Factify 2 中的 INO：基于结构连贯性的多模态事实核查

一诺zhang[1], 竹林涛[1], 曦王[2] 和同月王[1]

[1]中国传媒大学，中国北京[2]中国科学院信息工程研究所，中国北京

arXiv:2303.01510v1 [cs.LG] 2 Mar 2023

摘要

本文介绍了我们参加 AAAI 2023 多模态事实核查（FACTIFY）挑战赛的方法。近年来，随着社交媒体的广泛使用，虚假新闻能够迅速传播，并对社会安全产生负面影响。自动化主张核验对于打击虚假新闻变得越来越重要。在涉及多项模态数据的事实核查中，主张与文档之间应当具有结构连贯性。因此，我们提出了一种基于结构连贯性的多模态事实核查方案，用于对虚假新闻进行分类。我们的结构连贯性包括以下四个方面：句子长度、词汇相似度、语义相似度和图像相似度。具体而言，我们结合 CLIP 和 Sentence BERT 提取文本特征，并使用 ResNet50 提取图像特征。此外，我们还提取文本长度以及词汇相似度。随后，将这些特征拼接起来，并输入随机森林分类器。最终，我们的加权平均 F1 分数达到 0.8079，在 FACTIFY2 中获得了第 2 名。代码可见于 https://github.com/Catrin-baze/INO-of-factify。

关键词

事实核查、假新闻检测、多模态、DE-FACTIFY

1. 引言

近年来，随着社交媒体的广泛应用，以及深度伪造等新技术的发展，虚假新闻的数量激增。例如，研究 [1]指出，在大选前的三个月里，支持特朗普的虚假信息在 Facebook 上总共被分享了 3000 万次，而支持克林顿的信息被分享了 800 万次。社交媒体的快速发展导致虚假新闻广泛传播，不仅影响人们的生活、引发公众恐慌、扰乱社会秩序、影响公众舆论并操纵公众关注焦点，还损害了社交媒体平台的公信力。2018年，发表于《科学》的文章《假新闻科学》 [2] 指出，在所有信息类别中，虚假信息的传播范围显著更广、速度显著更快、渗透更深、影响更广。因此，有效检测

社交媒体上的虚假新闻传播受到遏制，对于维护社会稳定和网络空间安全具有重要意义。

Woodstock'22：不可复现科学研讨会，*2022*年*6*月*7*日至*11*日，纽约州伍德斯托克 *通讯作者。 catrinbaze@gamil.com (Y. Zhang);taozhulin@gmail.com (Z. Tao); wangxiboss@163.com (X. Wang); apri.sunmoon@gmail.com (T. Wang) 0000-0001-7910-7983 (Y. Zhang); 0000-0001-9011-8464 (Z. Tao); 0000-0002-6642-8160 (X. Wang); 0000-0003-1897-4420 (T. Wang)©2022 年本文作者版权所有。允许依据知识共享署名 4.0 国际许可协议 (CC BY 4.0) 使用。CEUR 研讨会论文集 (CEUR-WS.org)

为应对虚假新闻带来的挑战，人工智能和深度学习技术有望发挥关键作用。自2017年以来，美国国防高级研究计划局（DARPA）举办了“媒体取证挑战赛” [3]竞赛，以推动虚假信息检测技术快速发展。值得注意的是，碎片化的数字媒体环境为虚假信息不受控制地传播提供了温床。结合图像与文本的多模态虚假新闻更难识别，对自动多模态虚假新闻检测构成了严峻挑战。与此同时，多种模态的结合也已应用于推荐系统[4][5]等多个领域。为解决多模态虚假新闻检测问题，研究人员提出了许多有代表性的方法[6][7][8] 和数据集。利用多模态信息检测虚假新闻具有诸多优势，因为不同模态能够捕捉新闻文章的不同维度，在评估文章真实性时可以相互补充。

在本文中，我们的核心思想是从多个角度全面比较主张与文档之间的相关性，并以此作为分类的基础。因此，我们提出了一种基于结构连贯性的多模态事实核查方法。我们从结构连贯性的角度全面比较主张与文档的一致性，其中包括四个方面：句子长度、词汇相似度、语义相似度和图像相似度。实验验证了我们方法的有效性。

本文其余部分的组织如下。第2节介绍数据描述和任务定义。第3节介绍相关工作。第4节描述我们的方法，第5节讨论实验及其结果。本文最后给出我们的结论。

2. 任务与数据集

2.1. 任务定义

多模态事实验证任务 FACTIFY2[9][10] 本质上被建模为多模态蕴含。每个数据点包含一个待检测的主张和一份可靠的来源文档。两者都是多模态的，且包含一个图文对。因此，该任务提供文本主张、主张图像、文档和文档图像。系统必须将数据样本分类为五个类别之一。标签说明如下。

- `Support_Text`： 主张文本相似或蕴含，但文档与主张的图像并不相似；
- `Support_Multimodal`： 主张的文本和图像都与文档中的相似；
- `Insufficient_Text`： 文档既不支持也不反驳主张的文本和图像，尽管主张文本可能与文档文本包含一些相同的词；
- `Insufficient_Multimodal`： 文档既不支持也不反驳该主张文本，但图像与文档相似；

- `Refute`： 主张和文档中的图像和/或文本完全矛盾，即该主张是错误的/虚假的。

Fig结构1 展示了 FACTIFY 中的数据示例，FACTIFY2 的数据格式与之相同 .



(a) (b)

图 1： 数据示例。

2.2. 数据描述

数据来自印度和美国新闻媒体在推特上的账号：印度的《印度斯坦时报》[1],ANI[2] 和美国的 ABC[3], CNN[4] ，账号的选择依据是可访问性、受欢迎程度和每日发帖量。此外，这些推特账号以客观、中立的态度著称。该数据集共包含 50000 个样本，五个类别的样本数量均相同。数据集按 70:15:15 的比例划分为训练集-验证集-测试集。

3. 相关工作

3.1. 假新闻检测

近年来，研究人员提出了许多有效的虚假新闻检测方法。Qian等人[11]展示了文本内容和语义信息如何帮助检测虚假新闻。在 [12], 中，作者使用一种名为 TM 的新框架，捕捉真实和虚假新闻文本的词汇及语义特征，以检测虚假新闻。此外，一些违法者会通过伪造图片来篡改新闻。因此，Qi等人[13]提出了一种新颖的模型 MVNN，该模型利用频率域和像素域中的视觉信息来检测虚假新闻。

尽管上述所有单模态技术都取得了良好效果，但它们仍然忽视了这样一个事实：新闻事件通常同时包含文本和图片，而二者可以相互补充。这表明，我们需要一种用于虚假新闻检测的多模态系统。现有的基于多模态信息的虚假新闻检测方法可分为三

[1]https://twitter.com/htTweets [2]
https://twitter.com/ANI [3]
https://twitter.com/ABC [4]
https://twitter.com/CNN

类。有些方法将文本信息与视觉信息相结合。Singhal 等人[14] 提出了一种名为 SpotFake 的新模型，将从在 ImageNet 上预训练的 VGG-19 模型中提取的视觉特征与文本特征进行拼接，并通过全连接层进行分类。Wang 等人[15]构建了一种名为 E ANN（事件对抗神经网络）的端到端模型，用于学习并拼接文本和视觉的潜在表示，然后将其输入两个全连接神经网络分类器。其他研究则侧重于对比多模态信息。SAFE[7], 是一种感知相似性的多模态方法，通过比较跨模态提取特征之间的相关性来检测虚假新闻。此外，一些方法利用多模态信息增强来预测虚假新闻。Jin 等人[8]使用带有注意力机制的循环神经网络，增强多模态特征之间的信息理解，从而实现有效检测。

3.2. 事实核查

一般而言，事实核查是指依据经过验证的文档语料，评估某项主张真实性的任务。事实核查将客观参考资料和外部知识引入假新闻检测，其可靠性高于单纯依靠新闻文本和视觉特征作出判断。事实核查还能提高用户的防范意识，帮助他们在接触到虚假新闻时尝试查找事实核查信息。

事实核查数据集有多种类型，其中许多是单模态数据集。例如，Thorne 等人 [16] 将维基百科作为文本证据来源，并标注支持或反驳每项主张的句子。另一种常见的非结构化证据数据集通常采用元数据，它能提供文本来源的补充信息[17]。尽管大多数单模态数据集关注非结构化证据，结构化知识也得到过应用。一些数据集由半结构化数据表组成，能够简洁而灵活地传递信息。Wang 等人[18] 从科学文章中提取表格，并要求以从表格中选取单元格的形式进行证据选择。

近年来，事实核查也开始考虑构建多模态数据集，以提高虚假新闻检测的准确性。一个名为 Mocheg[19] 的数据集由文本和视觉证据组成，并使用三个标签：支持、反驳和信息不足。许多事实核查方法侧重于主张核验，这可以被视为自然语言推理任务或文本蕴含识别（RTE）的一种形式。Nie 等人[20]使用 ESIM 对主张进行核验，将所有证据片段拼接后作为输入，并使用最大池化来聚合信息。RTE 的典型检索策略包括搜索 API、Lucene 索引等 [21]。在 [22], 中，使用学习所得表示和快速点积索引的稠密检索器表现良好。Fan 等人[23] 提出了另一种检索证据的方法，即通过搜索引擎结果进行问题生成和问答。

4. 方法论

在我们的论文中，我们设计了一种基于结构连贯性的事实核查方法，用于计算主张与文档之间的结构连贯性。结构连贯性体现在以下四个方面：字面文本相似度、文本语义相似度、文本长度和图像相似度。具体而言，我们使用 ROUGE[24] 提取文本的字面相似度，并使用两种

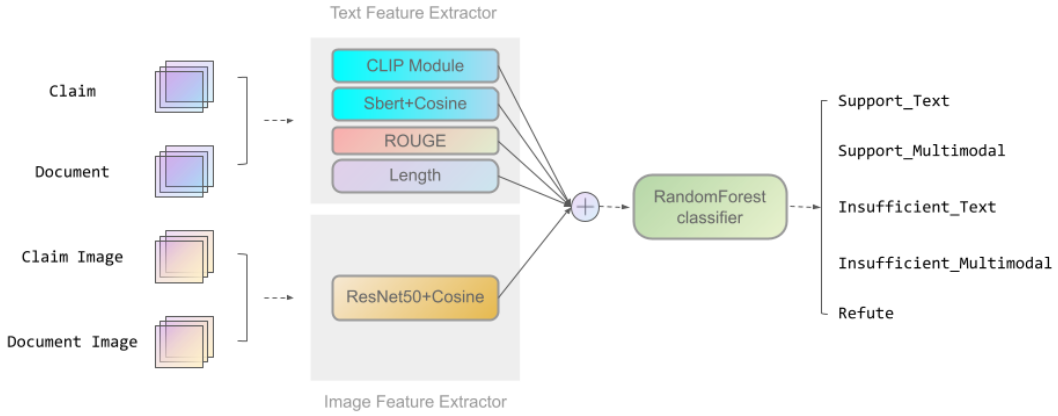


图2： 我们方法的总体架构。该架构包括两个部分：文本特征提取和图像特征提取。文本特征提取中的 CLIP 模块将在第4节中详细介绍。

CLIP[25] 和 Sentence BERT[26] 预训练模型计算文本的语义相似度，并计算主张和文档的长度。随后，使用 ResNet50 提取图像特征。将上述特征拼接后输入随机森林分类器，以获得最终分类结果。

多模态模型的架构如图 2所示。

4.1. 文本特征提取器

在特征提取模块中，我们提取了 ROUGE 和文本长度这两项特征。ROUGE 是面向摘要评估的召回率研究的缩写。它是一种评估指标，通过将摘要与人工创建的其他（理想）摘要进行比较，包含用于自动确定摘要质量的测量方法。这些测量方法统计待评估的计算机生成摘要与人工创建的理想摘要之间重叠单元的数量，例如 n元语法、词序列和词对。由于我们观察到反驳类别中存在大量词汇重叠，因此统计重叠单元数量的 ROUGE 非常适合我们的任务。同时，受 Gao等人[27], 的启发，我们还计算了文本长度，并发现文本长度与类别之间存在特定的相关性，因此也提取了文本长度特征。

在预训练模型模块中，我们将 Sentence BERT 和余弦相似度用作基线[25]。此外，我们还尝试使用预训练模型 CLIP、SimCSE[28], 和 RoBERTa[29]来提取文本特征。CLIP 是一种双流模型，它通过 Transformer 编码器处理文本和图像，并通过线性投影计算不同图文对之间的相似度。同时，它利用对比学习将图像分类转化为图文匹配任务。CLIP 使用来自网络的 4 亿组图文数据集，并将文本用作图像标签进行训练。SimCSE 是一种用于比较句子向量表示的简单框架，在句子嵌入任务中表现出色。RoBERTa 是 BERT 的增强版本，而经过微调的 BERT 模型在模型规模和训练方法方面有所改进

BERT[30] 。在这一部分，我们尝试使用 CLIP 文本编码器、SimCSE 和 RoBERTa 提取文本特征，并使用余弦相似度计算语义相似度。我们还将 CLIP 提取的特征输入多层感知机层，并输出三分类结果。最后，我们发现，将 CLIP 与多层感知机结合得到的三分类结果作为特征，更有助于提高最终 F1 分数。

4.2. 图像特征提取器

在本部分中，我们使用预训练的 ResNet50[31] 和余弦相似度。同时，我们还尝试了另外两种方案：使用 CLIP 提取图像特征并计算余弦相似度；将 CLIP 特征输入多层感知机，以输出三分类结果。

4.3. 分类器

最后，将提取的特征进行拼接，并输入随机森林分类器，以输出五分类结果。随机森林是一种元估计器，它在数据集的不同子样本上拟合多个决策树分类器，并通过取平均值来提高预测准确率和控制过拟合。在基线提供的分类器中，我们选择了在我们的特征上表现最佳的随机森林分类器。

5. 实验与结果

5.1. 实验设置

为评估我们的基线方案的性能，我们使用加权平均 F1 在验证集上进行基准测试。所有实验均在一块 NVIDIA Tesla T4 GPU 上完成，最多配备 15 GiB 随机存取存储器。对于 CLIP 模块，我们使用预训练文本编码器，仅训练后续的多层感知机层。多层感知机包含一个具有 100 个节点的隐藏层和 Adam[32] 优化器。由于文本长度与其他特征的数量级差异很大，我们对最终特征进行归一化。最终的随机森林分类器将 max_depth 值设为 40 时表现最佳。

5.2. 结果

我们的模型在测试集上取得了 0.8079 的 F1 分数，竞赛排行榜如表1所示。我们的方法在本次竞赛中排名第 2。

在探索解决方案的过程中，我们的尝试大致分为以下两个方面：1）文本预训练模型的选择；2）CLIP 的使用方式。

在实验过程中，我们尝试使用 SimCSE、RoBERTa 和 CLIP 的文本编码器替代 Sentence BERT。具体而言，我们将这些预训练模型用作文本特征提取器，然后使用余弦相似度计算主张与文档嵌入之间的相似度。我们将仅替换 Sentence BERT 的实验结果与基线进行了比较。验证集上的 F1 分数如表 2所示。

可以看出，使用文本预训练模型提取文本特征并直接计算余弦相似度时，其他方法的 F1 分数均低于 Sentence BERT。

表 1 Factify 官方排行榜

Rank	Team	支持_Text	支持_多项	不足_Text	不足_多项	反驳	最终
1	Triple-Check	0.8277	0.9138	0.8518	0.8922	1	0.8182
2	**INO**	**0.8124**	**0.9003**	**0.8881**	**0.8523**	**0.9993**	**0.8080**
3	从逻辑上讲	0.8038	0.9051	0.8439	0.8563	0.9851	0.7897
4	zhang	0.7664	0.8785	0.8161	0.8783	0.9993	0.7742
5	gzw	0.7849	0.8632	0.8142	0.8327	1	0.7605
6	coco	0.7725	0.8649	0.8152	0.8300	1	0.7570
7	Noir	0.7710	0.8726	0.7849	0.8156	0.9970	0.7452
8	Yet	0.7075	0.8263	0.7859	0.7190	1	0.6909
9	TeamX	0.5822	0.7091	0.5366	0.5556	0.6979	0.4562
-	基线	0.5	0.8272	0.8024	0.7593	0.9882	0.6499

表2将 Sentence BERT 替换为其他预训练模型

模型名称	验证集 F1 分数
SimCSE+ResNet50	0.6442
RoBERTa+ResNet50	0.6424
CLIP_文本+ResNet50	0.6404
CLIP_文本+CLIP_图像	0.5103
Sentence BERT+ResNet50（基线）	0.6664

同时，我们还尝试使用 CLIP 提取图像特征并计算余弦相似度，以替代 ResNet50。结果仅达到 0.5103 的 F1 分数，远低于基线。因此，在不进行微调的情况下，仅替换预训练模型并计算余弦相似度，并不能提升分类性能。因为不同的预训练模型本质上只是改变了向量的表示形式，无论向量的维度有多大，计算余弦相似度后都只能得到一个分数。然而，一个分数所包含的信息是有限的。因此，我们仍然保留 Sentence BERT 和 ResNet50 的结构。

此外，我们还尝试了以不同方式使用 CLIP。由于 CLIP 是一个多模态模型，包含一个文本编码器和一个图像编码器，二者均输出 512 维向量。因此，使用 CLIP 从数据中提取特征后，我们得到主张、文档、主张图像、文档图像，以及四个 512 维向量。因此，我们尝试了三种特征组合方法来使用 CLIP 模块，希望将其作为特征来提升最终分类效果。1）将文本特征向量拼接后输入多层感知机层，进行三分类；2）将图像特征向量拼接后输入多层感知机层，进行三分类；3）将所有图像和文本特征向量拼接后输入多层感知机层，进行五分类。前两种方法如图 3 所示，第三种方法如图4 所示。表 3 展示了模型使用不同 CLIP 模块时在验证集上的 F1 分数。

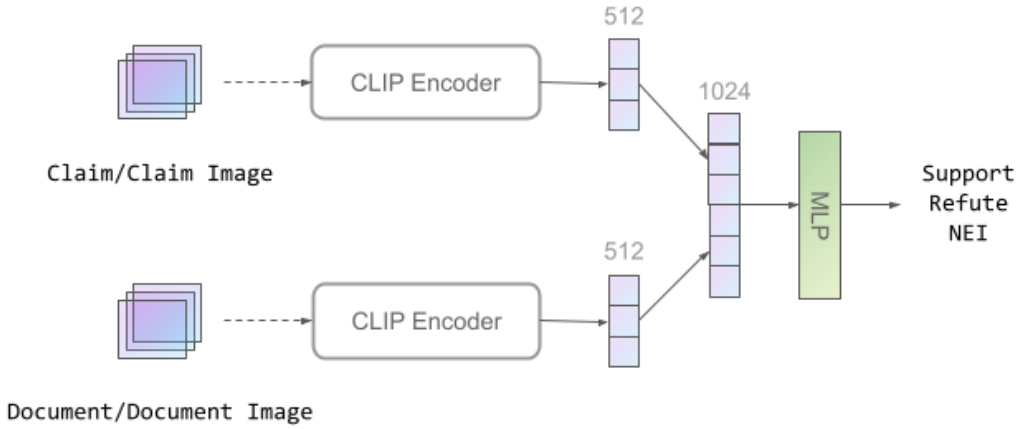


图 3： 在领域内使用 CLIP 的方法 1/2。具体而言，使用 CLIP 的文本/图像编码器提取主张和文档/主张_图像和文档_图像的嵌入，并将其输入多层感知机进行三分类。

表 3不同的 CLIP 模块

CLIP 模块	验证集 F1 分数
CLIP+多层感知机	0.7330
CLIP_文本+多层感知机+CLIP_图像+多层感知机	0.7291
CLIP_图像+多层感知机	0.7092
CLIP_文本+多层感知机（我们的方法）	0.8078

可以看出，使用 CLIP 模块的不同方式对分类结果有较大影响。其他使用方式无法提升模型效果，甚至会使效果变差。因为该模块提取的特征实际上只是一个多类别标签。因此，分类不准确可能会影响其他特征。仅使用 CLIP 提取图像特征的方法效果最差。我们认为，CLIP 和 ResNet50 对图像的分类结果相互排斥，因此加入这一特征会降低效果。

最后，我们针对模型中的不同模块进行了消融实验，结果如表 4 所示。可以看出，移除 ResNet50 模块对模型效果的影响最为显著。这一结果在意料之中，因为 ResNet50 特征是模型中唯一的图像特征。此外，对模型贡献最大的特征是 ROUGE 和文本长度特征。同时，移除任何文本特征都不会对结果产生显著影响。图5 展示了最终结果在验证集和测试集上的混淆矩阵。

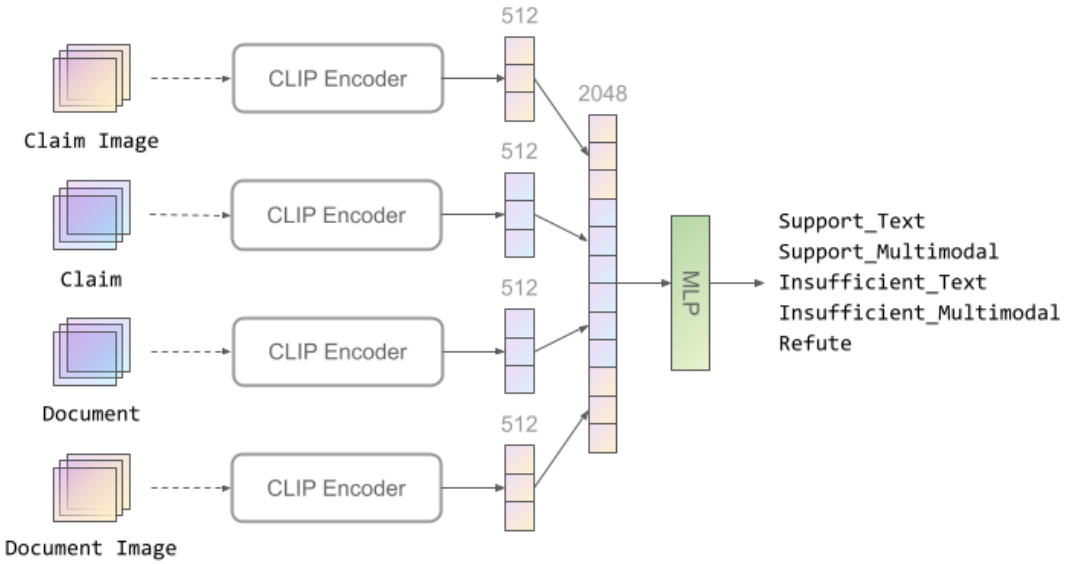


图 4：第三个 CLIP 模块直接拼接 CLIP 提取的所有嵌入，并训练多层感知机输出五种分类结果。

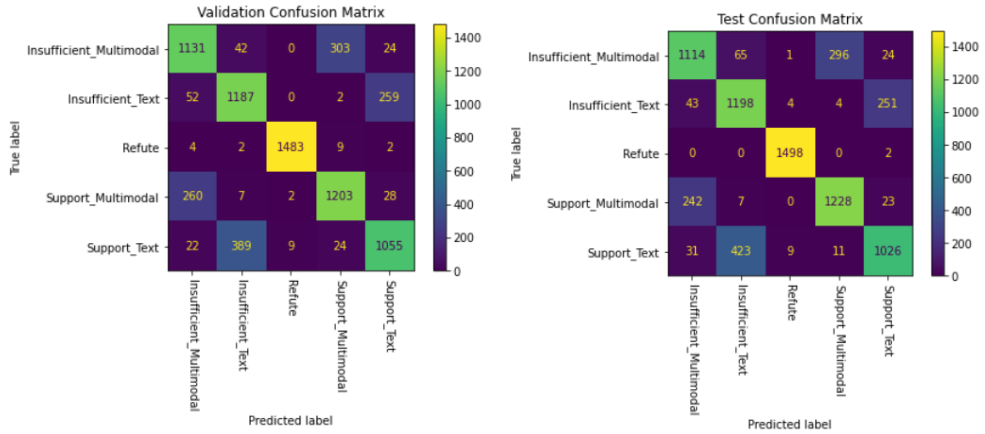
表 4消融研究

模型名称	验证集 F1 分数
不使用 Sentence BERT	0.7926
不使用 CLIP	0.7911
不使用 ROUGE+长度	0.7709
不使用 ResNet50	0.6007
基线	0.6664
Ours	0.8078

6. 结论

本文介绍了一种用于多模态事实核验任务的基于结构连贯性的方法。我们尝试了不同的预训练模型，例如 SimCSE、CLIP、RoBERTa，以及使用 CLIP 多模态特征的不同方式。最终，结合 Sentence BERT 和 CLIP 提取文本特征，并使用 ResNet50 提取图像特征，取得了最佳 F1 分数。

未来的工作可以从四个方向展开：1）改进数据预处理，以解决数据噪声和多语言性等问题；2）我们的方法未能充分利用图像特征，可考虑引入其他特征来补充 ResNet50，或在未来尝试其他方法；3）探索更有效的 CLIP 使用方式，或采用其他方法增强模态融合；4）提高假新闻检测模型的通用性，并将其扩展到更多类型的数据集。



(a) 在验证集上　　(b) 在测试集上

图 5：验证集和测试集的混淆矩阵。

7. 致谢

本研究得到国家重点研发计划（编号：2020YFB1406800）、中央高校基本科研业务费（编号：CUC22WH002）和国家自然科学基金（61702502）的支持。感谢 DE-FACTIFY 2023 的组织者允许我们使用该数据集开展研究。我们还感谢 Google Colab 提供图形处理器和深度学习代码运行平台，并感谢 scikit-learn 和 Hugging Face 提供高效便捷的深度学习库。

参考文献

[1] H. Allcott，M. Gentzkow，《2016年选举中的社交媒体与虚假新闻》，《经济展望杂志》31（2017年）211–36。[2] D. M. Lazer，M. A. Baum，Y. Benkler，A. J. Berinsky，K. M. Greenhill，F. Menczer，M. J. Metzger，B. Nyhan，G. Pennycook，D. Rothschild等，《假新闻科学》，《科学》359（2018年）1094–1096。[3] X. Jin，Y. Lee，J. G. Fiscus，H. Guan，A. Yates，A. Delgado，D. F. Zhou等，《媒体取证挑战赛：大型基准数据集上的图像溯源评估与最新技术分析》（2020年）。[4] Z. Tao，X. Liu，Y. Xia，X. Wang，L. Yang，X. Huang，T.-S. Chua，《用于多媒体推荐的自监督学习》，《IEEE多媒体汇刊》（2022年）。[5] X. Liu，Z. Tao，J. Shao，L. Yang，X. Huang，《Elimrec：消除多媒体推荐中的单模态偏差》，载于：《第30届ACM国际多媒体会议论文集》，MM '22，计算机协会，2022年。[6] S. Singhal，A. Kabra，M. Sharma，R. R. Shah，T. Chakraborty，P. Kumaraguru，《SpotFake+：一种通过迁移学习检测虚假新闻的多模态框架（学生摘要），

载于：AAAI人工智能会议论文集，第34卷，2020年，第13915–13916页。[7] X. Zhou，J. Wu，R. Zafarani，《SAFE：面向相似性感知的多模态虚假新闻检测》（2020年），预印本。arXiv 200304981（2020年）。[8] Z. Jin，J. Cao，H. Guo，Y. Zhang，J. Luo，《利用循环神经网络融合多模态信息以检测微博上的谣言》，载于：第25届ACM国际多媒体会议论文集，2017年，第795–816页。[9] S. Suryavardan，S. Mishra，P. Patwa，M. Chakraborty，A. Rani，A. Reganti，A. Chadha，D. Amitava，A. Sheth，C. Manoj，S. Ekbal，Asif Kumar，《Factify 2：多模态虚假新闻与讽刺新闻数据集》，载于：DeFactify 2：第二届多模态事实核查与仇恨言论检测研讨会论文集，CEUR，2023年。[10] S. Suryavardan，S. Mishra，M. Chakraborty，P. Patwa，A. Rani，A. Chadha，A. Reganti，A. Das，A. Sheth，M. Chinnakotla，A. Ekbal，S. Kumar，《Factify 2研究成果：多模态虚假新闻检测》，载于：DeFactify 2：第二届多模态事实核查与仇恨言论检测研讨会论文集，CEUR，2023年。[11] F. Qian，C. Gong，K. Sharma，Y. Liu，《神经用户响应生成器：利用群体用户智能检测虚假新闻》，载于：IJCAI，第18卷，2018年，第3834–3840页。[12] B. Bhattarai，O.-C. Granmo，L. Jiao，《用于虚假新闻检测及可信度评分评估的可解释特斯林机框架》，arXiv预印本 arXiv:2105.09114（2021年）。[13] P. Qi，J. Cao，T. Yang，J. Guo，J. Li，《利用多领域视觉信息检测虚假新闻》，载于：2019年IEEE国际数据挖掘会议（ICDM），IEEE，2019年，第518–527页。[14] S. Singhal，R. R. Shah，T. Chakraborty，P. Kumaraguru，S. Satoh，《SpotFake：一种虚假新闻检测多模态框架》，载于：2019年IEEE第五届国际多媒体大数据会议（BigMM），IEEE，2019年，第39–47页。[15] Y. Wang，F. Ma，Z. Jin，Y. Yuan，G. Xun，K. Jha，L. Su，J. Gao，《EANN：用于多模态虚假新闻检测的事件对抗神经网络》，载于：第24届ACM SIGKDD知识发现与数据挖掘国际会议论文集，2018年，第849–857页。[16] J. Thorne，A. Vlachos，C. Christodoulopoulos，A. Mittal，《FEVER：用于事实抽取与核验的大规模数据集》，arXiv预印本 arXiv:1803.05355（2018年）。[17] M. Potthast，J. Kiesel，K. Reinartz，J. Bevendorff，B. Stein，《对超党派新闻与虚假新闻的文体计量研究》，arXiv预印本 arXiv:1702.05638（2017年）。[18] N. X. Wang，D. Mahajan，M. Danilevsky，S. Rosenthal，《SemEval-2021 任务9：科学文献中表格数据的事实核查与证据查找（Sem-Tab-Facts）》，arXiv预印本 arXiv:2105.13995（2021年）。[19] B. M. Yao，A. Shah，L. Sun，J.-H. Cho，L. Huang，《端到端多模态事实核查与解释生成：一项具有挑战性的数据集与模型》，arXiv预印本 arXiv:2205.12487（2022年）。[20] Y. Nie，H. Chen，M. Bansal，《结合事实抽取与核验的神经语义匹配网络》，载于：AAAI人工智能会议论文集，第33卷，2019年，第6859–6866页。[21] J. Thorne，A. Vlachos，O. Cocarascu，C. Christodoulopoulos，A. Mittal，《事实抽取与核验（FEVER）共享任务》，arXiv预印本 arXiv:1811.10971（2018年）。[22] J. Maillard，V. Karpukhin，F. Petroni，W.-t. Yih，B. Oğuz，V. Stoyanov，G. Ghosh，多任务

面向知识密集型任务的检索，arXiv预印本 arXiv:2101.00117（2021年）。[23] A. Fan、A. Piktus、F. Petroni、G. Wenzek、M. Saeidi、A. Vlachos、A. Bordes、S. Riedel，生成事实核查简报，arXiv预印本 arXiv:2011.05448（2020年）。[24] C.-Y. Lin，Rouge：用于摘要自动评估的软件包，载于：文本摘要分支拓展，2004年，第74–81页。[25] N. Reimers、I. Gurevych，Sentence-BERT：使用孪生BERT网络的句子嵌入，arXiv预印本 arXiv:1908.10084（2019年）。[26] J. Gao、H.-F. Hoffmann、S. Oikonomou、D. Kiskovski、A. Bandhakavi，Logically at the Factify 2022：多模态事实核查，arXiv预印本 arXiv:2112.09253（2021年）。[27] S. Mishra、S. Suryavardan、A. Bhaskar、P. Chopra、A. Reganti、P. Patwa、A. Das、T. Chakraborty、A. Sheth、A. Ekbal等，Factify：一个多模态事实验证数据集，载于：第一届多模态事实核查与仇恨言论检测研讨会（DE-FACTIFY）论文集，2022年。[28] T. Gao、X. Yao、D. Chen，SimCSE：句子嵌入的简单对比学习，arXiv预印本 arXiv:2104.08821（2021年）。[29] Y. Liu、M. Ott、N. Goyal、J. Du、M. Joshi、D. Chen、O. Levy、M. Lewis、L. Zettlemoyer、V. Stoyanov，RoBERTa：一种稳健优化的BERT预训练方法，arXiv预印本 arXiv:1907.11692（2019年）。[30] J. Devlin、M.-W. Chang、K. Lee、K. Toutanova，BERT：面向语言理解的深度双向变换器预训练，arXiv预印本 arXiv:1810.04805（2018年）。[31] P. Kasnesis、R. Heartfield、X. Liang、L. Toumanidis、G. Sakellari、C. Patrikakis、G. Loukas，使用文本与图像相似度识别社交网络中的随机信息级联，《应用软计算》108（2021年）107413。[32] D. Kingma、J. Ba，Adam：一种随机优化方法，计算机科学（2014年）。