基于多粒度信息融合的社交媒体多模态假新闻检测

Yangming Zhou
y
mzhou21@m.fudan.edu.cn 计算机科学学院 复旦大学 中国上海

Yuzhou Yang
yz
yang22@m.fudan.edu.cn 计算机科学学院 复旦大学 中国上海

Qichao Ying
shinydotcom@163.com
计算机科学学院 复旦大学
中国上海

Zhenxing Qian*
zxqian@fudan.edu.cn
计算机科学学院 复旦大学
中国上海

Xinpeng Zhang*
zhangxinpeng
@fudan.edu.cn 计算机科学学院
复旦大学 中国上海

arXiv:2304.00827v1 [cs.CV] 3 Apr 2023

摘要

社交媒体上多媒体内容的便捷分享导致假新闻迅速传播，威胁着社会的稳定与安全。因此，虚假新闻检测已成为社交取证领域广受关注的研究课题。当前方法主要侧重于整合文本和视觉特征，却未能有效利用细粒度和粗粒度层面的多模态信息。此外，由于模态之间缺乏相关性，或不同模态各自作出的判断相互矛盾，这些方法还面临歧义问题。为应对这些挑战，我们提出了一种用于虚假新闻检测的多粒度多模态融合网络（MMFN）。受人类评估新闻真实性时所采用的多粒度过程启发，我们分别使用两个基于 Transformer 的预训练模型，对文本和图像中的令牌级特征进行编码。多模态模块融合细粒度特征，同时考虑由 CLIP编码器编码的粗粒度特征。为解决歧义问题，我们设计了带有基于相似度的加权机制的单模态分支，以自适应地调整多模态特征的使用。实验结果表明，在三个常用数据集上，所提出的框架优于最先进方法。

CCS概念

• **信息系统** →**数据挖掘**。

关键词

虚假新闻检测，多模态学习，多模态融合

*通讯作者。所有作者也隶属于复旦大学文化和旅游部文化和旅游智能计算重点实验室。

允许免费为个人或课堂使用而制作本作品全部或部分内容的数字副本或纸质副本，前提是这些副本不得用于营利或商业利益，也不得为此目的进行复制或分发，并且副本首页须保留本声明和完整引文。对于本作品中由 ACM 以外的其他方拥有版权的组件，必须尊重其版权。注明出处后允许摘录。若要以其他方式复制、重新发布、发布到服务器或重新分发给列表，则须事先获得特定许可和/或支付费用。请向 permissions@acm.org 申请许可。*ICMR '23，2023*年*6*月*12*日至*15*日，塞萨洛尼基，希腊© 2018年计算机协会。ACM ISBN 978-1-4503-XXXX-X/18/06...$15.00 https://doi.org/XXXXXXX.XXXXXXX

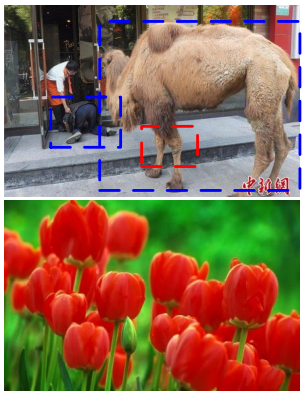
ACM参考文献格式：

Yangming Zhou、Yuzhou Yang、Qichao Ying、Zhenxing Qian 和 Xinpeng Zhang。2018年。通过多粒度信息融合进行社交媒体上的多模态假新闻检测。载于论文集 请填写版权确认邮件中的正确会议名称 *(ICMR '23)*。ACM，纽约，纽约州，美国，共10页。https://doi.org/XXXXXXX.XXXXXXX

1 引言

Twitter 和微博等在线社交网络，在很大程度上取代了以报纸和杂志为代表的传统信息传播方式。尽管这些网络为用户提供了诸多便利，例如寻找朋友和分享观点，但也助长了假新闻的快速广泛传播 [16, 25, 47]。与传统书面材料相比，在线新闻帖子更容易受到操纵。此外，读者很容易受到精心编造的假新闻影响，甚至可能在不经意间进一步传播这些内容。假新闻会造成恐慌、误导公众舆论，并对社会产生负面影响。数以亿计的社交媒体用户不断生成海量帖子，因此，采用传统且效率低下的人工审查方法来遏制假新闻威胁并不可行。为此，计算机科学领域的许多研究人员近期开始专注于开发假新闻检测方法 [2, 7, 21, 24, 29]。

早期的虚假新闻检测研究主要关注分析纯文本或纯图像内容 [5, 9]。这些研究通常会验证输入内容的逻辑性和语义连贯性，并考虑一些简单指标，例如语法错误或图像篡改痕迹。尽管单模态方法有效，但现代新闻文章和帖子往往包含彼此相关的多种模态，因此这些方法忽略了这类相关性。例如，真实图像可能与完全虚假的谣言结合，而正确的文字也可能被用来描述一张篡改图像。因此，多模态特征分析对于虚假新闻检测至关重要，能够带来互补优势。近年来，已有许多研究结合多模态特征来检测新闻中的异常 [7, 16, 35, 41]。尽管该领域取得了进展，当前方法仍面临两项重大挑战。首先，尽管许多研究提出了新颖的融合方法，但它们



虚假：网民爆料称，街头有乞丐乞讨，身边还牵着骆驼。图片中的骆驼四肢被截断了，从膝盖处起无法行走，只能躺在地上。

真实：爱仍在传递！让爱充满世界！

图1：当前研究面临的两项挑战示例。上图：检测虚假信息需要多粒度信息（蓝框表示匹配的元素，红框表示不匹配的元素）。下图：文本与图像之间的相关性较弱，可能会妨碍用户/模型发现跨模态线索。

它们仅在整体帖子层面进行融合，因此不可避免地会遗漏一些细节信息；或者只考虑实体、词元或区域之间的匹配，而忽视全局语义相关性。这导致不同粒度的信息未能得到充分利用。其次，许多研究过度依赖多模态融合特征，因此受到歧义问题的影响，即模态间冲突或相关性较弱所导致的不一致。

图1展示了微博数据集中两个案例，说明了上述两项挑战。图1上方的图像描绘了多粒度虚假新闻检测的过程，其中，无论是文本还是图像的单模态特征，都无法验证其真实性。首先，读者会关注图像中的物体和文本中的实体。对于给出的案例，人们首先会看到图片中的乞丐和骆驼，以及文本中的“乞丐”、“乞讨”、“骆驼”、“截肢”、“膝盖 ”和“谎言 ”等词语。此时，除了匹配的元素（以蓝色区域标示）外，他们还会发现“截肢”、“膝盖 ”和“谎言 ”这三个词与图像内容不匹配（以红色区域标示）。随后，他们会理解句子和图像整体所表达的语义，分析二者是否匹配，最终得出关于新闻真实性的结论。许多现有研究往往忽略了这一点。图1下方的图像展示了一个体现歧义的案例，这也是社交媒体上常见的一类帖子。帖子中的视觉物体与文本实体之间没有匹配关系，语义上也互不相关。发布该帖的用户可能只是在表达自己的感受。人工审核可以轻易判定这不是假新闻，也没有危害。然而，过度强调多模态融合的模型可能会因多模态特征不匹配而误判。因此，该帖子可能无法通过社交软件的自动审核流程，难以顺利发布。在这种情况下，分析单模态内容和情感就足以评估帖子的可信度；多模态融合特征并非必需，甚至可能给分类带来噪声

任务。为解决这一问题，Chen等人 [7] 提出了CAFE：首先对图像和文本进行压缩，并通过最小化Kullback–Leibler（KL）散度，对具有正确图文对的新闻进行对比学习。随后，利用相应的跨模态歧义分数对多模态特征重新加权。然而，由于输入网络的数据有限，其多模态特征的对齐无法得到保证。此外，他们通过使用非共享权重的变分自编码器处理单模态特征，因此我们无法确定KL散度是否足以可靠地衡量跨模态歧义。因此，解决歧义问题仍是本研究的动机之一。

为解决上述问题，本文提出多粒度多模态融合网络（MMFN）。MMFN 方法融合单模态特征与多粒度多模态融合特征，以更准确地检测虚假新闻。MMFN 分别使用预训练的 BERT [11] 和 Swin Transformer（Swin-T） [19] 模型对文本和图像进行编码，从而提取词元级细粒度信息。预训练的 CLIP [27]模型用于编码粗粒度特征并捕获帖子级语义信息，进而解决歧义问题。我们提出一种多粒度多模态融合模块，对多模态特征进行粒度级融合：将细粒度特征输入协同注意力 Transformer（CT）模块，以生成对齐的细粒度多模态特征；利用粗粒度特征评估跨模态相关性。该模块动态调整多模态特征的使用方式，从而缓解歧义问题。

本文的贡献主要体现在以下三个方面：

- 我们提出了 MMFN，通过在不同粒度层级上处理多模态特征，构建全面的表征，以反映新闻的细节和整体特征。
- 我们专门设计了两个单模态分支，并采用 CLIP 预训练模型评估跨模态相关性，从而进一步解决高跨模态歧义场景带来的问题。
- 我们在三个知名数据集上进行了全面实验，结果表明，MMFN 的表现优于当前最先进的虚假新闻检测方法。此外，消融实验验证了粒度层级处理和多模态特征调整的有效性。

2 相关工作

2.1 虚假新闻检测

模态是指人类交流所适应的不同传播渠道所传递的信息 [3]。根据所使用的模态，虚假新闻检测方法可分为单模态和多模态虚假新闻检测。近年来，基于深度学习的这两类方法都已取得大量进展。

尽管文本内容等单模态特征在区分假新闻方面发挥着关键作用， [22, 46], 多模态特征之间的相关性和一致性也至关重要。早期研究设计了复杂但不透明的注意力机制，用于多模态特征融合 [4, 5, 16]。Singhal 等人分别将预训练的 XLNet 和 ResNet 用于文本和特征提取。随后拼接这些特征，以进行最终的二分类 [30]。Singhal

等人 [32] 提出了 Spotfake，使用 VGG 和 BERT 提取特征，之后又在 Spotfake+ [30] 中进行了改进，以检测全文文章。Wang 等人 [35] 提出了一种事件咨询神经网络，即 EANN，旨在提取不同事件新闻之间的共享特征。许多其他研究 [7, 16, 41] 也提出了更好的方法，在将特征送入分类器之前融合从不同模态中提取的特征。

与单模态方法相比，多模态方法不仅拥有额外的图像特征，还能挖掘多模态内容中蕴含的更多潜在信息。一些研究尝试从数据集中挖掘潜在信息。Qi 等人 [24]指出，以往研究忽略了无法通过预训练提取器提取的信息，例如实体和图像内文本。因此，他们手动提取这些信息，作为文本内容的补充。Zhang 等人 [43] 设计了一种新颖的双重情感特征描述符，用于衡量发布者与评论之间的情感差异，并进一步验证了双重情感在假新闻与真实新闻之间具有显著差异。Allein 等人 [2] 提出了 DistilBert，它利用新闻文章和用户生成内容的潜在表示来指导模型学习。Wang 等人 [36] 提出 KMGCN，将文本、视觉和知识信息整合到统一框架中，以建模语义表示并提高准确率。此外，Abdelnabi 等人 [1] 使用在线搜索引擎收集相关网络证据，并提出一致性检查网络，以模拟人类推理过程。

除了挖掘信息外，规划更好的多模态表示交互对于提升检测性能也至关重要。SAFE [44] 计算新闻文本与视觉信息之间的相关性。MCAN [38] 堆叠多个协同注意力层，以融合多模态特征。Qian 等人 [26] 使用 BERT 生成文本的层次语义，并使用 ResNet 生成区域图像表示。随后，将不同层次的语义与区域图像特征输入协同注意力层，以实现层次化多模态特征融合。FND-CLIP [45]使用 CLIP 提取的特征作为多模态表示，并设计了模态级注意力来聚合特征。

2.2 预训练模型

Vaswani 等人 [34] 提出了 Transformer，该模型迅速成为许多自然语言处理任务中的最先进方法。随后，Devlin 等人 [11] 提出了 BERT，这是一种在大型语料库上预训练的大型基于 Transformer的模型。基于Transformer的图像处理方法也在各自领域取得了成果。然而，将自注意力简单应用于图像，需要的计算开销会随像素数量呈二次方增长，这使其难以处理尺寸符合实际应用需求的图像。Dosovitskiy 等人 [12] 将 Transformer 架构直接应用于互不重叠的中等大小图像块进行图像分类，这一方法称为 ViT。在此基础上，Liu 等人 [19] 提出了 Swin-T，将卷积层与其提出的 Swin-T 层相结合，使计算复杂度随图像尺寸线性增长。Swin-T 更注重通用性能，而非像 ViT 那样专门针对分类；但在图像分类任务中，它在其他方法中仍实现了最佳的速度与准确率权衡。

过去十年间，多模态机器学习受到了研究社区的广泛关注 [3]。神经网络架构被用于超越单一模态的任务，例如，视觉问答（VQA） [10], 、视觉常识推理（VCR） [40], 等。在这些任务中，需要利用来自不同模态的先验知识和特征；如果仅提供单一模态的信息，算法或深度网络便无法有效发挥作用。人们开发了多种通用技术，用于学习图像内容与自然语言的联合表示。例如，CLIP [27] 是一种多模态模型，它将语言概念知识与图像的语义知识相结合。得益于对比学习范式，CLIP 提取的文本和视觉特征可视为在同一语义空间中对齐，从而反映新闻文本内容与视觉内容之间的相关性。基于 CLIP 的多模态学习已应用于许多下游任务 [8, 37]。其他多模态模型，如 GLIDE [23] 和 VilBERT [20]，也已用于文本生成图像和跨模态表示学习等任务。

3 方法

本文提出 MMFN，通过多粒度多模态融合提升多模态虚假新闻检测的准确率。MMFN 的网络设计如图2所示，由多模态特征编码器、多粒度多模态融合、单模态分支、基于 CLIP 相似度的模态加权以及假新闻分类器组成。

3.1 多模态特征编码器

基于 BERT 的文本特征编码。BERT [11] 是一种基于 Transformer 构建的热门预训练语言模型，通过在大型语料库上进行无监督学习来训练，并在许多自然语言处理下游任务中取得了出色的结果。因此，我们使用 BERT 模型对T中的特征进行编码。新闻帖子的文本内容由文本和从图像中提取的光学字符识别（OCR）结果拼接而成，是一个由词组成的序列，记作$\mathbf{T} = \left[t_1, t_2, \ldots, t_{n_w}\right]$，其中 n_w表示词数。将 BERT 应用于T后，可得到编码后的文本特征$\mathbf{T}^b = \left[t_1^b, t_2^b, \ldots, t_{n_w}^b\right]$，其中 $t_i^b \in \mathbb{R}^{d_b}$表示文本嵌入中第 i–个词元对应的最后一层隐藏状态输出，d_b 表示词嵌入的维度。

通过 Swin-T 进行视觉特征编码。Swin-T [19] 是一种视觉 Transformer，基于移位窗口自注意力生成分层特征表示，并在目标检测和语义分割等多项计算机视觉任务中取得了最先进的性能。本文将 Swin-T 引入虚假新闻检测任务。给定视觉内容 $\mathbf{V} \in \mathbb{R}^{w \times h}$，Swin-T 将其转换为序列嵌入 $\mathbf{V_s} = \left[v_1^s, v_2^s, \ldots, v_{n_p}^s\right]$，其中 w 和 h 分别是图像的宽度和高度，$v_i^s \in \mathbb{R}^{d_s}$是模型最后一层在输入中第 i–个窗口对应的隐藏状态，n_p 是 Swin-T 中图像块的数量，d_s是视觉嵌入的隐藏维度。**通过 CLIP 的多模态特征编码**。CLIP 在海量且多样化的数据集上进行预训练，能够将文本和图像嵌入统一的数学空间，因此天然适合计算跨模态相关性。此外，实验表明，该模型在应对分布偏移时具有鲁棒性，

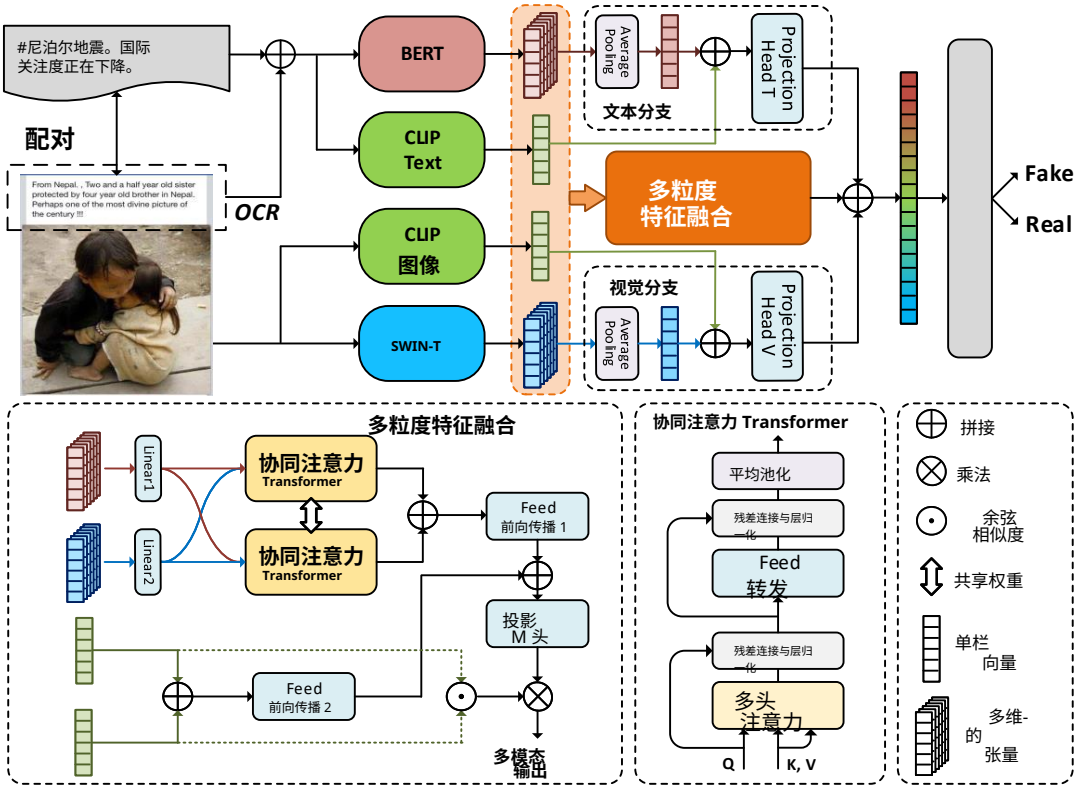


图2：MMFN 的架构。使用 BERT、Swin-T 和 CLIP 对多模态新闻中不同模态的特征进行编码。不同粒度的编码特征通过协同注意力 Transformer 和投影头进行融合。计算 CLIP 相似度分数，以自适应地为不同模态的特征赋权，从而使分类器能够对假新闻进行分类。

因此，即使在零样本场景下，它也适用于虚假新闻检测 [27]。许多研究 [13, 18] 已展现出 CLIP 在泛化到未知领域方面的强大能力。因此，我们使用 CLIP 多模态特征来丰富文本和视觉特征的全局相关信息。给定多模态新闻 $\mathbf{X} = \{\mathbf{T}, \mathbf{V}\}$，我们将 CLIP 编码后的特征表示为 $\mathbf{X}^c = [t^c, v^c]$，其中 t^c，$v^c \in \mathbb{R}^{d_c}$ 是两个长度为 d_c 的向量。

3.2 多粒度特征融合

通过 Transformer 实现细粒度融合。 由于 BERT 和 Swin-T 并非多模态模型，它们提取的特征之间存在较大差异，无法直接实现信息交互。为有效融合帖子的文本和视觉特征，我们使用协同注意力 Transformer [20] 实现模态间信息互补。如图2所示，协同注意力 Transformer 由一个多头注意力网络和一个前馈神经网络组成，二者后面均接有残差连接和层归一化。

我们分别将不同模态的输入记为 I_1 和 I_2。在协同注意力 Transformer 中，I_1 用作查询 Q，而 I_2 用作键 K 和值 V。协同注意力 Transformer 计算每个头的协同注意力矩阵如下：

$$h_i = Softmax(\frac{(I_1 W_i^q)(I_2 W_i^k)^T}{d^h}) I_2 W_i^v, \tag{1}$$

其中，投影矩阵为 $W_i^q, W_i^k, W_i^v \in \mathbb{R}^{d_m \times d_h}$，$d_h = d_m/m$，$d_m$ 是协同注意力 Transformer 模型的维度，而 m 是头的数量。

多头注意力是所有协同注意力矩阵经过一个投影矩阵后的拼接：

$$H = (h_1; h_2; \cdots; h_m) W^o, \tag{2}$$

其中，符号；表示拼接操作，且 $W^o \in \mathbb{R}^{d_m \times d_m}$。

在 H 和 I_1 经过带有两层层归一化的前馈神经网络后，可得到基于注意力的多模态表示 H'：

$$H' = \text{Norm}(I_1 + \text{FFN}(\text{Norm}(I_1 + HI_1))). \tag{3}$$

最后，将多模态表示 H' 进行平均池化，得到特征向量 F，作为协同注意力 Transformer 的输出：

$$F = \text{CT}(I_1, I_2) = \overline{H'}. \tag{4}$$

在我们的模型中，将输入的 BERT 和 Swin-T 特征 $\mathbf{T}^b$，$\mathbf{V}^s$ 通过线性层映射到相同维度 $\mathbb{R}^{d_m}$ 后，按照不同的前后顺序将它们输入共享加权CT，分别表示为 I_1 和 I_2，以获得输出特征：视觉注意力加权文本特征 F^{vt} 和文本注意力加权视觉特征 F^{tv}：

$$\begin{cases} F^{vt} = \text{CT}((\mathbf{T}^b W^t), (\mathbf{V}^s W^v)) \\ F^{tv} = \text{CT}((\mathbf{V}^s W^v), (\mathbf{T}^b W^t)) \end{cases}, \tag{5}$$

其中， $W^t \in \mathbb{R}^{d_b \times d_m}$ 和 $W^v \in \mathbb{R}^{d_s \times d_m}$。

MCAN [38] 和 HMCAN [26] 也使用 Transformer 融合特征。MMFN 中的协同注意力 Transformer 与 HMCAN 和 MCAN 中的 Transformer 在结构和功能上存在差异。MCAN 将池化后的特征向量作为堆叠协同注意力模块的输入，输出粗粒度的模态相关性向量；而 MMFN 的注意力结构在词元级进行计算，并输出细粒度融合特征。HMCAN 中的 Transformer 由一个用于增强模态内信息的单头自注意力模块，以及另一个用于捕获模态间信息的模块组成。考虑到 BERT 和 Swin-T 已经使用自注意力实现模态内交互，为避免影响预训练模型并缓解过拟合问题，MMFN 中的协同注意力 Transformer 仅包含多头协同注意力，以实现模态间交互。此外，HMCAN 中的 Transformer 不共享权重，而 MMFN 中的 Transformer 共享权重，使协同注意力 Transformer 具备模态对齐功能。

通过 CLIP 进行粗粒度融合与多模态表示生成。我们将 CLIP 编码的特征融合为稳健的粗粒度特征，以反映全局语义相关性信息。可以认为，CLIP编码器的输出 t^c 和 v^c 通过对比学习有效消除了模态间差异。这使得后续网络学习能够有效利用来自不同模态的信息。

具体而言，将 t^c 和 v^c 拼接后输入一个包含线性层、批归一化层和*ReLU*激活函数的前馈网络。此外，将 F^{vt} 和 F^{tv} 也进行拼接，并输入另一个具有相同架构的前馈网络。因此，经过两个前馈网络后，我们得到融合的多模态细粒度特征 M^f和粗粒度特征 M^c:

$$\begin{cases} M^f = \mathrm{FFN}_1(F^{vt}; F^{tv}) \\ M^c = \mathrm{FFN}_2(t^c; v^c) \end{cases}. \tag{6}$$

最后，将多模态特征 M^f和 M^c 拼接后输入投影头 Φ_M，以生成多模态表示。输出向量中的每个元素都将乘以一个相似度分数，该分数将在下一段中介绍。生成的多模态表示记为 F^m:

$$F^m = similarity \cdot \Phi_M(M^f; M^c). \tag{7}$$

3.3 基于 CLIP 相似度的单模态分支与模态加权

多模态融合特征通常反映两种模态之间的相关性信息，但很容易受到歧义的影响。为了解决模态歧义程度较高时多模态融合的特征表示能力下降的问题，我们分别设计了单模态文本分支和单模态视觉分支。

对于文本分支，我们将 BERT 特征沿词元级维度池化为一个特征向量，并将其与 CLIP文本特征向量拼接，然后通过一个由两个全连接网络组成的投影头，并使用 *ReLU* 激活函数，得到单模态文本表示；类似地，对于视觉分支，我们将池化后的 Swin-T 特征与 CLIP图像特征拼接，再通过以下方式获得单模态视觉表示：

一个映射结构相同但参数不同于文本分支的投影头：

$$\begin{cases} F^t = \Phi_T(\overline{\mathbf{T}^b}; t^c) \\ F^v = \Phi_V(\overline{\mathbf{V}^s}; v^c) \end{cases}, \tag{8}$$

其中， Φ_T 和 Φ_V 是投影头。

如果直接将单模态分支表示输入分类器进行决策，分类器可能更倾向于使用网络更深的多模态表示来拟合结果，而单模态分支则可能干扰决策并导致更严重的歧义问题。为克服这些局限，受 CAFE 启发，我们使用 CLIP 余弦相似度作为加权系数，对多模态特征进行加权，从而引导分类器的学习过程。余弦相似度的计算方法如下：

$$similarity = \frac{t^c \cdot (v^c)^T}{\|t^c\| \, \|v^c\|}. \tag{9}$$

3.4 假新闻分类器

获得融合多模态表示、单模态文本表示和单模态视觉表示后，我们将它们拼接起来作为分类器的输入，并得到输出 $\hat{y}$，表示新闻为假新闻的概率：

$$\hat{y} = \mathrm{FNC}(F^m; F^t; F^v), \tag{10}$$

其中，FNC(·) 是假新闻分类器，由一个带有 *ReLu* 激活函数的两层全连接网络构成。

目标函数是最小化交叉熵损失，以正确预测真实新闻和假新闻。

$$\mathcal{L} = y \log(\hat{y}) + (1 - y) \log(1 - \hat{y}), \tag{11}$$

其中， y 是真实标签。

4 实验

4.1 实验设置

数据集。 我们使用了三个从社交媒体收集的公开真实世界数据集，分别是微博 [14],Twitter [6], 和 Gossipcop [29]。

微博是虚假新闻检测中广泛使用的中文数据集。真实新闻采集自新华社，这是中国一家权威新闻来源。在实验中，筛除了没有图像或没有文本描述的单模态新闻帖子。训练集包含 3,783 条真实新闻和 3,675 条假新闻，测试集包含 1,685 条新闻。

Twitter 数据集是为 MediaEval 多媒体使用验证任务发布的 [6]，也是一个知名的虚假新闻检测多模态数据集。在实验中，我们遵循现有研究，筛除了附带视频的推文和非英语推文。筛选后，训练集包含 4,031 条真实新闻和 5,139 条假新闻，测试集包含 1,406 条帖子。

Gossipcop 数据集是从 FakeNewsNet [29] 代码库的娱乐领域收集的英语全文新闻文章数据集。Gossipcop 包含 10,010 条新闻，其中 7,974 条为真实新闻，2,036 条为假新闻；训练新闻数量为 2,285 条。测试集中有真实新闻和 545 条假新闻。

实现细节。 在我们的实验中，CT 的 $d_m = 512$ 和 $m = 8$ 分别设为。BERT 的文本嵌入维度

表 1：MMFN 与最先进方法在微博、Twitter 和 Gossipcop 数据集上的性能比较。最佳性能以粗体突出显示。

	方法	准确率	假新闻			真实新闻		
			精确率	召回率	F1分数	精确率	召回率	F1分数
微博	Spotfake [32]	0.892	0.902	**0.964**	**0.932**	0.847	0.656	0.739
	CAFE [7]	0.840	0.855	0.830	0.842	0.825	0.851	0.837
	MCAN [38]	0.899	0.913	0.889	0.901	0.884	0.909	0.897
	LIIMR [31]	0.900	0.882	0.823	0.847	0.908	**0.941**	**0.925**
	HMCAN [26]	0.885	0.920	0.845	0.881	0.856	0.926	0.890
	MMFN	**0.923**	**0.921**	0.926	0.924	**0.924**	0.920	0.922
Twitter	Spotfake [32]	0.777	0.751	**0.900**	0.820	0.832	0.606	0.701
	CAFE [7]	0.806	0.807	0.799	0.803	0.805	0.813	0.809
	MCAN [38]	0.809	0.889	0.765	0.822	0.732	0.871	0.795
	LIIMR [31]	0.831	0.836	0.832	0.830	0.825	0.830	0.827
	HMCAN [26]	0.897	**0.971**	0.801	0.878	0.853	0.979	0.912
	MMFN	**0.935**	0.960	0.856	**0.905**	**0.924**	**0.980**	**0.951**
Gossipcop	SAFE [44]	0.838	0.758	0.558	0.643	0.857	0.937	0.895
	Spotfake+ [30]	0.858	0.732	0.372	0.494	0.866	0.962	0.914
	DistilBert [2]	0.857	**0.805**	0.527	0.637	0.866	0.960	0.911
	CAFE [7]	0.867	0.732	0.490	0.587	0.887	0.957	0.921
	MMFN	**0.894**	0.799	**0.598**	**0.684**	**0.910**	**0.964**	**0.936**

设为 $d_b = 768$。对于中文数据，我们使用“bert-base-chinese”模型；对于英语数据，我们使用“bert-base-uncased”模型。输入文本长度设为 300 个词，即 $n_w = 300$。对于 Swin-T，我们使用“swin-base-patch4-window7-224”模型对视觉特征进行编码，并将输入图像尺寸设为 224×224。Swin-T 中的图像块数量为 $n_p = 49$，视觉嵌入的维度为 $d_s = 1024$。CLIP 的输入图像尺寸也设为 224×224。Science CLIP 没有预训练的中文文本模型，因此我们使用 Google 翻译 API [15] 将中文文本翻译成英语。对于超过 50 个词的文本，我们使用摘要生成模型 [28]生成摘要，以满足 CLIP 的输入尺寸要求。所使用的预训练 CLIP 模型为“ViT-B/32”，特征维度为 $d_c = 512$。训练阶段对 BERT 和 Swin-T 进行微调；由于 CLIP 难以在小型数据集上训练，因此冻结其参数。与 MCAN [38], 类似，我们在 Twitter 数据集上冻结 BERT 模型，以缓解过拟合。前馈网络 1 和前馈网络 2 的隐藏维度均为 256，两个投影头的隐藏单元数分别为 256 和 16。分类器中两个全连接层的隐藏维度分别为 48 和 2。批次大小设为 16。对数据集进行预处理，移除所有位于“@”、“#”和“http:”之后的无效消息。使用 Adam 优化器 [17]，并采用默认参数。除微调 BERT 和 Swin Transformer 时使用学习率 1×10^{-5}外，网络的学习率设为 1×10^{-3}。模型训练 100 个周期，并采用提前停止以防止过拟合。

4.2 性能比较

我们将 MMFN 的性能与其他先进方法进行比较，比较结果见表 1。评估指标包括准确率、精确率、召回率和 F1分数，这些指标通常用于衡量二分类问题的性能。如表 1 所示，就准确率而言，MMFN 在三个数据集上的表现均优于其他方法。具体而言，MMFN 在三个真实世界数据集上分别取得了最高准确率：**92.3%**、**93.5%**和**89.4%**，分别超过最先进方法**2.3%**、**3.8%**和**2.7%**。在精确率、召回率和 F1 分数方面，MMFN 在几乎所有测试中均位列第一或第二，展现了其有效性。

所比较的方法 SAFE、Spotfake、DistilBert 和 Spotfake+ 在虚假新闻检测方面均存在局限。SAFE 学习文本与视觉信息之间的相似性，但由于忽略了歧义，可能会将相关性较弱的真实帖子误判为假新闻。DistilBert 通过检查用户相关约束对文章潜在空间的影响来指导检测，但它忽略了新闻的视觉信息，因此性能竞争力较弱。Spotfake 和 Spotfake+ 只是简单拼接文本与视觉表征，缺乏充分的跨模态交互与融合，导致性能未达最优。CAFE 定义并利用跨模态歧义，以缓解不同模态之间意见不一致的问题。在 Twitter 和 Gossipcop 数据集上，其实验结果优于 Spotfake、Spotfake+ 和 SAFE。

LIIMR 能够捕捉细粒度的显著图像与文本特征，因此在微博数据集上取得了更好的结果。MCAN 和 HMCAN 的良好实验结果证明了基于 Transformer 的多模态融合网络的有效性。然而，这些方法仅专注于细粒度特征挖掘，忽略了有助于洞察全局语义的粗粒度信息。

MMFN优于其他方法可归因于三个因素。1）Swin-T组件能够提取细粒度特征，以补充BERT编码器生成的特征。此外，预训练CLIP编码器能够生成粗粒度文本与图像特征，这些特征在共享的

表 2：对 MMFN 在微博、Twitter 和 Gossipcop 数据集上的模态、粒度及架构设计进行消融研究。最佳性能以粗体标出。

	方法	准确率	假新闻			真实新闻		
			精确率	召回率	F1分数	精确率	召回率	F1分数
微博	MMFN 去除 T	0.825	0.846	0.815	0.830	0.803	0.836	0.819
	MMFN 去除 V	0.901	0.859	0.940	0.898	0.944	0.868	0.904
	MMFN 去除 F	0.818	0.793	0.839	0.815	0.844	0.800	0.821
	MMFN 去除 C	0.909	0.914	0.907	0.911	0.904	0.912	0.908
	MMFN（去除协同注意力 Transformer）	0.905	0.849	**0.959**	0.900	**0.963**	0.861	0.909
	MMFN（去除 U）	0.912	**0.939**	0.891	0.915	0.884	**0.934**	0.908
	MMFN（去除 W）	0.906	0.921	0.895	0.908	0.890	0.917	0.903
	MMFN	**0.923**	0.921	0.926	**0.924**	0.924	0.920	**0.922**
Twitter	MMFN（去除 T）	0.933	0.963	0.850	0.903	0.919	0.981	0.949
	MMFN（去除 V）	0.739	0.691	0.580	0.630	0.762	0.839	0.799
	MMFN（无 F）	0.866	**0.993**	0.708	0.826	0.805	**0.996**	0.891
	MMFN（无 C）	0.882	0.817	0.817	0.817	0.913	0.913	0.913
	MMFN（无协同注意力 Transformer）	0.900	0.965	**0.885**	**0.923**	0.793	0.932	0.857
	MMFN（无 U）	0.918	0.929	0.835	0.880	0.913	0.965	0.938
	MMFN（无 W）	0.915	0.960	0.810	0.879	0.893	0.979	0.934
	MMFN	**0.935**	0.960	0.856	0.905	**0.924**	0.980	**0.951**
Gossipcop	MMFN 去除 T	0.836	0.702	0.255	0.374	0.846	0.974	0.905
	MMFN 去除 V	0.888	0.818	0.536	0.648	0.898	0.972	0.933
	MMFN 去除 F	0.862	0.691	0.517	0.592	0.891	0.945	0.917
	MMFN 去除 C	0.885	**0.844**	0.495	0.624	0.890	**0.978**	0.932
	MMFN 去除协同注意力 Transformer	0.888	0.756	**0.615**	0.678	**0.912**	0.953	0.932
	MMFN 去除 U	0.885	0.799	0.539	0.644	0.898	0.965	0.932
	MMFN 去除 W	0.889	0.764	0.611	0.679	0.911	0.968	0.933
	MMFN	**0.894**	0.799	0.598	**0.684**	0.910	0.964	**0.936**

语义空间中蕴含丰富的语义信息。这使得模型能够在细粒度和粗粒度上实现互补的特征表示。2）CT组件支持词元级的模态间交互，从而促进多种模态的细粒度融合。3）采用基于CLIP加权的单模态分支，能够有效缓解歧义问题。

4.3 消融实验

我们通过采用模型的各种部分配置进行实验，评估 MMFN 中关键组件对其性能的影响。对于每项实验，我们移除一个不同的组件，并从头开始重新训练模型。MMFN 的对比变体实现如下：

1) MMFN 去除 T。移除与文本相关的模块，仅使用由 Swin-T 和 CLIP图像编码器编码的单模态视觉特征。

2) MMFN 去除 V。移除与视觉相关的模块，仅保留由 BERT 和 CLIP 文本编码器编码的单模态文本特征。

3) MMFN 去除 F。移除与 BERT 和 Swin-T 相关的模块，不使用细粒度特征。相反，将 CLIP 编码的文本和图像特征直接拼接为多模态表示，并将两个 CLIP 特征分别用作两个独立的单模态表示。

4）MMFN（不含 C）。移除与 CLIP 相关的模块，不使用粗粒度特征。仅使用细粒度特征进行分类。

5）MMFN（不含 CT）。移除 CT模块，并将 BERT 和 Swin-T 编码的特征直接拼接。

6) MMFNw/o U。移除文本和视觉单模态分支，仅使用多模态融合表示进行分类。

7) MMFN w/o W。移除CLIP加权模块，且不对多模态融合特征进行加权。

不同模态的贡献。为评估不同模态对MMFN模型整体性能的贡献，我们将MMFN w/o T和MMFN w/o V的结果与完整MMFN模型进行比较。结果表明，当缺少视觉模态或文本模态时，MMFN的性能会下降。这说明两种模态对于模型的最终性能都至关重要。

此外，我们发现，MMFN w/o T 在微博和 Gossipcop 数据集上的表现较差，但在 Twitter 数据集上排名第二。另一方面，MMFN w/o V 在 Twitter 数据集上的表现不佳，但在微博和 Gossipcop 数据集上优于 MMFN w/o T 及一些经过消融的多模态模型。这些观察结果表明：1）不同数据集和不同社交媒体平台上的新闻分布各不相同。例如，微博新闻的文本包含丰富的虚假新闻检测线索，而 Twitter 用户倾向于通过图像表达信息。与此同时，Gossipcop 上长篇新闻的主要信息位于文本中，图像所起的作用较小。这说明单模态模型可能难以适应不同的社交环境。2）单模态模型的表现优于

若干经过消融的多模态模型，这表明，只有在通过有效机制充分挖掘并整合不同模态的线索信息时，多模态学习才有效。

不同粒度的影响。为分析不同粒度对模型性能的影响，我们将 MMFN w/o C 和 MMFN w/o F 的结果与完整的 MMFN 模型进行比较。结果表明，粗粒度特征和细粒度特征都有助于提升模型性能。移除细粒度特征后，模型性能的下降幅度大于移除粗粒度特征时的下降幅度。这说明细粒度特征在模型的最终性能中发挥着更重要的作用。这一推断符合我们的直觉，近期的研究工作，如 LIIMR 和 HMCAN 等，正是通过挖掘更多细粒度特征来检测假新闻。然而，消融研究结果表明，粗粒度特征有助于细粒度特征提升检测能力。

各组件的有效性。为评估 MMFN 模型中各组件的有效性，我们将完整的 MMFN 模型与消融模型 MMFN w/o CT、MMFN w/o U 和 MMFN w/o W 的结果进行比较。如表 2 所示，可以得出以下结论：1）通过 CT 模块融合视觉和文本模态，增强了表征能力和细粒度特征融合能力，这体现在完整模型的性能优于 MMFN w/o CT。2）完整的 MMFN 模型优于 MMFN w/o U，证明了为处理歧义而设计的单模态分支能够提升模型性能。3）结果还表明，在微博和 Twitter 数据集上，完整的 MMFN 模型和 MMFN w/o U 均优于 MMFN w/o W，这说明对融合特征加权有助于缓解歧义。另一方面，在不加权的情况下仅添加单模态分支可能会损害性能。总的来说，结果表明，只有完整组合所有组件，才能使 MMFN 模型取得最佳性能。

4.4 T-SNE可视化

在图3中，我们进一步通过 t-SNE [33] 可视化分析了 MMFN 及其变体在微博上学习到的不同模态表示的可分性。视觉、文本、多模态和完整表征分别来自 MMFN w/o T、MMFN w/o I、MMFN w/o U 和完整的 MMFN。相同颜色的点表示具有相同标签的实例。

MMFN w/o T 学习到的视觉表征与不同标签的表征存在一定重叠，这表明仅凭视觉信息不足以完成微博分类。另一方面，与视觉表征相比，MMFN w/o I 学习到的文本表征具有较好的可分性，说明文本信息能够为分类任务提供一定的判别信息。

与视觉表征相比，MMFN w/o U 学到的多模态表征具有更好的可分性；与文本表征相比，其可分性也略好。这表明，文本信息与视觉信息的结合能够提供互补信息，并提高表征的可分性

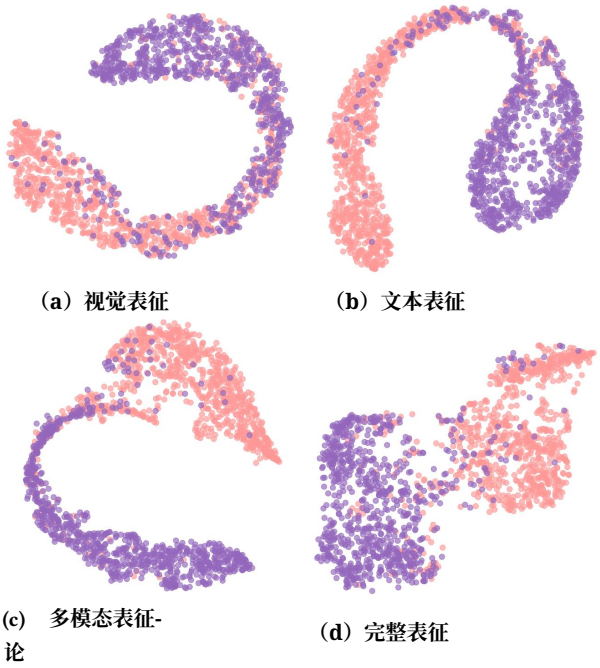


图3：微博测试数据集上不同模态表示的 T-SNE 可视化。

。然而，由于缺少单模态分支，MMFN w/oU 的效果不如完整模型。最后，完整 MMFN 学到的表征在所有变体中可分性最佳，甚至优于多模态表征。这表明，所提出的方法能够有效利用多模态表征来解决歧义问题，从而显著提升可分性。

总体而言，这些结果不仅证明了结合视觉信息和文本信息对微博虚假新闻检测的重要性，也表明所提出的融合方法能够有效组织多粒度多模态表征。

5 结论

本文提出了一种名为 MMFN 的新型多模态假新闻检测方法，该方法使用 BERT 和 Swin-T，并结合 Transformer 获取细粒度多模态特征；同时使用 CLIP 获取粗粒度多模态特征。此外，我们引入了结合 CLIP 相似度加权的单模态分支，以自适应地辅助多模态分类。我们在两个知名的多模态假新闻检测数据集上进行了全面实验。结果表明，MMFN 优于许多最先进方法。

致谢

本工作得到中国国家自然科学基金项目（批准号：U20B2051、62072114、U22B2047 和 U1936214）的资助。

参考文献

[1] Sahar Abdelnabi、Rakibul Hasan 和 Mario Fritz。2022年。利用在线资源对脱离语境的图像进行开放领域、基于内容的多模态事实核查。载于《IEEE/CVF计算机视觉与模式识别会议论文集》。14940–14949。[2] Liesbeth Allein、Marie-Francine Moens 和 Domenico Perrotta。2021年。《文章相似，受众相似：通过强化多模态相关性进行虚假信息检测》。arXiv预印本 arXiv:2108.13892（2021年）。[3] Tadas Baltrušaitis、Chaitanya Ahuja 和 Louis-Philippe Morency。2018年。多模态机器学习：综述与分类体系。《IEEE模式分析与机器智能汇刊》第41卷第2期（2018年），423–443。[4] Gaurav Bhatt、Aman Sharma、Shivam Sharma、Ankush Nagpal、Balasubramanian Raman 和 Ankush Mittal。2018年。结合神经、统计和外部特征进行假新闻立场识别。载于《万维网大会论文集》。1353–1357。[5] Tian Bian、Xi Xiao、Tingyang Xu、Peilin Zhao、Wenbing Huang、Yu Rong 和 Junzhou Huang。2020年。利用双向图卷积网络检测社交媒体上的谣言。载于《AAAI人工智能会议论文集》，第34卷。549–556。[6] Christina Boididou、Symeon Papadopoulos、Markos Zampoglou、Lazaros Apostolidis、Olga Papadopoulou 和 Yiannis Kompatsiaris。2018年。Twitter上误导性内容的检测与可视化。《国际多媒体信息检索期刊》第7卷第1期（2018年），71–86。[7] Yixuan Chen、Dongsheng Li、Peng Zhang、Jie Sui、Qin Lv、Lu Tun 和 Li Shang。2022年。用于多模态虚假新闻检测的跨模态歧义学习。载于《万维网大会论文集》。2897–2905。[8] Marcos V Conde 和 Kerem Turgutlu。2021年。CLIP-Art：用于细粒度艺术分类的对比预训练。载于《IEEE/CVF计算机视觉与模式识别会议论文集》。3956–3960。[9] Nadia K Conroy、Victoria L Rubin 和 Yimin Chen。2015年。自动欺骗检测：发现假新闻的方法。《信息科学与技术协会会议论文集》第52卷第1期（2015年），1–4。[10] Corentin Dancette、Remi Cadene、Damien Teney 和 Matthieu Cord。2021年。超越基于问题的偏差：评估视觉问答中的多模态捷径学习。载于《IEEE/CVF国际计算机视觉大会论文集》。1574–1583。[11] Jacob Devlin、Ming-Wei Chang、Kenton Lee 和 Kristina Toutanova。2018年。BERT：用于语言理解的深度双向Transformer预训练。arXiv预印本 arXiv:1810.04805（2018年）。[12] Alexey Dosovitskiy、Lucas Beyer、Alexander Kolesnikov、Dirk Weissenborn、Xiaohua Zhai、Thomas Unterthiner、Mostafa Dehghani、Matthias Minderer、Georg Heigold、Sylvain Gelly 等。2020年。一张图像胜过16×16个词：用于大规模图像识别的Transformer。arXiv预印本 arXiv:2010.11929（2020年）。[13] Xiuye Gu、Tsung-Yi Lin、Weicheng Kuo 和 Yin Cui。2021年。通过视觉与语言知识蒸馏实现开放词汇目标检测。arXiv预印本 arXiv:2104.13921（2021年）。[14] Zhiwei Jin、Juan Cao、Han Guo、Yongdong Zhang 和 Jiebo Luo。2017年。利用循环神经网络进行多模态融合，以检测微博上的谣言。载于《ACM国际多媒体会议论文集》。795–816。[15] Gregory Johnson。2012年。谷歌翻译 http://translate. google. com。《技术服务季刊》第29卷第2期（2012年），165–165。[16] Dhruv Khattar、Jaipal Singh Goud、Manish Gupta 和 Vasudeva Varma。2019年。MVAE：用于虚假新闻检测的多模态变分自编码器。载于《万维网大会论文集》。2915–2921。[17] Diederik P Kingma 和 Jimmy Ba。2014年。Adam：一种随机优化方法。arXiv预印本 arXiv:1412.6980（2014年）。[18] Boyi Li、Kilian Q Weinberger、Serge Belongie、Vladlen Koltun 和 René Ranftl。2022年。语言驱动的语义分割。arXiv预印本 arXiv:2201.03546（2022年）。[19] Ze Liu、Yutong Lin、Yue Cao、Han Hu、Yixuan Wei、Zheng Zhang、Stephen Lin 和 Baining Guo。2021年。Swin Transformer：采用移位窗口的分层视觉Transformer。载于《IEEE/CVF国际计算机视觉大会论文集》。10012–10022。[20] Jiasen Lu、Dhruv Batra、Devi Parikh 和 Stefan Lee。2019年。VilBERT：面向视觉与语言任务的预训练任务无关视觉语言表征。载于《神经信息处理系统进展》第32卷（2019年）。[21] Jing Ma、Wei Gao、Prasenjit Mitra、Sejeong Kwon、Bernard J Jansen、Kam-Fai Wong 和 Meeyoung Cha。2016年。利用循环神经网络检测微博中的谣言。载于《国际人工智能联合会议论文集》。3818–3824。[22] Qiong Nan、Juan Cao、Yongchun Zhu、Yanyan Wang 和 Jintao Li。2021年。MD-FEND：多领域假新闻检测。载于《ACM国际信息与知识管理会议论文集》。3343–3347。[23] Alex Nichol、Prafulla Dhariwal、Aditya Ramesh、Pranav Shyam、Pamela Mishkin、Bob McGrew、Ilya Sutskever 和 Mark Chen。2021 年。GLIDE：基于文本引导的扩散模型实现照片级真实感图像生成与编辑。arXiv预印本 arXiv:2112.10741（2021年）。[24] Peng Qi、Juan Cao、Xirong Li、Huan Liu、Qiang Sheng、Xiaoyue Mi、Qin He、Yongbiao Lv、Chenyang Guo 和 Yingchao Yu。2021年。使用实体增强框架融合多样化多模态线索以改进虚假新闻检测。载于《ACM国际多媒体会议》论文集。1212–1220。[25] Peng Qi、Juan Cao、Tianyun Yang、Junbo Guo 和 Jintao Li。2019年。利用多领域视觉信息进行虚假新闻检测。载于《IEEE国际数据挖掘会议》论文集。IEEE，518–527。[26] Shengsheng Qian、Jinguang Wang、Jun Hu、Quan Fang 和 Changsheng Xu。2021年。用于虚假新闻检测的分层多模态上下文注意力网络。载于《国际ACM SIGIR信息检索研究与发展会议》论文集。153–162。[27] Alec Radford、Jong Wook Kim、Chris Hallacy、Aditya Ramesh、Gabriel Goh、Sandhini Agarwal、Girish Sastry、Amanda Askell、Pamela Mishkin、Jack Clark 等。2021年。从自然语言监督中学习可迁移视觉模型。载于《国际机器学习会议》。8748–8763。[28] Colin Raffel、Noam Shazeer、Adam Roberts、Katherine Lee、Sharan Narang、Michael Matena、Yanqi Zhou、Wei Li 和 Peter J Liu。2019年。探索迁移学习的局限：统一的文本到文本转换器。arXiv预印本 arXiv:1910.10683（2019年）。[29] Kai Shu、Deepak Mahudeswaran、Suhang Wang、Dongwon Lee 和 Huan Liu。2020年。FakeNewsNet：一个包含新闻内容、社会语境及时空信息、用于研究社交媒体上假新闻的数据仓库。《大数据》8，3（2020），171–188。[30] Shivangi Singhal、Anubha Kabra、Mohit Sharma、Rajiv Ratn Shah、Tanmoy Chakraborty 和 Ponnurangam Kumaraguru。2020年。Spotfake+：通过迁移学习进行虚假新闻检测的多模态框架（学生摘要）。载于《AAAI人工智能会议》论文集，第34卷。13915–13916。[31] Shivangi Singhal、Tanisha Pandey、Saksham Mrig、Rajiv Ratn Shah 和 Ponnurangam Kumaraguru。2022年。利用模态内与模态间关系进行多模态假新闻检测。载于《2022年万维网大会》伴随论文集。726–734。[32] Shivangi Singhal、Rajiv Ratn Shah、Tanmoy Chakraborty、Ponnurangam Kumaraguru 和 Shin'ichi Satoh。2019年。Spotfake：用于虚假新闻检测的多模态框架。载于《IEEE国际多媒体大数据会议》论文集。39–47。[33] Laurens Van der Maaten 和 Geoffrey Hinton。2008年。使用 t-SNE 可视化数据。《机器学习研究杂志》9，11（2008年）。[34] Ashish Vaswani、Noam Shazeer、Niki Parmar、Jakob Uszkoreit、Llion Jones、Aidan N Gomez、Łukasz Kaiser 和 Illia Polosukhin。2017年。注意力就是你所需要的一切。《神经信息处理系统进展》30（2017年）。[35] Yaqing Wang、Fenglong Ma、Zhiwei Jin、Ye Yuan、Guangxu Xun、Kishlay Jha、Lu Su 和 Jing Gao。2018年。EANN：用于多模态假新闻检测的事件对抗神经网络。载于《ACM国际知识发现与数据挖掘会议》论文集。849–857。[36] Youze Wang、Shengsheng Qian、Jun Hu、Quan Fang 和 Changsheng Xu。2020年。基于知识驱动的多模态图卷积网络进行虚假新闻检测。载于《国际多媒体检索会议》论文集。540–547。[37] Tianyi Wei、Dongdong Chen、Wenbo Zhou、Jing Liao、Zhentao Tan、Lu Yuan、Weiming Zhang 和 Nenghai Yu。2021年。Hairclip：通过文本和参考图像设计发型。arXiv预印本 arXiv:2112.05142（2021年）。[38] Yang Wu、Pengwei Zhan、Yunjian Zhang、Liming Wang 和 Zhen Xu。2021年。利用协同注意力网络进行假新闻检测的多模态融合。载于《计算语言学协会：ACL-IJCNLP 2021》论文集。2560–2569。[39] Junxiao Xue、Yabo Wang、Yichen Tian、Yafei Li、Lei Shi 和 Lin Wei。2021年。通过探索多模态数据的一致性检测虚假新闻。《信息处理与管理》58，5（2021年），102610。[40] Keren Ye 和 Adriana Kovashka。2021年。视觉常识推理中的捷径效应案例研究。载于《AAAI人工智能会议》论文集，第35卷。3181–3189。[41] Qichao Ying、Xiaoxiao Hu、Yangming Zhou 等。2023年。为虚假新闻检测自举多视图表示。载于 AAAI。[42] Feng Yu、Qiang Liu、Shu Wu、Liang Wang、Tieniu Tan 等。2017年。用于虚假信息识别的卷积方法。载于 IJCAI。3901–3907。[43] Xueyao Zhang、Juan Cao、Xirong Li、Qiang Sheng、Lei Zhong 和 Kai Shu。2021年。挖掘双重情绪进行虚假新闻检测。载于《2021年万维网大会》论文集。3465–3476。[44] Xinyi Zhou、Jindi Wu 和 Reza Zafarani。2020年。SAFE：相似性感知的多模态假新闻检测。载于《亚太知识发现与数据挖掘会议》。Springer，354–367。[45] Yangming Zhou、Yuzhou Yang、Qichao Ying、Zhenxing Qian 和 Xinpeng Zhang。2022年。通过 CLIP 引导学习进行多模态假新闻检测。arXiv预印本 arXiv:2205.14304（2022年）。[46] Yongchun Zhu、Qiang Sheng、Juan Cao、Qiong Nan、Kai Shu、Minghui Wu、Jindong Wang 和 Fuzhen Zhuang。2022年。记忆引导的多视图多领域虚假新闻检测。《IEEE知识与数据工程汇刊》

（2022年）。[47] Arkaitz Zubiaga、Ahmet Aker、Kalina Bontcheva、MariaLiakata 和 Rob Procter。2018年。社交媒体中的谣言检测与核查：一项综述。*ACM* 计算机调查（*CSUR*）51，2（2018年），1–36。