于一个娱乐新闻事实核查网站，是虚假信息检测大型数据集 FakeNewsNet[42]的一部分。Gossipcop 数据集包含 14673 条新闻，其中包括 10126 条真实新闻和 4547 条假新闻，按 7:1:2 的比例划分为训练集、验证集和测试集。

4.2 用于实验的基线模型

实验选用已开源的具有代表性的多模态虚假信息检测模型作为实验的基线模型，如下所示：

（1）CARMN[21]：该模型通过跨模态注意力残差网络和多通道卷积神经网络进行虚假信息检测，其中跨模态注意力残差网络有选择地从另一个来源的模态中提取与目标模态相关的相关信息，融合不同模态之间的相关信息，同时保持每个模态的独特特性。多通道卷积神经网络能够通过同时从原始和融合的文本信息中提取文本特征表示，来减轻由跨模态融合组件可能产生的噪声信息的影响，从而提高虚假信息检测准确率。

（2）CAFE[27]：提出了一种模糊感知的多模态方法，首次提出跨模态歧义学习理论，自适应聚合跨模态与单模态特征，通过注意力加权抑制无关噪声，并使用单模态公共语义共享辅助任务解决模态不一致导致的误分类问题。但该方法对小语种或专业领域（如科学谣言）数据适应性较差，需大规模预训练数据支持。

（3）MMFN[28]：首次采用基于 CLIP 模型加权的方法，动态平衡单模态与多模态分支对虚假信息检测的贡献率，优化了模态歧义性评估方法，同时通过分层融合文本与图像的细粒度（token 级）和粗粒度（全局语义）特征，较好解决多模态信息特征利用不足问题。

（4）MSACA[29]：提出了一种基于多尺度语义对齐和跨模态注意力的虚假新闻检测框架，主要解决多模态信息（文本、图像）融合中的语义鸿沟问题。通过分层特征提取（如卷积神经网络的不同层）捕捉图像从全局构图到局部细节的语义信息，动态匹配文本描述的粒度，并在潜在空间中计算图文片段的相关性权重，抑制无关噪声，从而有效识别对信息有用的图像，增强模态间一致性特征的提取，提高虚假信息检测准确率。

4.3 实验设置

4.3.1 数据预处理

首先，对于原始的 Weibo 数据集，一篇新闻文章的文本可能伴有多张图片或者没有图片。为了构建标准的新闻样本，将每篇新闻文章保留一张高质量的图片，或者过滤掉没有图片的文章。然后对文本和图像数据进行预处理，使用 BERT tokenizer 对中英文文本分别处理：中文按字切分，英文按 WordPiece 切分。图像统一调整为 224 × 224 分辨率大小。此外，由于有些文本内容语法规则可能混乱，导致模型准确率下降，我们对文本嵌入进行了投影梯度下降（Projected Gradient Descent，PGD）对抗性训练[43]，以增强模型的鲁棒性，实现更准确的预测。在此过程中，需要计算对抗样本损失和反向传播梯度来更新模型参数。

4.3.2 参数设置

在单模态编码器中，所提供的词向量被用于维数为 300 的词嵌入[44]。对于跨模态编码器，文本被截断或填充为 77 个 token。使用的预训练 CLIP 模型是“CLIP-ViT-B/16”，尺寸为 512。可学习内存信息的长度设置为 50。实验中，将训练时的批大小设置为 64，测试时的批大小设置为 50，本文提出的 CMFFA 模型使用 Adam 算法[45]对损失函数进行优化，Weibo 数据集的学习

率为 0.001，Pheme 数据集的学习率为 0.002。MHA机制的头编号设置为8。在所有实验中，模型在单个 NVIDIA RTX 3080 GPU 上训练了 20 epoch，并且随机种子设为固定的 42，以确保结果的再现性。

4.3.3 基线模型设置

在基线方法参数设置方面，为确保对比实验的公平性，所有基线模型尽量严格遵循原论文报告的最优参数配置，并通过开源代码验证实现一致性。CARMN 模型使用原文提供的开源代码，按原文进行参数设置，其中文本嵌入维度 300，图像输入同本实验一致调整为 224×224（ResNet-50），并通过归一化解决其跨模态注意力层梯度消失问题。CAFE 模型，将原文余弦相似度度量的系数 τ 的取值定为 0.5。 MMFN 的 CLIP 特征加权模块调整 ViT-B/16 的输入尺寸为 224×224（原论文未明确说明）。 MSACA 的多尺度图像处理模块由于消耗显存较大，将尺寸设为全局（224×224）、区域（112×112）、局部（56×56）。所有基线模型均采用 Adam 优化器，跨模态注意力头数统一为 8，学习率保持一致（Weibo: 0.001, Pheme: 0.002），并在相同硬件环境（RTX 3060 GPU）下完成训练，随机种子固定为 42 以保证可复现性。参数选择依据验证集 F1 值进行早停判定（连续 3 轮无提升终止训练）。

4.4 实验环境

CPU R7-5800，显卡 NVIDIA GeForce RTX 3080 。编程语言：Python 3.12.7。深度学习框架：Pytorch 2.6。

4.5 实验评价指标

本文选取准确率（Accuracy/Acc）和 F1 值作为评价指标。同时，对于真（Real）假（Fake）两种类型的信息，分别将精确率（Precision/Pre）、召回率（Recall/Rec）和F1值作为对比指标。真（Real）假（Fake）信息的精确率、召回率和 F1 值取平均值作为模态总体的精确率、召回率和F1值。

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \tag{24}$$

$$Precision = \frac{TP}{(TP+FP)} \tag{25}$$

$$Recall = \frac{TP}{(TP+FN)} \tag{26}$$

$$F1 = \frac{2*Precision*Recall}{(Precision+Recall)} \tag{27}$$

4.6 实验结果分析

CMFFA 模型在 3 个公开数据集上的准确率均优于所有比较方法，分别为 0.868、0.834 和 0.902。

首先对基线模型的实验结果体现的性能差异进行分析。CARMN 和 CAFE 方法的结果较差,主要是因为它们仅通过其特定模式的编码器提取单模态特征，内在地将异质性引入用于对齐和加权的特征中导致模态之间存在显著的语义鸿沟。此外，CAFE 通过余弦相似度衡量模态歧义性，在检测非对称语义冲突场景（如文本夸大但图像真实）可能效果一般。MMFN 模型首次使用 CLIP 模型进行模态歧义性学习，将模态歧义性作为虚假信息检测的重要依据，从而检测效果对比 CARMN 和 CAFE 有了较大提升，但由于其仅通过 CLIP 进行静态加权，无法动态调整权重，因此检测效果不如本文所提模型。MSACA 模型相较其他 3 个模型检测效果最好，主要原因是其通过跨模态注意力，试图识别判别信息真假最有效的图像，从而将文本信息与图像信息结合，提高识别的准确率。

表 2 实验结果对比

Table 2 Comparison of experimental results

Datasets	Models	Acc	Pre	Recall	F1	Fake News			Real News		
						Pre	Rec	F1	Pre	Rec	F1
Weibo	CARMN[21]	0.784	0.779	0.788	0.783	0.781	0.789	0.785	0.777	0.786	0.781
	CAFE[27]	0.733	0.728	0.741	0.734	0.732	0.743	0.737	0.724	0.739	0.732
	MMFN[28]	0.815	0.815	0.817	0.816	0.814	0.820	0.817	0.815	0.813	0.814
	MSACA[29]	0.846	0.838	0.846	0.842	0.835	0.848	0.841	0.840	0.844	0.842
	CMFFA(本文模型)	**0.868**	**0.864**	**0.867**	**0.866**	**0.852**	**0.862**	**0.857**	**0.877**	**0.872**	**0.874**
Pheme	CARMN[21]	0.746	0.725	0.762	0.743	0.713	0.758	0.735	0.736	0.765	0.750
	CAFE[27]	0.710	0.698	0.712	0.705	0.685	0.734	0.709	0.711	0.689	0.700
	MMFN[28]	0.785	0.779	0.795	0.787	0.776	0.793	0.784	0.782	0.796	0.803
	MSACA[29]	0.831	**0.831**	0.833	0.832	**0.826**	0.832	**0.829**	0.835	0.833	0.834
	CMFFA(本文模型)	**0.834**	0.830	**0.838**	**0.833**	0.821	**0.834**	0.827	**0.838**	**0.841**	**0.839**
Gossip cop	CARMN[21]	0.741	0.762	0.750	0.740	0.854	0.619	0.718	0.670	0.880	0.761
	CAFE[27]	0.867	0.785	0.674	0.717	0.732	0.490	0.587	0.837	0.857	0.847
	MMFN[28]	0.894	0.805	0.731	0.760	0.799	0.598	0.684	0.810	0.864	0.836
	MSACA[29]	0.887	0.857	0.705	0.766	0.816	0.538	0.648	0.897	0.871	0.884
	CMFFA(本文模型)	**0.902**	**0.908**	**0.788**	**0.839**	**0.905**	**0.673**	**0.772**	**0.911**	**0.902**	**0.906**

CMFFA 方法优于其他方法的主要原因有 3 个方面。（1）跨模态特征互补性。CMFFA 通过预训练的 CLIP 生成的共享语义空间特征与单模态特征形成互补，既保留模态内部的特异性（如文本逻辑矛盾、图像局部篡改痕迹），又强化跨模态关联（如图文语义一致性）。（2）多层次特征交互。跨模态特征融合模块通过双层联合注意力，实现了文本和图像特征的深度融合，从而促进模型对不同模态间的隐含语义的深入理解。（3）跨模态歧义动态分析。CMFFA 使用的跨模态歧义学习可以对不同模态特征进行加权聚合。使用基于 CLIP 的 KL 散度分数来评估模态间歧义性，将其作为多模态特征融合的权重，能自适应感知不同模态对检测的贡献，从而动态选择模态特征进行检测。

在所有实验中，CMFFA 模型在准确率、精密度、召回率和 F1 得分方面均排名第一，证明了其在虚假信息检测方面的有效性。

4.7 消融实验

为了进一步评估 CMFFA 中各个重要模块对整体性能的影响，在数据集上进行了 4 组实验。对于每个实验，删除不同的组件并重新训练模型。用于实验的 CMFFA 变体设置如下：

（1）CMFFA w/o A：去除模态间歧义分析模块，不使用歧义性分数。将单模态特征和多模态特征直接连接起来进行分类。

（2）CMFFA w/o E：去除单模态特征增强模块，直接进行跨模态特征融合。使用跨模态融合特征作为虚假信息分类器的输入。

（3）CMFFA w/o C：去除基于 CLIP 的跨模态特征，只使用模态内特征进行虚假信息检测。

（4）CMFFA w/o F：去除跨模态融合模块，采用单层跨模态特征融合获取多模态特征进行虚假信息检测。

表 3 消融实验结果

Table 3 Results of ablation experiments

Dataset	Method	Acc	F1-score	
			Fake News	Real News
Weibo	CMFFA	**0.868**	**0.857**	**0.874**
	CMFFA w/o A	0.802	0.809	0.812
	CMFFA w/o E	0.825	0.824	0.827
	CMFFA w/o C	0.833	0.841	0.836
	CMFFA w/o F	0.817	0.813	0.815
Pheme	CMFFA	**0.834**	**0.827**	**0.839**
	CMFFA w/o A	0.776	0.768	0.771
	CMFFA w/o E	0.805	0.804	0.810
	CMFFA w/o C	0.811	0.802	0.805
	CMFFA w/o F	0.783	0.799	0.792
Gossipcop	CMFFA	**0.902**	**0.772**	**0.906**
	CMFFA w/o A	0.794	0.688	0.787
	CMFFA w/o E	0.813	0.714	0.824
	CMFFA w/o C	0.849	0.736	0.850
	CMFFA w/o F	0.825	0.703	0.792

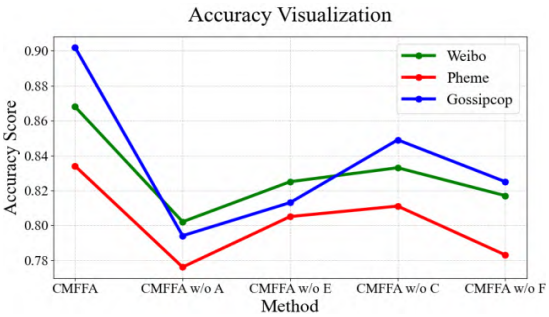

图 2 消融实验准确率统计图

Fig.2 Statistical plot of the accuracy of ablation experiments

消融实验结果如表 3 和图 2 所示，从中可以看出删除模型中的任何模块会导致检测结果下降。其中 CMFFA w/o A 结果最差，说明了使用模态歧义分数对跨模态特征使用的重要性。CMFFA w/o E 表现出性能下降,这有效验证单模态特征增强在特征融合中的重要性。CMFFA w/o C，实验结果比标准模型要差，说明将图像和文本嵌入映射到共享语义空间的效果是明显的。CMFFA w/o F 性能下降第二多，证明通过跨模态融合能够深入理解复杂的隐式信息，从而助于显著提高性能。

5 讨论

本文所提 CMFFA 模型虽然在 3 个公开数据集上的综合表现优于基线模型，但仍有一些技术层面的不足和可以改进之处，具体讨论如下：

（1）模型架构技术仍有局限。一方面，注意力机制还存在可优化空间。当前双层联合注意力机制虽然能够捕捉跨模态关联，但对细粒度语义对齐的建模仍不够精确，特别是在处理长文本与复杂图像场景时，注意力权重分布可能不够聚焦关键语义区域。另一方面，CLIP 编码器存在适应性瓶颈，依赖预训练的 CLIP 模型提取跨模态特征，可能导致领域适应性不足，特别是对专业领域(如医疗、金融)术语和特殊图像(如医学影像、工程图纸)的编码效果下降。

（2）特征处理技术仍有不足。一方面，模态歧义性评估上精确度可能不高，当前基于 KL 散度的歧义分数计算只考虑了特征分布差异，未能有效区分“语义冲突”与“信息互补”这两种不同性质的模态差异，可以进一步探索多维度模态歧义评价方法，结合语义相似度、逻辑一致性、情感极性等多个维度进行综合评判。另一方面，模型对抗训练策略还需要完善，本文方法仅利用 PGD 对抗训练[43]对文本模态进行对抗性训练，这可能使模型对精心设计的图像类对抗样本(如通过色彩、对比度、细节增强等，旨在不显著改变整体观感的前提下优化特定视觉特征的图像）的鲁棒性不足。

6 结束语

本文针对虚假信息检测中跨模态特征融合不充分和歧义性问题，提出基于跨模态特征融合与对齐的虚假信息检测技术，探索如何有效提取单模态特征，有目的地进行多模态特征融合，以解决当前基于信息内容的跨模态检测方法中常见的模态特征联系不紧密和歧义性问题。具体来说，在特征提取过程，将 CLIP 预训练模型作为跨模态编码器，与单模态编码器一起工作，以提取完整的信息。然后，在跨模态特征融合过程中实现单模态特征增强和跨模态特征融合，以挖掘模态中隐含的含义。最后，跨模态歧义分析模块根据歧义性分数自适应地使用各种特征。在 3 个公开使用的数据集上的实验表明，CMFFA 检测效果更出色。

本文不足之处在于，社交上下文信息利用不足，未利用用户评论、传播路径等社交特征，而社交媒体的新闻中通常包含丰富的社交上下文信息，若把这些信息特征集成到模型中，对模型的偏好感知和检测能力会有较大提升，这可能是未来关注的方向。此外，未来研究还可以在小样本

学习、领域迁移、时空动态建模、多模态生成式检测等角度进一步优化模型的检测效果。

参考文献

[1] HU Z R, HUANG C X. China New Media Development Report No.15[M]// Beijing: Social Sciences Academic Press, 2024. (in Chinese)
胡正荣, 黄楚新. 中国新媒体发展报告 No.15[M]// 北京: 社会科学文献出版社, 2024.

[2] BONDIELLI A, MARCELLONI F. A Survey on Fake News and Rumor Detection Techniques[J]. Information Sciences, 2019, 497: 38-55.

[3] KUMAR K, GEETHAKUMARI G. Detecting Misinformation in Online Social Networks Using Cognitive Psychology[J]. Human-centric Computing and Information Sciences, 2014, 4(1): 14-26.

[4] TRUEMAN T E, KUMAR A, NARAYANASAMY P, et al. Attention-based C-BiLSTM for Fake News Detection[J]. Applied Soft Computing, 2021, 110: 107600.

[5] WANG W Y. "Liar,Liar Pants on Fire": a New Benchmark Dataset for Fake News Detection[C]// Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, Vancouver, Canada. Stroudsburg: ACL, 2017, pp. 422-426.

[6] CHAWDA S, PATIL A, et al. A Novel Approach for Clickbait Detection[C]// Proceedings of the 3rd International Conference on Trends in Electronics and Informatics, India. 2019: 1318-1321.

[7] RAI N, KUMAR D, KAUSHIK N, et al. Fake news classification using transformer based enhanced LSTM and BERT[C]//International Journal of Cognitive Computing in Engineering.2022;3:98–105.

[8] AHN Y, JEONG C. Natural Language Contents Evaluation System for Detecting Fake News Using Deep Learning[C]// Proceedings of the 16th International Joint Conference on Computer Science and Software Engineering, IEEE, 2019, pp. 289-292.

[9] DEVLIN J, CHANG M W, et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding[C]// Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies,, Minneapolis, Minnesota, 2019, pp. 4171–4186.

[10] YU F,LIU Q, et al. A Convolutional Approach for Misinformation Identification[C]// Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, 2017, pp.3901-3907.

[11] RAZA N, ABDULKADIR SJ, et al. Enhancing fake news detection with transformer-based deep learning: A multidisciplinary approach[J]. PLoS One , 2025, 20(9):e0330954.

[12] XU F, LI M H,HUANG Q, et al. Rumor Detection Model based on Graph Convolutional Neural Network Driven by Knowledge Graph [J]. Science China Information Science, 2023, (04): 663-681. (in Chinese)
徐凡, 李明昊, 黄琪, 鄢克雨, 王明文, 周国栋. 知识图谱驱动的图卷积神经网络谣言检测模型[J]. 中国科学: 信息科学, 2023, (04): 663-681.

[13] LIANG X, ZHANG Q, SHI C, et al. MSynFD: Multi-hop Syntax Aware Fake News Detection[C]// Proceedings of the ACM on Web Conference, 2024, pp. 4128-4137.

[14] JIN Z W, CAO J, GUO H, et al. Multimodal Fusion with Recurrent Neural Networks for Rumor Detection on Microblogs[C]// Proceedings of the 25th ACM international conference on Multimedia. 2017, pp. 795-816.

[15]WANG Y, MA F, JIN Z, et al. EANN: Event Adversarial Neural Networks for Multi-Modal Fake News Detection[C]// Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2018, pp. 849-857.

[16] KIM Y. Convolutional Neural Networks for Sentence Classification[J]. EMNLP, 2014, arXiv:1408.5882.

[17] SIMONYAN, K., ZISSERMAN, A. Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. arXiv preprint, 2014 arXiv:1409.1556.

[18] ZHANG H, FANG Q, QIAN S, and XU C.. Multi-modal Knowledge-Aware Event Memory Network for Social Media Rumor Detection[C]// Proceedings of the 27th ACM Conference on Multimedia, 2019, pp. 1942–1951.

[19] WU Y, ZHAN P , ZHANG Y, WANG L, and Xu Z. Multimodal Fusion with Co-Attention Networks for Fake News Detection[C]// Proceedings of the Findings Association for Computational Linguistics, 2021, pp. 2560–2569.

[20] QIAN S, WANG J, HU J, FANG Q, and XU C. Hierarchical Multi-Modal Contextual Attention Network for Fake News Detection[C]// Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2021, pp. 153–162.

[21] SONG C, NING N, et al. A Multimodal Fake News Detection Model Based on Crossmodal Attention Residual And Multichannel Convolutional Neural Networks[J]// Information Processing & Management, 2021, 58(1): Art. no. 102437.
[22] HE K, ZHANG X, REN S and SUN J. Deep Residual Learning for Image Recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770-778.
[23] LIU Y, WU Y F. Early Detection of Fake News on Social Media Through Propagation Path Classification with Recurrent And Convolutional Networks[C]// Proceedings of the AAAI conference on artificial intelligence. 2018, pp. 354-361.
[24] JING J, Wu H, SUN J, et al. Multimodal Fake News Detection via Progressive Fusion Networks[J]. Information Processing & Management, 2023, 60(1): Art. no. 103120.
[25] LIU Z, LIN Y, CAO Y, et al. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows[C/OL]// Proceedings of IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 2021, pp. 9992-10002.
[26] ZHOU X Y, WU J D, Zafarani R. SAFE: Similarity-Aware Multi-modal Fake News Detection[C]// Proceedings of Pacific-Asia Conference on Knowledge Discovery and Data Mining. 2020, pp. 354–367.
[27] CHEN Y. et al. Cross-modal Ambiguity Learning for Multimodal Fake News Detection[C]// Proceedings of the ACM Web Conference, 2022, pp. 2897–2905.
[28] ZHOU Y, YANG Y, YING Q, QIAN Z, and ZHANG X. Multimodal Fake News Detection via CLIP-guided Learning[C]// Proceedings of the IEEE International Conference on Multimedia and Expo. 2023, pp. 2825–2830.
[29] WANG J, ZHANG H, LIU C, YANG X. Fake News Detection via Multi-scale Semantic Alignment and Cross-modal Attention[C/OL]// Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval, New York(NY), 2024,pp. 2406-2410.
[30] KUMAR V. Generative Adversarial Networks-aided Intrusion Detection System[J/OL]. Generative Adversarial Networks and Deep Learning. 2024, pp. 79-98.
[31] HE P, ZHANG H, CAO S, WU Y. Cross-Modal Fake News Detection Method Based on Multi-Level Fusion Without Evidence[J]. Algorithms. 2025; 18(7):426.
[32]YIFAN LIU, YAOKUN LIU, et al. Modality Interactive Mixture-of-Experts for Fake News Detection[C]// In Proceedings of the ACM on Web Conference 2025 (WWW '25). 2025, 5139–5150.
[33] HU J, YANG M, et al. Integrating Message Content and Propagation Path for Enhanced False Information Detection Using Bidirectional Graph Convolutional Neural Networks[J]. Appl. Sci. 2025, 15,3457.
[34]XUEQIN CHEN, et al. Enhancing text-centric fake news detection via external knowledge distillation from LLMs[J]. Neural networks : the official journal of the International Neural Network Society vol. 187 (2025): 107377.
[35] RADFORD A, KIM J, HALLACY C, et al. Learning Transferable Visual Models from Natural Language Supervision[C]// Proceedings of the International Conference on Machine Learning. 2021, pp. 8748-8763.
[36] ASHISH V, et al. Attention Is All You Need[C]// Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017, pp. 5998-6008.
[37] VINOD N, GEOFFREY E. H. Rectified Linear Units Improve Restricted Boltzmann Machines Vinod Nair[C]// Proceedings of the 27th International Conference on Machine Learning, 2010, pp. 807-814.
[38] MCCLELLAND J, RUMELHART D. Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Psychological and Biological Models[M]// MIT Press, 1987.
[39] KULLBACK S and LEIBLER R. A. On Information and Sufficiency[J]. The Annals of Mathematical Statistics, 1951, 22(1): 79-86.
[40] KINGMA D P, WELLING M. Auto-Encoding Variational Bayes[C]// Proceedings of the International Conference on Learning Representations. 2014, CoRR abs/1412.6980.
[41] ZUBIAGA A, LIAKATA M, PROCTER R. Exploiting Context for Rumour Detection in Social Media[M/OL]//Lecture Notes in Computer Science, vol 10539, Social Informatics. 2017: pp. 109-123.
[42] KAI S, DEEPAK M, et al. FakeNewsNet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media[J]. Big Data, 2020,8(3):171–188.
[43] MADRY A, MAKELOV A, SCHMIDT L, et al. Towards Deep Learning Models Resistant to Adversarial Attacks.[C]// Proceedings of the International Conference on Learning Representations. 2017, arXiv.1706.06083.
[44] YUAN C, MA Q, ZHOU W, et al. Jointly Embedding the Local and Global Relations of Heterogeneous Graph for Rumor Detection[C]//

Proceedings of IEEE International Conference on Data Mining, 2019, pp. 796-805.
[45] DIDRIK P. K, JIMMY B. Adam: A Method for Stochastic Optimization[C]//Proceedings of the International Conference on Learning Representations Computer Science, 2014. arXiv.1412.6980.

杨晨光，出生于 1996 年，硕士研究生，主要研究方向为虚假信息检测、大数据分析。

卢记仓，出生于 1985 年，博士，副教授，主要研究方向为知识推理和社交网络分析。

郭嘉兴，出生于 1995 年，硕士研究生，主要研究方向为情感分析、立场检测和大数据分析。

Yang Chenguang，born in 1996, postgraduate. His main research interests include fake news detection and big data analysis .

LU Jicang, born in 1985, Ph.D, associate professor. His main research interests include knowledge reasoning and social network analysis.

GUO Jiaxing, born in 1995, postgraduate. His main reserarch interests include sentiment analysis, stance detection and big data analysis.