

智能情报融合综述：对抗视角下的开源情报融合分析

袁唯淋¹, 赵卫伟¹, 胡振震², 曹巍¹, 何俊¹, 董绍进¹, 王程远¹, 王盛青¹

(1. 国防科技大学信息通信学院, 湖北 武汉 430000;

2. 国防科技大学智能科学学院, 湖南 长沙 410073)

摘要: 在“信息开放共享、万物互通互联”的数字社会, 互联网中“海量、多源、爆炸式增长”的信息痕迹, 为开源情报提供了富饶的“矿藏”。自然语言处理和计算机视觉等先进人工智能技术赋能的情报处理, 极大提升了情报产出效率, 被广泛应用于安防领域。然而, 以深度伪造为代表的生成式人工智能打开了人工智能的“潘多拉魔盒”, 不断被用于制造数字野火, 投放虚假信息混淆视听, 给智能情报融合带来巨大的挑战。为此, 以开源情报为主要研究对象, 深入分析开源情报特点和现有挑战; 然后, 综述先进人工智能技术在开源情报融合中的发展现状, 从对抗视角总结现有的情报欺骗攻击方法和情报对抗防御方法; 最后, 提出未来智能情报融合新方向, 旨在为可信情报融合分析提供借鉴, 为智能态势分析与辅助决策提供支撑。

关键词: 开源情报融合; 情报欺骗攻击; 情报对抗防御; 深度造假; 舆情检测

中图分类号: TP181

文献标志码: A

doi: 10.11959/j.issn.2096-6652.202429

Research on intelligence fusion: a holistic analysis of open-source intelligence fusion from the perspective of confrontation

YUAN Weilin¹, ZHAO Weiwei¹, HU Zhenzhen², CAO Wei¹, HE Jun¹, DONG Shaojin¹,

WANG Chengyuan¹, WANG Shengqing¹

1. College of Information and Communication, National University of Defense Technology, Wuhan 430000, China

2. College of Intelligence Science and Technology, National University of Defense Technology, Changsha 410073, China

Abstract: In the digital society of “open sharing of information and interconnection of everything”, the information traces characterized with “massive, multi-source and explosive growth” in the internet provide rich “mineral deposits” for open-source intelligence collecting. Intelligence processing enabled by advanced artificial intelligence technologies, such as natural language processing and computer vision, greatly improves the efficiency of intelligence production and is widely used in the field of security. However, generative artificial intelligence represented by deepfake has opened the “Pandora’s box” of artificial intelligence and is used to create digital wildfires and put false information, which confuse the public and bring crucial challenges to the integration of intelligent intelligence. This paper focuses on open-source intelligence and deeply analyzes its characteristics and existing challenges. Then, this paper summarizes the development of advanced artificial intelligence technology in opensource intelligence fusion, summarizes the existing intelligence spoofing attack methods and intelligence anti-spoofing defense methods from the perspective of confrontation. Finally, new directions of intelligence fusion in the future are proposed to provide reference for the trusted intelligence fusion and provide support for intelligent situation analysis and auxiliary decision making.

Key words: open-source intelligence fusion, intelligence spoofing attack, intelligence anti-spoofing defense, deepfake, public opinion detection

收稿日期: 2024-04-01; 修回日期: 2024-08-10

通信作者: 赵卫伟, zhaozww@163.com

0 引言

“用兵之要，必先察敌情”。在巨复杂、高动态、不确定的强对抗环境中，情报是消除“战争迷雾”和破解对手“行动密码”的关键，旨在尽可能地为指挥员提供透明的决策环境，为获取信息和决策优势提供保障。从情报信息来源看，情报可以分为隐秘情报和开源情报。在“信息开放共享、万物互通互联”的数字社会，开源情报蕴藏着丰富的价值——互联网中各组织、机构和人员活动留下的“海量、多源、爆炸式增长”信息痕迹，为开源情报提供了富饶的“矿藏”。与传统的隐秘情报（如人力情报）相比，开源情报具有“低成本、低风险、高收益”等优点，能够为分析态势的发展和演变提供更加系统化的支撑，逐渐成为情报活动中的重要成分。美国中央情报中心曾指出“高达80%的情报可以从公开信息源获得”。然而，碎片化的情报数据还难以被当作情报产品直接使用。针对多模态和多平台传感器所获得的情报信息，如何“有效融合海量、多源异构情报数据，解决所获得的信息残缺、不精准等问题，完成对时空要素的扩展，从而得到准确、可靠的情报过程”成为当前情报研究的热点。

近年来，结合人工智能、大数据、云计算、物联网、5G等颠覆性技术，情报融合技术在民用安全和军事对抗领域不断取得显著成果。尤其是2022年爆发的俄乌冲突成为互联网时代开源情报最新的大型应用场景。现有情报研究综述缺乏对开源情报的综合性分析。文献[1]针对开源情报在网络安全检测中的应用，从网络安全角度综述了网络威胁情报挖掘技术。文献[2]综述了情报信息检索方法发展现状，概括了Web信息检索、智能信息检索、跨语言检索的一般方法。瑞士I-intelligence团队整理了较为全面的开源情报分析工具与资源，涵盖网站、事件与话题、数据库、图片与视频、新闻与文章、代码与文件等类别。文献[3]聚焦网络舆情情感分析问题，介绍当前多模态舆情分析算法，并与文本、图像单模态方法进行对比分析。文献[4]从情报对抗的角度分析总结了虚假新闻生成模型和检测方法。文献[5]综述了面向图像和视频的深度造假方法和对抗检测方法。与现有情报相关综述不同，本文主要以开源情报为主要研究对象，介绍人工智能技术在开源情报融合中的进展，从对抗

视角综述现有的情报欺骗攻击方法和情报对抗防御方法，提出未来智能情报融合方向，旨在为智能态势分析与辅助决策提供支撑。

1 开源情报融合问题特性

1.1 开源情报的发展特点

以任务需求为牵引，开源情报（open source intelligence）具体流程包括^[5]：开源情报收集、开源情报整理、开源情报融合、情报产品整合分发等环节。开源情报能够快速提供背景资料，提高分析质量。美国中央情报局曾指出“高达80%的情报可以从开源信息获取”。然而，互联网用户日益剧增，网络活动复杂多样，人工处理分析难以应对“海量、多源、爆炸式增长”的数据。计算机视觉和自然语言处理^[6]等新技术的出现，为情报自动处理分析提供了新的机遇与方向——目标检测、图像分割、图像标注、语义分割，以及文本分类、摘要提取、机器翻译等关键技术大幅提升了海量情报的处理效率。然而，深度伪造（deepfake）技术打开了人工智能的“潘多拉魔盒”——伪造文本、伪造音频、伪造图像、伪造视频等数据越来越逼真，给智能情报融合带来了新的挑战。此外，基于大规模预训练模型构建的生成式AI^[7]（如ChatGPT），在生成高质量新内容的同时，也存在出现信息偏差的风险——生产出虚假信息 and 错误信息，甚至被恶意用于制造假新闻或冒充个人。数据层面的欺骗与反欺骗技术在对抗中不断发展，成为当前开源情报的鲜明特点。

1.2 开源情报问题中的关键挑战

1.2.1 多源多模态情报信息难关联

情报融合类似拼图游戏，任一模态的描述不一定全面，多种模态数据互补有助于拼接出全方位立体的情报“地图”。互联网情报呈现出多模态特征，以社交平台数据为典型代表，文本、图像、音频、视频、矢量、混合数据等多种模态相互关联、错综复杂。此外，情报信息的获取手段多样，来源也十分广泛，不同来源的传感器获得的情报信息往往是零散的，碎片化的，甚至是残缺的，无法提供完整的情报“地图”。并且，不同源、不同类型的数据往往结构差异较大、粗精不一、语义复杂。发展先进的情报信息融合技术不仅需要多源多模态的数据支持，还需要将这些数据进行有效融合，统一格式，对原始信息进行“去粗取精、由此及彼、由

表及里”的判断。只有通过数据之间的相互支持、补充和校对,才能提供更加准确、全面和及时的信息支撑情报感知工作^[8]。

1.2.2 欺骗攻击情报信息难检测

情报欺骗自古就是军事对抗中的重要手段。信息化条件下,情报欺骗性质没变,但模式已经发生了很大的变化。近年来,基于生成式深度学习(generative deep learning)技术的深度伪造开始崭露头角,使虚假情报的编辑、修改、合成变得相当简便,甚至无技术门槛限制,任何人都可以利用深度伪造多媒体AI工具制造虚假信息。深度伪造能够操控图像和声音,将虚假视频与假音频结合起来,让有影响力的公众人物说出预先设定的虚假内容,制造谎言和混乱。基于大规模开源数据集,在生成式对抗网络、图像转换网络、解码编码网络等现有算法支撑下,深度伪造生成的虚假内容已经精细到对表情、嘴型、眼神、人脸朝向等细节的操控,足以混淆视听、以假乱真。虚实混合的信息环境,为情报融合带来了巨大的挑战。情报对抗过程中,对抗双方都采取各种欺骗措施(如释放大量虚假信息、生产相互矛盾的情报信息)进行数据攻击。如何检测对手的欺骗攻击情报,成为智能情报融合技术发展的关键问题。在信息泛滥的环境中剔除“信息淤泥”,甄别“信息矛盾”,获得真实可靠,与决策行动强相关的高价值、关键性的信息,生成面向用户任务需求的为指挥决策服务的情报产品,是指挥控制领域需要突破的重大技术方向。

2 开源情报融合方法

2.1 基于知识图谱的开源情报融合

知识图谱^[9]通过结合基于规则的模式匹配、机器学习、深度学习^[10]等技术,规范化语义融合多源数据,从结构化、半结构化、非结构化开源信息抽取出实体和实体之间的链接关系,并能通过图谱推理能力支持复杂关联数据的挖掘分析,为海量、异构、碎片化的开源情报的表达、组织、管理以及利用提供了一种更为有效的方式。李玥琪^[11]对社交网络舆情多平台进行了要素拆解,通过对多平台中用户、信息、时空等要素主题图谱的构建与分析,构建了社交网络舆情多平台风险识别模型。Zhang等^[12]基于推特(Twitter)平台上新型冠状病毒感染相关推文,构建主题知识图谱,对全球新型冠状病毒感染疫情进行舆情分析,通过提取话题的多维属

性(如“主语-谓语-宾语”三元组、话题热度、话题情感强度等),探索网络舆情话题之间的语义关系。文献[13]将超图概念引入知识图谱,提出分层海战场情报态势知识超图体系,建立事理知识图谱、概念知识超图、实例知识超图三层结构,使用超关系对多元情报信息进行知识表达。现有基于知识图谱的开源情报融合研究集中在以单模态文本数据为数据源的知识图谱构建。仅以文本、图像等单一模态信息为知识来源的知识图谱难以全面客观地反映现实世界的真实状态。多模态知识图谱^[14]融合文本、图片、视频、音频等不同模态通道,实现多模态实体消歧,利用模态间的知识互补,提升情报融合模型的感知和理解能力,大幅度提高了情报产品质量。例如,语音识别任务中,将人的唇部运动图片与发音信息关联,有助于提高语音识别正确率^[15]。

本文从数据级融合、信息级融合、知识级融合、情报级融合等不同层级梳理了基于多模态知识图谱的情报融合框架,如图1所示。信息融合层面,图像、视频和音频数据分别转换为文本模态,和文本信息一起进行三元组信息提取。知识融合层面,图像中的对象和关系通过构建场景图进行描述,场景图中图像实体与文本抽取的命名实体跨模态对齐,实现不同模态之间的知识交互和共享。情报融合层面,不同模态知识整合后,为下游应用提供全面的支撑。

2.2 基于复杂网络的开源情报融合

目前,复杂网络^[16]已经成为社交网络舆情分析的重要手段之一,常用于社交网络中的群体偏好、意见领袖(opinion leader)、群体意见及社群结构等研究。胡艳丽^[17]针对某高校大型BBS论坛建立了基于兴趣的在线社会网络,并基于复杂网络理论分析了BBS兴趣网络的拓扑结构。黄宝成^[18]针对新浪微博(评论数量和维数大规模的)复杂主题进行意见领袖挖掘研究,基于文本相似性与情感因素,构建复杂主题文本网络,根据受众积分排序确定意见领袖。艾川^[19]基于共享单车出行数据和微信网页分享传播数据,分析时空规律特性,提出了基于无标度特性和距离因素的大规模社会网络建模方法。文献[20]基于超网络理论,对微博舆情传播中的各子网进行建模,挖掘微博舆情中的关键节点。党中恺^[21]构建了多种类型参与者的机构内外谣言传播模型,求解模型的无谣言状态,分析平衡

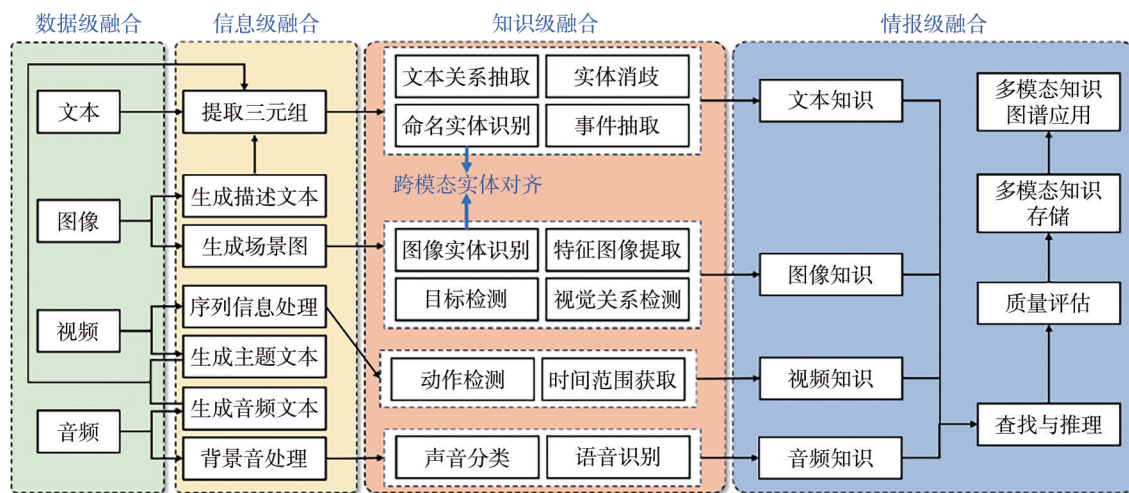


图1 基于多模态知识图谱的情报融合框架

点处的局部稳定性和全局稳定性。

2.3 基于深度学习的开源情报融合

在开源情报分析中，由于社交媒体的核心功能是让人们自由地表达观点和意见，情感分析（sentiment analysis）自然成为智能情报融合研究的核心问题之一。深度学习^[22]已经成为情感分析的主流方法^[23-25]，一些具有代表性的神经网络，如卷积神经网络（convolutional neural network, CNN）、循环神经网络（recursive neural network, RNN）、长短时记忆神经网络等，结合注意力机制，应用于情感分析，能够高效表征低级特征或准确学习高级抽象特征。

2.3.1 文本情感分析

现有的情感分析研究与应用多是针对文本数据。根据所处理文本的颗粒度，文本情感分析研究可以划分为文档级、句子级和属性级。文献[25]对不同层级下基于深度学习的情感分类方法进行了分析，文献[26]针对方面级文本情感分析方法进行了综述。文本情感分析可以看作一个语义分析问题，相比传统的语义分析，文本情感分析聚焦于理解与情感有关的语义内容，而不需要理解句子或文档的全部信息。然而，文本情感分析需要解决众多与自然语言处理相关的子任务，包括实体识别、指代消歧、词极性消歧、主观性检测、人格识别和方面提取等。

随着深度学习在各领域的成功应用，深度学习应用于本情感分析方法也逐渐成为热点研究。相比于基于情感词典和基于传统机器学习的情感分析方法，基于深度学习的情感分析方法可以学习上下文

顺序信息，主动学习本文特征，提取相关词语的语义信息。Li等^[27]提出了一种基于CNN的中文微博观点摘要方法，应用CNN自动学习识别微博极性（polarity）的重要特征，然后对中文微博系统中的热点事件进行句子级情感分析。Da'U^[28]从文本评论中提取与产品相关的情感词汇，结合共同注意力机制聚焦与用户和商品相关的重要信息，设计LSTM编码器捕获文本中词汇的上下文依赖关系，实现用户对商品的情感感知的表示学习。还有更多的研究^[29-31]将注意力神经网络和预训练模型引入文本情感分析。然而，作为一种数据驱动的学习范式，深度学习模型训练依赖于大量高质量数据。

2.3.2 视觉情感分析

随着计算机视觉的兴起，视觉情感分析也逐渐成为热点问题。视觉情感分析可以看作一个图像分类问题，但与一般的图像识别任务相比，视觉情感分析问题更具有挑战性，涉及认知过程中更高层次的抽象性与主观性，属于目标识别、场景识别、动作识别和事件识别的上层任务。Song等^[32]提出一种基于图像显著性的情感分类方法，用两个深度网络分别估计视觉关注度和提取局部特征，并将视觉关注度与区域特征融合，强调不同图像区域的重要性，提升特征表征能力，最后融合成一个端到端的网络提升情感分类表现。文献[33]将注意力网络用于刑事侦查中的微表情识别（micro-expression detection），以揭示犯罪嫌疑人试图隐藏的真实情绪。其次，在视觉情感分析任务中，根据心理学中的情感状态模型^[34]，可以将情感分为十几种类别，对图像情绪的手工标注容易带入人的主观性。一些采

用众包形式标记的数据集,需要考虑手工标定的噪声处理问题。You等^[35]改进监督学习机制,通过提高学习框架的“容忍性”来减少数据集中噪声带来的影响,并设计了一种面向视觉情感分析的渐进式卷积神经网络框架,根据预测得分反向挑选训练数据,使用更“干净”的数据微调模型参数,依次渐进训练,逐步降低噪声实例对训练的影响。

2.3.3 语音和视频情感分析

语音情绪识别应用在社交媒体分析中也较为常见,其通常做法有两种思路。一种是间接转换,即将语音转换为对应的文本,然后利用文本情感分析方法处理;另一种是直接处理,即利用语音特征提取模型(如Wav2vec),将原始语音转换为向量特征,然后设计神经网络,基于标注数据集训练网络模型,实现语音情感分类。视频情感识别做法类似于图像情感识别,通过对视频中的每帧图像进行情绪识别,最后加权求和得到视频的整体情感分析结果^[36]。

2.3.4 语音和视频情感分析

在社交平台中,结合文字、图像、音频、视频表达对事件的情绪成为一种普通的社交活动。如何融合不同模态特征,更准确地实现用户群体的情感分析,逐渐成为研究的新方向。目前,由于“文本-图像-语音-视频”4个模态的语义空间对齐和联合表征方面存在的挑战,现有技术还难以实现四模态数据的融合。此外,预测过程中,容易出现某种模态缺失的现象,这也为3种以上模态的融合提出挑战。

针对以上挑战,现有研究也有一些探索性的工作。例如,Zhang等^[37]提出多模态学习统一框架Meta-Transformer,使用同一组参数同时对来自12种模态的数据(包括图像、自然语言、点云、音频频谱图、视频、红外、高光谱、X射线、惯性测量单元等)进行编码,为通用模态情报融合技术发展提供新方向。Tang等^[38]针对预测时模态的缺失问题,构建双向转换融合网络,确保输入中某一模态缺失场景下多模态情感分析的鲁棒性。基于多模态深度学习的的情感分析研究大部分使用文本-图像双模态数据^[39]。基于图文融合的多模态情感分析分为特征层融合、决策层融合和线性回归融合。蔡国永等^[40]聚焦特征层融合,将不同层次的文本特征与图像基于CNN进行特征融合,得到词语级、短语级和句子级的图文特征,最后基于集成学习进

行决策融合,将不同层级的图文特征融合以获得准确的微博图文情感极性。You等^[41]提出了一种跨模态一致回归(cross-modality consistent regression)模型,在不同模态之间施加一致约束,挖掘多个模态之间的相关性。Majumder等^[42]提出一种分层融合框架,在使用文本、视觉和语音特征双向融合的基础上再进行三向融合,消除不同模态携带的相互矛盾的信息。此外,多模态情感分析的研究还包含虚假新闻检测和反讽识别等子任务。讽刺是一种高级的语言艺术,单从文本情感分析的角度应对反讽识别,需要考虑用户表达习惯、语境等因素,而引入图片后,可以为讽刺检测提供更多线索,提升讽刺检测性能。

2.4 基于大模型的开源情报融合

深度学习技术已日趋成熟,大型语言模型的出现对于深度学习而言是“如虎添翼”^[43]。2022年,OpenAI公司推出聊天机器人ChatGPT,它在文案写作、文字转图片、视频智能配音、智能海报生成、视频智能特效、语音人机交互等多种任务上展现出强大的数智能力,引起社会各界广泛关注。此后,生成式AI(generative AI)急速发展,OpenAI公司的GTP-3、GTP-4系列,中科院自动化所的“紫东太初”,百度的“文心一言”等大模型应用落地,推动人工智能发展进入人类内容生成和人工智能内容生成的并行阶段^[44]。以ChatGPT为代表的大语言模型(large language model)成为全球普遍关注的战略前沿技术,给开源情报工作带来了前所未有的变革,让开源情报工作更具想象力和创造力。随后,大模型发展迅速,推动形成LLaVA^[45]、LVM^[46]等视觉大模型,以及LLaMA-Adapter V2^[47]、VisualGLM-6B10、mPLUG-Owl11、VisualGLM-6B、InstructBLIP^[48]等多模态大模型。大模型改变了原本“多源收集、多步研判”的复杂情报产出方式,向具备自然语言处理、视觉理解、语音分析处理、内容生成能力的智能问答式情报产出方向发展,充分发挥海量开源数据的优势,提高情报工作的便捷性,并且通过多模态数据加强情报对象全息塑造,大幅度提升开源情报信息智能融合分析能力。

大模型有望成为智能情报的基础设施,如图2所示,大模型基座赋能多模态情报处理,有效表征并对齐多模态,通过问答方式生成情报摘要,输出情报分类结果,并对意图和行为进行预测分析。

Hua等^[49]把大模型用作一种国际复杂形式分析和预测工具，提出了一种基于大模型的多智能体世界大战仿真交互系统，用于模拟历史事件和分析复杂的决策行为。华为公司提出盘古气象大模型^[50]，实现了对全球气象（包括位势、湿度、风速、温度、海平面气压等数值）的快速、精准预报。美国SchaleAI公司推出基于大模型的“多诺万（Donovan）”军事决策辅助系统，应用于美国陆军第18空降师的加密网络，实现了开源情报与战场隐秘情报的结合。Palantir公司推出了基于大模型的人工智能平台AIP，用于提供情报分析和辅助决策服务。

然而，当前多模态大模型还并不能替代情报分析员。多模态大模型在实际应用过程中依旧存在“模型训练数据集的更新迟延、模型参数的更新、模型的认知偏差、机器幻觉”等局限^[51]，以及“模型可解释性、模型反事实推理能力^[52]、模型抗篡改能力”等开放问题。多模态大模型仍旧只是一个高效的情报分析工具，人在回路的情报模式还将长期存在，将人类纳入多模态大模型训练循环来开发更加安全的情报分析大模型，以及加速多模态大模型与其他开源情报系统形成铰链，在推动智能情报前向发展中都具有重要意义。

3 开源情报对抗

3.1 情报欺骗攻击

3.1.1 对抗样本攻击

智能情报融合中，海量数据资源的收集与分析处理强烈依赖机器学习。对抗样本攻击通过对学习训练样本蓄意添加人类感官难以识别的细微扰动，使机器学习模型以高置信度给出一个错误的分类输出。如图3所示，通过在原有图片添加肉眼不可见的对抗样本，可以以更高的置信度得到与原样本完全不同的结果，将“熊猫”识别为“长臂猿”。在极端条件下，对抗样本甚至可以微小到只改变单个像素，就能成功干扰大量图片的识别结果。例如：针对Kaggle CIFAR-10自然图像测试数据集，单像素攻击^[53]通过修改单个像素就可以扰乱数据集中67.97%的图像的判别结果。

从攻击方式上来划分，对抗样本的攻击方式可分为白盒攻击、灰盒攻击和黑盒攻击。白盒攻击是指攻击者能知道被攻击的神经网络算法及模型参数（如神经网络的架构、超参数），然后攻击者根据已知的神经网络信息设计具有针对性的对抗样本，使神经网络失效。典型的白盒攻击方法包括快速

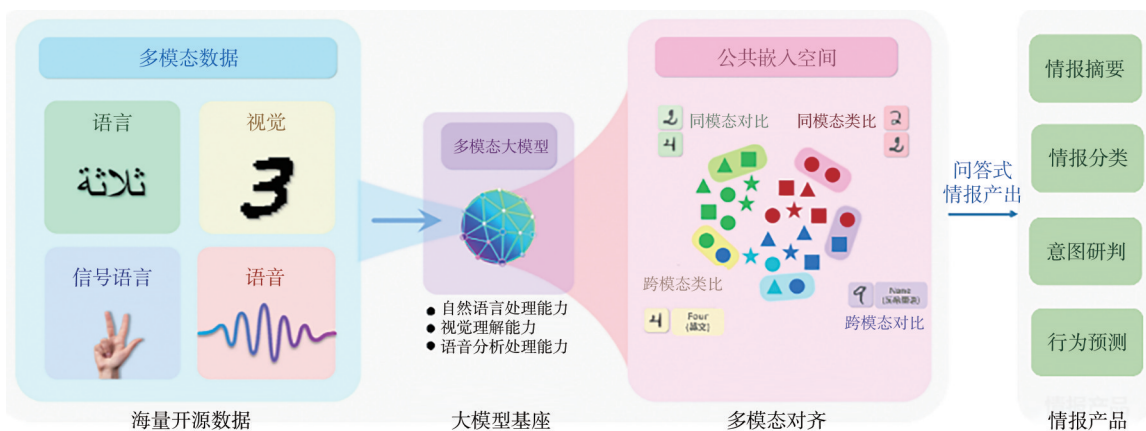


图2 多模态情报处理大模型

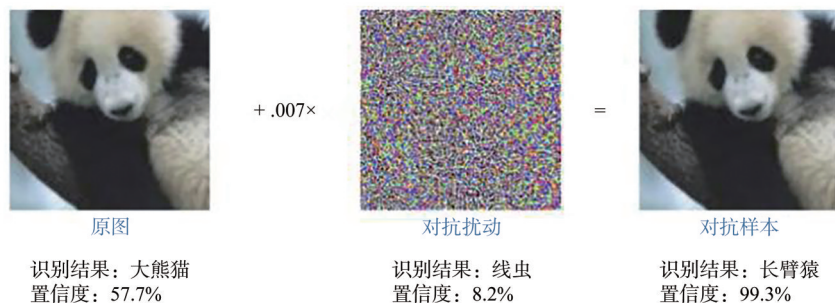


图3 对抗攻击样本示例^[54]

梯度攻击^[55] (fast gradient sign method, FGSM)、基于FGSM的迭代攻击^[56]、迭代最小可能性类攻击 (ILLCM)^[56]、深度欺骗 (Deepfool) 攻击^[57]、雅克比映射攻击^[58]等。灰盒攻击是使用了目标模型的任意信息。黑盒攻击^[59-60]是指攻击者对攻击的模型的内部结构、训练参数等都未知, 仅能通过模型的输入输出与模型交互。黑盒攻击是比白盒攻击更为接近真实环境的一种攻击方式, 但是攻击方法难度更大, 攻击成功率低。从攻击输出的特点看, 对抗样本攻击可以分为定向攻击和非定向攻击。定向攻击能够使模型输出指定的误判类别。从扰动的适用范围看, FGSM、ILLCM、Deepfool等方法都是对一个图片产生特定的扰动, 从而欺骗这个网络, 而通用对抗扰动^[61]可以作用到大批图片上, 从而实现对深度神经网络的欺骗。

除了视觉欺骗攻击, 对抗样本攻击也可以用于欺骗智能语音识别系统。Chen等^[62]针对说话者识别系统, 提出了一种黑盒攻击方法Fakebob, 它不仅能够成功欺骗开源商业语音识别系统, 并且在物理世界中播放, 人耳也很难区分原始语音和对抗攻击语音。Cisse等^[63]提出一种灵活性对抗框架Houdini, 它不仅可以生成对抗音频, 在黑箱攻击场景中成功欺骗谷歌语音 (Google Voice), 还能对视觉人体姿态估计模型^[64]进行定向和非定向攻击。

3.1.2 对抗补丁攻击

类似于对抗样本攻击, 对抗补丁 (adversarial patch) 攻击^[65]的目标也是寻求可以改变模型分类结果的扰动。然而, 不同的是对抗样本攻击是在数字世界完成对机器学习算法的攻击, 而对抗补丁攻击不局限于数字世界, 可以在物理世界对图像识别算法进行攻击。对抗样本攻击希望尽可能减少扰动程度, 以避免被发现, 而对抗补丁攻击不再追求将扰动限制在人眼难以察觉的范围内, 而是仅针对智能图像识别算法, 生成一个与原图像无关的补丁, 攻击分类器, 让其输出指定的目标类别。该补丁甚至可以出现在图像中的任何位置且以多个角度出镜, 都能起到攻击效果。一张简单打印出来的对抗补丁, 就可以使深度学习模型产生致命的分类错误, 从而将识别目标“隐藏”起来, 甚至可以生成视觉保真度较好且与场景语义相关度较高的对抗伪装^[66-68] (adversarial camouflage), 如图4所示。不仅限于单模态, 对抗补丁也能在多模态场景发挥作用。Wei等^[69]提出了一种统一对抗性补丁生成方法

执行跨模态物理域欺骗攻击, 通过单个补丁同时欺骗可见光和红外目标探测器的目标检测识别。并且, 带有对抗补丁贴片的拍摄物体在不同角度、距离、姿势和场景等不同设置中, 都能实现接近70%的攻击成功率。



图4 对抗伪装示例^[70]

3.1.3 生成式攻击

生成式攻击使用深度神经网络生成高可信度的媒体, 最早出现在视觉领域, 能够以假乱真, 操控舆论。2017年12月, 国外社交网站上一名为“DeepFake”的用户发布了一段利用名人面孔合成的色情视频, 引发各界关注。由于其娱乐性和社交媒体的助力, 深度伪造开始在全世界快速传播。起初, 深度伪造产品还可以通过图像中的伪影或者视频中人物五官、眨眼频率等特征分辨出来, 随着技术不断演进, 无论是实现“制伪”的计算能力还是逼真程度都有了很大提升。深度伪造可以合成、修改图片或视频中的人脸、表情、肢体动作等, 其伪造程度细致周密。以人脸深度伪造为例, 如图5所示, 深度伪造技术可以重定制人脸的表情、嘴型、眼神、姿态, 甚至编辑人的发型、肤色、年纪、美丑。

现有的深度伪造主流方法都基于生成式对抗网络 (generative neural networks, GAN)^[71]以及其变体 (如pix2pix^[72]、CycleGAN^[73]) 实现。在GAN训练过程中, 两个神经网络 (生成网络和鉴别网络) 相互博弈。生成网络负责生成伪造数据, 复制原始数据集的属性。鉴别网络的任务是识别伪造数据。根据每次迭代的结果, 生成器网络不断调整参数, 以生成更加逼真的数据, 欺骗鉴别网络。而鉴别网络的目标是不断提高鉴别能力。两个网络会持续对抗, 进行数千或数百万次迭代, 最终生成网络的性能提高到鉴别网络无法再区分真实数据和伪造数据。除GAN之外, 深度伪造模型在图像合成上也取得了巨大的进展, 如变分自编码器^[74]、基于流的模型 (flow-based models)^[75]、基于能量的模型、扩散模型 (diffusion model)^[76]等。最近, 扩散模型在图像合成上的表现惊人, 性能已经超越

图5 不同类型的人脸深度伪造示例^[54]

GAN, 成为生成模型的一个热点研究方向。目前, 深度伪造已经发展为包括视频伪造、声音伪造、文本伪造和微表情合成等多模态视频欺骗技术, 由最初的操纵单模态媒体 (包括音频、图像、视频和文本) 向操纵多模态媒体转变, 呈现出“算法自动生成、制作成本和门槛较低”“产品逼真而多元, 识别难度大”和“与社交媒体结合、传播速度快”等特点。以扩散模型、大模型等为代表的生成模型, 能够有效实现跨模态伪造攻击。例如, 可以根据输入的文本生成与之描述相匹配的图像、音频、视频^[77-80]。人工智能生成内容 (AI generated content, AIGC) 正在以超出人们预期的速度成为科技革命历史上的重大事件。典型情报欺骗攻击方法见表1。

3.2 情报对抗防御

情报欺骗攻击制造虚假信息, 已经成为误导舆论的武器, 不断侵蚀社会信任, 甚至导致政治危机^[84]。美军高度关注情报欺骗攻击手段的发展, 不断探索应对之策。2018年, 美国海军发布项目“利用对抗样本欺骗计算机视觉分类器”, 旨在深入研究对抗样本欺骗的本质, 抵御敌人对己方分类器的伪装和误导。2019年, DARPA发布项目“确保人工智能抵御欺骗的鲁棒性”, 目标是开发针对对抗性欺骗攻击的防御。2019年, DARPA发起了语义取证 (semantic forensics, SemaFor) 项目, 旨在对文本、图像和视频等多种类型的伪造数据进行检测, 抵御虚假情报数据的攻击。2020年, 美国

表1 典型情报欺骗攻击方法

模型	核心方法	源模态	目标模态	欺骗对象	伪造类型		
					对抗样本	对抗伪装	深度造假
DeepFool ^{[57][81]}	基于梯度迭代方法	图像	图像	GoogleNet、CaffeNet	√		
Houdini ^{[63][82]}	Houdini 损失	语音 图像	语音 图像	Google Voice、姿态估计模型、语义分割模型	√		
UAP ^{[61][83]}	通用扰动	图像	图像	GooLeNet、CaffeNet VGG-19、ResNet-152	√		
ILLCM ^[56]	ILLCM	图像	图像	ImageNet Inception	√		
OnePixel ^[53]	差分进化	图像	图像	DNN	√		
PS-GAN ^[66]	PS-GAN	图像	图像	VGG16、ResNet		√	
FakeBob ^[62]	分数阈值估计、基本迭代法、基于演化的梯度估计	语音	语音	说话人识别系统	√		
AdvCam ^[68]	对抗伪装损失	图像	图像	VGG-19		√	
Grad-tts ^[77]	扩散模型 单调对齐搜索	文本	语音	(合成高质量语音)			√
Imagen ^[78]	扩散模型	文本	图像	(生成高保真图像)			√
DOPatch ^[67]	DOP-LO、DOP-DTA DOP-DMAT	图像	图像	FaceNet、SphereFace、 MobileFaceNet、ArcFace		√	
UniCroM ^[69]	差分进化 边界限制形状优化	可见光图像、 红外图像	可见光图像、 红外图像	YOLOv3、Faster RNN		√	
Dreamix ^[79]	级联视频扩散模型 混合视频图像微调	图像 视频	视频	(生成高保真高分辨率视频)			√

国防部发布《数据战略》，重点强调了数据可信在决策中的重要性。为了减少深度伪造的社会危害，学术界和工业界也在不断探索检测反制方法。脸书(Facebook)、微软和麻省理工联合斥资1 000多万美元举办“deepfake检测挑战赛”，并发布大规模deepfake视频识别数据集和合成语音数据集^[85]，旨在提供开源平台和数据集推动深度伪造反制技术发展。商汤科技发布了具有更多人种和合成效果的深度伪造数据集DeeperForensics-1.0^[86]。典型情报对抗防御方法见表2。

3.2.1 检测识别性防御

(1) 伪造图像/视频检测

深度伪造难以完美再现实面部所有细节，一种直接的检测识别方法是在脸上寻找视觉伪影。例如，人眼颜色、耳朵特征等左右不对称，眼睛中的镜面反射表现为白斑，甚至缺失等。还有一些方法依赖于检测面部扭曲伪影^[88]、面部标识(如眼角、嘴角、鼻尖)位置^[89]和头部姿势^[102]等人工特征。此外，也有研究从颜色统计特征分析深度伪造图像与真实图像的差异。例如，深度伪造图像在RGB

颜色空间中没有颜色相关性约束，但在HSV和YCbCr等其他颜色空间中伪造图像与真实图像有较为明显的差异^[90,103]。基于人工特征的检测识别方法具有较好的可解释性^[104]，但精心设计的底层特征缺乏对数据的深入分析理解。结合深度学习的检测识别方法能够自动提取额外的语境特征，配合自注意力机制，聚焦关键伪造特征。基于深度学习的伪造检测方法进一步可分为^[104]：基于对比的检测方法^[105]、基于注意力的检测方法^[106]、基于局部相关性的检测方法^[107]和基于频率的检测方法^[108-109]。

(2) 伪造音频检测

音频伪造可以分为文字转语音(text-to-speech)、语音转换、语音情绪造假、语音背景造假、语音片段造假5种类别。伪造音频检测的关键在于提取语音信号特征，构建分类器或学习器判别音频伪影。常用的语音信号特征包括：短时频谱特征、长时频谱特征、韵律特征和深度特征。从近年举办的伪造音频识别比赛(ASV spoof和ADD)结果看^[110]，使用短时频谱特征和原始波长，基于残差神经网络(ResNet)^[111]和LightCNN^[112]及其变体，能够取得

表2 典型情报对抗防御方法

模型	时间	核心方法	测试数据集	检测效果	检测范围			
					伪造图像	伪造视频	伪造音频	谣言
att-RNN ^[87]	2017	注意力机制RNN	Twitter数据集	ACC:68.2%				√
DFWA ^[88]	2019	VGG16、ResNet50、ResNet101、ResNet152	DeepfakeTIMIT UADFV	AUROC:84.6~99.9 AUROC:84.5~97.4	√	√		
IH ^[89]	2019	面部标记特征、SVM	CelebA、PGGAN	AUROC:91.2~94.1	√			
DEDRC ^[90]	2019	高通滤波、色度分量、CNN、随机森	CelebA、PGGAN	ACC:98.77%~100%	√			
OGAN ^[91]	2020	OGAN	YouTube数据集	-	√			
Deep4SNet ^[92]	2021	CNN、图像增强	-	ACC:99.7%			√	
COSMOS ^[93]	2021	自监督学习、图文匹配模型	COSMOS	ACC(内容):85% ACC(匹配):72%	√			√
FakeTagger ^[94]	2021	图像标签	CelebA-HQ	ACC:>95%	√			
Dual-BF ^[95]	2021	双频段混合	ASVspoof 2019	EER:0.035			√	
CVT ^[96]	2021	CNN、ViT	DFDC	ACC:91.5%		√		
GADE ^[97]	2022	共同注意力、图注意力	ESD、EmoFake、ASVspoof 2019	ACC:97.3%~98.8% ACC:87.1%~96.1% EER:1.65			√	
EANN ^[98]	2022	Text-CNN VGG19	Twitter数据集 Weibo数据集	ACC:71.5% ACC:82.7%	√			√
CCN ^[99]	2022	多模态一致性检查	NewsCLIPpings	ACC:80.3%~84.7%	√			√
HAMMER ^[100]	2023	HAMMER	DGM4	AUROC:94.4 ACC:86.8% EER:13.18	√			√
DE-FAKE ^[101]	2023	CLIP、BLIP	MSCOCO Flickr30k	ACC:73.6% ACC:~95%	√			

较低的等误差率 (equal error rate, EER)。伪造音频检测分类大多基于深度学习方法, 其性能显著优于基于SVM和GMM^[113]的分类器。深度学习模型架构一般基于CNN、ResNet^[114]、Res2Net^[83]、压缩激励网络 (SENet)^[115]、图神经网络 (GNN)、可微架构搜索^[82]、Transformer及其对应变体。

(3) 谣言检测

谣言检测是虚假信息检测的一个重要内容。虽然谣言并不是互联网的产物, 但网络社交平台为谣言的传播“推波助澜”, 加速了其负面影响累积。谣言的收集分析通常基于一些开源平台, 如Twitter、新浪微博、FacebookAPI等^[116]。根据谣言的时间特性, 收集策略可以分为^[117]自上而下和自下而上两种。自上而下的策略^[118-119]适用于长期存在有争议性的话题 (如疫苗), 需要对谣言具有先验知识, 基于关键词和标签来收集谣言。相反, 自下而上的策略^[120-121]的核心思想是在特定时间窗口内收集尽可能多的社交媒体帖子, 专门用于收集突发新闻中的潜在谣言。谣言检测识别主要依赖于内容特征 (content feature) 和上下文特征^[122]。内容特征指的是可以直接从文本中提取的信息, 基于文本分析工具就能实现从文本 (词元、词性、词目、句子等) 中提取基本知识, 并获得结构化表示, 以支撑进一步的分析。基于上下文特征是指真实社交媒体帖子或假新闻周围的相关信息, 涉及对用户的分析、谣言或新闻的来源、社交媒体上信息的传播结构以及其他用户对新闻的反应。

(4) 多模态融合检测

社交平台中, 多模态媒体内容信息立体丰富, 不同模态从不同维度提供画像。多模态融合检测利用不同模态的相互印证可以实现对虚假新闻的检测。针对小规模的人工智能生成造假内容的研究相对较多。Wang等^[123]针对新出现事件提出基于事件对抗神经网络的多模态假新闻检测方法。Jin等^[87]使用带有注意力机制的RNN, 融合来自文本、图像和社交上下文的特征, 以完成谣言检测任务。值得注意的是, 具有更多的真实细节的虚假新闻往往更具有迷惑性。例如, 把真实图像脱离上下文, 用于支撑虚假事件进行误导性描述。一些围绕脱离上下文 (out-of-context) 多模态造假新闻的检测研究相继出现^[93,99]。其中, Shao等^[100]提出一种分层多模态伪造推理Transformer框架 (HAMMER), 它不仅能够实现深度伪造内容检测, 还能够定位伪造

的图像区域和篡改的单词。

3.2.2 扰乱阻断性防御

相比于在深度造假后检测识别, 深度造假扰乱阻断 (deepfake disruption and evasion) 注重对深度造假的预防。扰乱阻断性防御类似一种对深度伪造模型的数据投毒攻击, 其核心思想是对原图片加入对抗性扰动, 生成带有扰乱阻断防御能力的新图片, 在遭遇深度伪造攻击后新图片明显失真, 从源头上阻断造假信息的传播。Ruiz等^[123]向源视频帧注入微小的扰动以破坏人脸交换模型的自编码器输出。其核心思想是基于GAN框架, 迭代训练具有阻断能力的防御生成器和换脸模型, 实现防御生成器生成的图像失真最小化和防伪能力最大化, 能够明显辨别出是合成图片而非原图。然而, 由于假设深度伪造模型是静态的, 这种防御方法只能干扰阻断模型参数固化的伪造工具。现实场景下, 攻击者很容易用新收集的数据 (或者将带有扰乱阻断性防御的图片加入训练集) 训练一个新的深度造假模型来逃避这种防御。为解决这一问题, Segalis等^[91]通过时空扭曲攻击注入扰动, 连续帧图像在应用深度伪造算法后脸部发生明显变化, 以此实现针对动态伪造模型的防御。Yang等^[124]提出一种新的变换感知对抗性扰动来保护原始人脸图像, 即使深度造假模型使用受保护的人脸图像训练也无法生成高质量的伪造人脸, 合成人脸伴有明显的视觉伪影。

4 智能情报融合展望

4.1 情报融合模式变革

4.1.1 人机交互的情报融合

智能情报融合只是一种工具, 其本身并不具备自主意识、指挥艺术、自主创新能力等, 对一些隐含发现型、思维创新型情报处理问题仍需“人在回路”进行补充完善。未来很长一段时间内, 情报融合模式仍将保持“智能+人工”的模式不变, 形成“智能识别+人工干预”“智能融合+人在旁路”“智能分析+人工研判”等多种人机交互情报融合模式。探索高效人机协作机制, 增强人机交互情报融合的可靠性、可信性, 将是人机交互情报融合的难点。

4.1.2 任务驱动的主动情报感知

主动情报感知是利用情报方法实现情报任务的主动观察和主动发现的过程^[125]。情报需求往往由情报用户直接提出, 情报任务主体缺乏洞见能力和

感知能力。在复杂环境下,要求情报任务主体主动感知潜在的情报用户需求,打破信息孤岛,消解信息迷雾,实现面向全任务流的主动监测、主动发现。

4.1.3 集群分布式情报融合

集群分布式情报融合的研究已形成初步转化应用。美国海军构建无人潜航器(UUV)“分布式侦察与探测系统”^[126],协调水下异构无人潜航器,用于执行海底搜索、通信导航、水下救援等任务。欧盟委员会资助的多国合作项目“自主水下航行器协同感知控制项目”^[127],旨在开发用于多UUV协同控制的高级认知系统,聚焦无人集群在水下通信所面临的挑战,着重解决3D感知与绘图、协同态势感知、导航控制与行为控制等问题。然而,在数据链带宽有限、传输不稳定的情况下,集群分布式情报融合仍旧面临“不确定信息条件下的多情报节点智能融合”“基于异构情报信息的目标联合识别”等难题。针对战场数据海量、控制命令传输受限以及多源传感器协同瞄准等难题,如何构建分组分布式情报处理架构,有效降低数据传输带宽需求,提升各个传感器或情报应用节点情报融合和共享能力,将成为进一步提升战场态势应用效能亟待解决的问题。

4.2 情报融合技术加强

4.2.1 多源多模态时空情报融合

在时间和空间数据的辅助下,多源多模态情报的融合可以充分利用时空交互,将时序数据、地理空间信息融入整个数据分析和感知决策的过程。时序数据串联了情报感知中各个事件或要素的发展演化,一些看似离散叙述的事件,背后往往存在着一定的联系。地理空间数据既可以表示目标在一定时间范围内的位置变化,也可以从细粒度、多层面和多视角记录目标的变化状态。基于位置服务的应用平台能够产生海量全球定位系统(GPS)数据。例如, Twitter每天约产生1 000万条带有地理信息的数据,美国Foursquare中更是存储了超100亿条的位置签到信息,这些地理空间情报包含大量的人类活动、交通网络、通信网络,甚至自然特征(如地形、植被和天气)等信息。如何有效融合时空信息,集成不同格式数据、解决数据过载难题、建立数据安全和隐私协议将成为情报融合技术发展的挑战。

4.2.2 安全可信情报融合

深度造假攻击和防御相互适应、反复博弈,如何提高情报融合的安全性和可信度,成为提高情报产品质量的关键问题。情报融合安全性方面,研究人员开始探索提高情报融合安全性的新技术。例如,基于区块链的虚假新闻检测技术^[128]、基于强化学习的深度造假对抗防御^[129]、基于群体智慧的虚假情报检测^[130]。一些研究对情报可信分析机理进行了探索,例如:文献[131]面向深度伪造的情报可信分析机理研究,从信息源、信息模态与信息内容3个维度分别提取出情报信息的元特征、模态特征和语义特征,采用多模态情报关联印证机制对情报可信度进行分析评估。文献[132]以情报本身和情报传播者相关理论为基础,总结定量开源情报分析方法,包括基于知识的可信分析方法、基于风格语义的可信分析方法、基于传播模式的可信分析方法、基于情报源的可信分析方法。此外,一些研究也尝试从可解释性^[133]角度阐明深度造假攻击和防御模型的内部机理,以为情报融合提供有力的可信分析。在信息源头、传播过程和融合机制等不同层面,如何提高情报融合模型的立体防御能力将是未来情报融合的热点问题。

4.2.3 情报物联网云边协同

2020年,美国国际战略研究中心(CSIS)发布报告《收集优势:利用新兴技术建立情报收集优势》,指出配备人工智能和云计算功能的边缘传感器和处理设备等新兴技术可改变开展情报行动的方式,优化情报收集的速度与规模。2021年,CSIS发布报告《在安全竞争时代实现情报、侦察和监视的现代化》,指出美国要加强持续监视全球威胁能力必须加强边缘计算能力。复杂作战环境下,网络通信受限,构建云边协同的情报物联网具有“加速情报处理和响应、降低通信网络负担”的现实意义。计算能力推向网络边缘、情报跨边缘交互、云端资源调度管理、情报可视化管理、数据非实时热备等情报物联网云边协同研究,将是推动情报大数据交换的关键。

4.3 全域联合情报能力生成

深度学习、大模型、边缘计算等技术的飞速发展加速了武器装备革新与战争形态演变,体系对抗取代平台对抗成为决胜战争的关键。在这样的大背景之下,情报融合不单单是在技术层面情报综合处理的结果,现实意义的情报融合还应当包括情报机

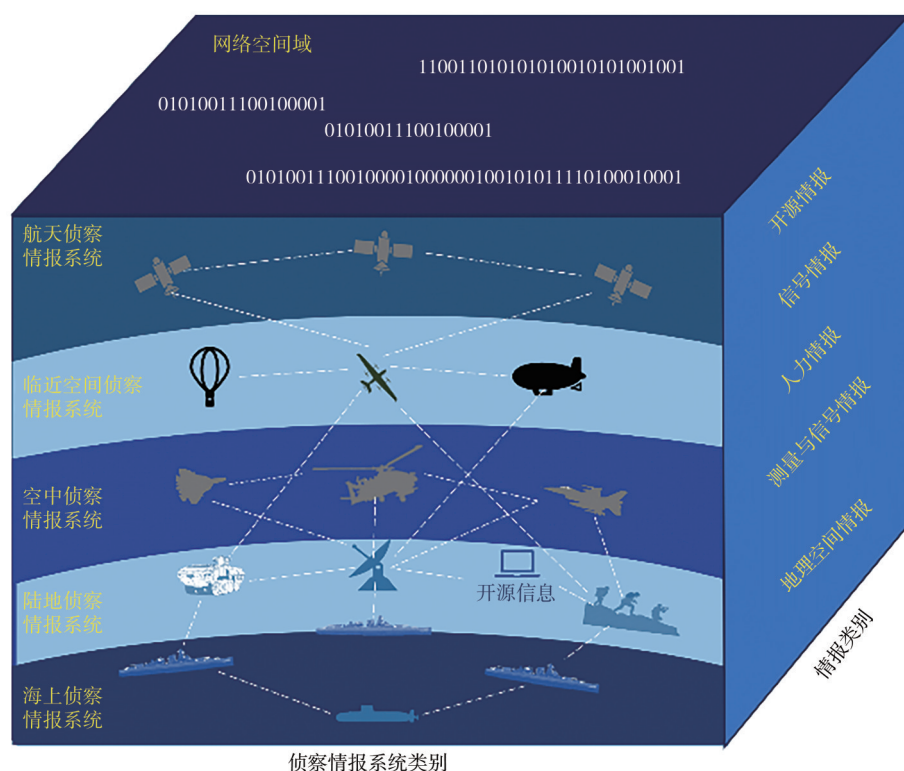


图6 全域联合情报能力

构层面的协作与融合，各情报系统（航天侦察情报系统、临近空间侦察系统、空中侦察情报系统、地面侦察情报系统、海上侦察情报系统）的互联互通，不同类别情报的相互印证，形成全域联合情报能力，如图6所示。未来的智能化对抗中，研究如何“加强地海天联合情报与网络开源情报大数据信息关联印证，推动侦察、指挥、通信、打击、保障等跨网络平台信息融合，加速共享数据”，将有助于在网络战、心理战、舆论战、自媒体战等跨领域跨媒体舆情塑造与认知对抗中赢得先机^[134]。

参考文献：

- [1] 崔琳, 杨黎斌, 何清林, 等. 基于开源信息平台的威胁情报挖掘综述[J]. 信息安全学报, 2022, 7(1): 1-26.
CUI L, YANG L B, HE Q L, et al. Survey of cyber threat intelligence mining based on open source information platform[J]. Journal of Cyber Security, 2022, 7(1): 1-26.
- [2] 赖茂生. 情报检索技术与方法的研究综述[J]. 情报学进展, 2004, 5: 164-202.
LAI M S. A review on information retrieval techniques and methods[J]. Advances in Information Science, 2004, 5: 164-202.
- [3] 刘颖, 王哲, 房杰, 等. 基于图文融合的多模态舆情分析[J]. 计算机科学与探索, 2022, 16(6): 1260-1278.
LIU Y, WANG Z, FANG J, et al. Multi-modal public opinion analysis based on image and text fusion[J]. Journal of Frontiers of Computer Science and Technology, 2022, 16(6): 1260-1278.
- [4] 刘欣, 张占月, 朱丰. 空中进攻作战中天基信息支援指挥[J]. 指挥与控制学报, 2021, 7(3): 319-325.
LIU X, ZHANG Z Y, ZHU F. Space-based information support command for air offensive operation[J]. Journal of Command and Control, 2021, 7(3): 319-325.
- [5] 范昊, 郑小川. 国内外开源情报研究综述[J]. 情报理论与实践, 2021, 44(10): 185-192, 201.
FAN H, ZHENG X C. A review of the research on open source intelligence at home and abroad[J]. Information Studies: Theory & Application, 2021, 44(10): 185-192, 201.
- [6] 桂韬, 奚志恒, 郑锐, 等. 基于深度学习的自然语言处理鲁棒性研究综述[J]. 计算机学报, 2024, 47(1): 90-112.
GUI T, XI Z H, ZHENG R, et al. Recent researches of robustness in natural language processing based on deep neural network[J]. Chinese Journal of Computers, 2024, 47(1): 90-112.
- [7] CAO Y H, LI S Y, LIU Y X, et al. A comprehensive survey of AI-generated content (AIGC): a history of generative AI from GAN to ChatGPT[EB]. arXiv preprint, 2023, arXiv: 2303.04226.
- [8] 白如江, 鞠孜涵, 张玉洁, 等. 面向情报感知的多源多模态数据融合方法研究[J]. 情报杂志, 2023, 42(10): 124-131.
BAI R J, JU Z H, ZHANG Y J, et al. Research on multi-source and multi-mode data fusion for intelligence perception[J]. Journal of Intelligence, 2023, 42(10): 124-131.
- [9] 赵苗乔, 承楠, 陈劼, 等. 基于知识图谱的6G网络场景认知研究[J]. 智能科学与技术学报, 2023(4): 494-504.
ZHAO Z Q, CHENG N, CHEN J, et al. Research on 6G network scenario cognition based on knowledge graph[J]. Chinese Journal of Intelligent Science and Technology, 2023(4): 494-504.
- [10] 王剑, 彭雨琦, 赵宇斐, 等. 基于深度学习的社交网络舆情信息抽取方法综述[J]. 计算机科学, 2022, 49(8): 279-293.

- WANG J, PENG Y Q, ZHAO Y F, et al. Survey of social network public opinion information extraction based on deep learning[J]. Computer Science, 2022, 49(8): 279-293.
- [11] 李玥琪. 社交网络舆情多平台主题图谱构建及风险识别研究[D]. 长春: 吉林大学, 2023.
- LI Y Q. Research on the construction of multi platform theme maps for social network public opinion and risk identification[D]. Changchun: Jilin University, 2023.
- [12] ZHANG Y C, WANG Y F, FAN H, et al. Worldwide COVID-19 topic knowledge graph analysis from social media[C]//Proceedings of the 2023 International Joint Conference on Neural Networks (IJCNN). Piscataway: IEEE Press, 2023: 1-10.
- [13] 苏琦, 史红权, 王义涛, 等. 海战场情报态势知识超图组织与运用[J]. 舰船电子工程, 2023, 43(8): 37-42.
- SU Q, SHI H Q, WANG Y T, et al. Organizing and application of intelligence situation knowledge hyper-graph in the sea battlefield[J]. Ship Electronic Engineering, 2023, 43(8): 37-42.
- [14] ZHU X R, LI Z X, WANG X D, et al. Multi-modal knowledge graph construction and application: a survey[J]. IEEE Transactions on Knowledge and Data Engineering, 2024, 36(2): 715-735.
- [15] 何俊, 张彩庆, 李小珍, 等. 面向深度学习的多模态融合技术研究综述[J]. 计算机工程, 2020, 46(5): 1-11.
- HE J, ZHANG C Q, LI X Z, et al. Survey of research on multimodal fusion technology for deep learning[J]. Computer Engineering, 2020, 46(5): 1-11.
- [16] DA MATA A S. Complex networks: a mini-review[J]. Brazilian Journal of Physics, 2020, 50(5): 658-672.
- [17] 胡艳丽. 在线社会网络中的舆论演化关键技术研究[D]. 长沙: 国防科学技术大学, 2011.
- HU Y L. Research on key technologies of public opinion evolution in online social network[D]. Changsha: National University of Defense Technology, 2011.
- [18] 黄宝成. 基于社交网络的舆情信息挖掘方法研究[D]. 哈尔滨: 哈尔滨工业大学, 2017.
- HUANG B C. Research on public opinion information mining method based on social network[D]. Harbin: Harbin Institute of Technology, 2017.
- [19] 艾川. 大规模复杂社会网络上传播行为建模仿真研究[D]. 长沙: 国防科技大学, 2021.
- AI C. Modeling and simulation of communication behavior on large-scale complex social network[D]. Changsha: National University of Defense Technology, 2021.
- [20] 张连峰, 周红磊, 王丹, 等. 基于超网络理论的微博舆情关键节点挖掘[J]. 情报学报, 2019, 38(12): 1286-1296.
- ZHANG L F, ZHOU H L, WANG D, et al. Key node mining of weibo public opinion based on super network theory[J]. Journal of the China Society for Scientific and Technical Information, 2019, 38(12): 1286-1296.
- [21] 党中恺. 基于复杂网络的信息传播建模与防控研究[D]. 北京: 北京邮电大学, 2022.
- DANG Z K. Research on modeling and prevention of information dissemination based on complex network[D]. Beijing: Beijing University of Posts and Telecommunications, 2022.
- [22] 张钺. 人工智能进入后深度学习时代[J]. 智能科学与技术学报, 2019, 1(1): 4-6.
- ZHANG B. Artificial intelligence is entering the post deep-learning era[J]. Chinese Journal of Intelligent Science and Technology, 2019, 1(1): 4-6.
- [23] GANDHI A, ADHVARYU K, PORIA S, et al. Multimodal sentiment analysis: a systematic review of history, datasets, multimodal fusion methods, applications, challenges and future directions[J]. Information Fusion, 2023, 91: 424-444.
- [24] ETAIWI W, SULEIMAN D, AWAJAN A. Deep learning based techniques for sentiment analysis: a survey[J]. Informatica, 2021, 45(7): 89-95.
- [25] ZHANG L, WANG S, LIU B. Deep learning for sentiment analysis: a survey[J]. WIREs Data Mining and Knowledge Discovery, 2018, 8(4): 1253-1258.
- [26] 王璐, 马宏伟, 吕欢欢. 方面级文本情感分析综述[J]. 计算机应用, 2022, 42(2): 1-9.
- WANG L, MA H W, LYU H H. Summary of aspect-based sentiment analysis[J]. Journal of Computer Applications, 2022, 42(2): 1-9.
- [27] LI Q D, JIN Z P, WANG C, et al. Mining opinion summarizations using convolutional neural networks in Chinese microblogging systems[J]. Knowledge-Based Systems, 2016, 107: 289-300.
- [28] DA'U A, SALIM N. Sentiment-aware deep recommender system with neural attention networks[C]//Proceedings of the IEEE Access. Piscataway: IEEE Press, 2019: 45472-45484.
- [29] MA D H, LI S J, ZHANG X D, et al. Interactive attention networks for aspect-level sentiment classification[C]//Proceedings of the Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence. California: International Joint Conferences on Artificial Intelligence Organization, 2017: 4068-4074.
- [30] WEI J Y, LIAO J, YANG Z F, et al. BiLSTM with multi-polarity orthogonal attention for implicit sentiment analysis[J]. Neurocomputing, 2020, 383: 165-173.
- [31] YANG H, ZENG B Q, YANG J H, et al. A multi-task learning model for Chinese-oriented aspect polarity classification and aspect term extraction[J]. Neurocomputing, 2021, 419: 344-356.
- [32] SONG K K, YAO T, LING Q, et al. Boosting image sentiment analysis with visual attention[J]. Neurocomputing, 2018, 312: 218-228.
- [33] 于洋, 孙芳芳, 吕华, 等. 基于多尺度时空注意力网络的微表情检测方法[J]. 计算机工程, 2023, 50(6): 228-235.
- YU Y, SUN F F, LYU H, et al. Micro-expression detection method based on multi-scale spatiotemporal attention network[J]. Computer Engineering, 2023, 50(6): 228-235.
- [34] PLUTCHIK R. A general psychoevolutionary theory of emotion[M]//Theories of Emotion. Devon: Academic Press, 1980: 3-33.
- [35] YOU Q Z, LUO J B, JIN H L, et al. Robust image sentiment analysis using progressively trained and domain transferred deep networks[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2015, 29(1): 381-388.
- [36] LIU B. Sentiment analysis: mining opinions, sentiments, and emotions[M]. Cambridge: Cambridge university press, 2020.
- [37] ZHANG Y Y, GONG K X, ZHANG K P, et al. Meta-Transformer: a unified framework for multimodal learning[EB]. arXiv preprint, 2023, arXiv: 2307.10802.
- [38] TANG J J, LI K, JIN X Y, et al. CTFN: hierarchical learning for multimodal sentiment analysis using coupled-translation fusion network[C]//Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2021: 5301-5311.
- [39] 彭勇, 彭旋, 郑志军, 等. 多模态大模型: 技术原理与实战[M]. 北京: 电子工业出版社, 2023.

- PENG Y, PENG X, ZHENG Z J, et al. Multimodal large model: technical principles and practical applications[M]. Beijing: Publishing House of Electronics Industry, 2023.
- [40] 蔡国永, 夏彬彬. 基于卷积神经网络的图文融合媒体情感预测[J]. 计算机应用, 2016, 36(2): 428-431, 477.
- CAI G Y, XIA B B. Multimedia sentiment analysis based on convolutional neural network[J]. Journal of Computer Applications, 2016, 36(2): 428-431, 477.
- [41] YOU Q Z, LUO J B, JIN H L, et al. Cross-modality consistent regression for joint visual-textual sentiment analysis of social multimedia[C]// Proceedings of the Ninth ACM International Conference on Web Search and Data Mining. New York: ACM, 2016: 13-22.
- [42] MAJUMDER N, HAZARIKA D, GELBUKH A, et al. Multimodal sentiment analysis using hierarchical fusion with context modeling[J]. Knowledge-Based Systems, 2018, 161: 124-133.
- [43] 黄峻, 林飞, 杨静, 等. 生成式AI的大模型提示工程: 方法、现状与展望[J]. 智能科学与技术学报, 2024, 6(2): 115-133.
- HUANG J, LIN F, YANG J, et al. From prompt engineering to generative artificial intelligence for large models: the state of the art and perspective[J]. Chinese Journal of Intelligent Science and Technology, 2024, 6(2): 115-133.
- [44] 李荣, 吴晨生, 董洁, 等. ChatGPT对开源情报工作的影响及对策[J]. 情报理论与实践, 2023, 46(5): 1-5.
- LI R, WU C S, DONG J, et al. Study on the Impact of ChatGPT on open source intelligence work and countermeasures[J]. Information Studies: Theory & Application, 2023, 46(5): 1-5.
- [45] LIU H T, LI C Y, WU Q Y, et al. Visual instruction tuning[EB]. arXiv preprint, 2023, arXiv: 2304.08485.
- [46] BAI Y T, GENG X Y, MANGALAM K, et al. Sequential modeling enables scalable learning for large vision models[EB]. arXiv preprint, 2023, arXiv: 2312.00785.
- [47] GAO P, HAN J M, ZHANG R R, et al. LLaMA-Adapter v2: parameter-efficient visual instruction model[EB]. arXiv preprint, 2023, arXiv: 2304.15010.
- [48] DAI W L, LI J N, LI D X, et al. InstructBLIP: towards general-purpose vision-language models with instruction tuning[EB]. arXiv preprint, 2023, arXiv: 2305.06500.
- [49] HUA W Y, FAN L Z, LI L Y, et al. War and peace (waragent): large language model-based multi-agent simulation of world wars[EB]. arXiv preprint, 2023, arXiv: 2311.17227.
- [50] BI K F, XIE L X, ZHANG H H, et al. Accurate medium-range global weather forecasting with 3D neural networks[J]. Nature, 2023, 619(7970): 533-538.
- [51] VYKOPAL I, PIKULIAK M, SRBA I, et al. Disinformation capabilities of large language models[EB]. arXiv preprint, 2023, arXiv: 2311.08838.
- [52] EMRE K, ROBERT O, AMIT S, et al. Causal reasoning and large language models: opening a new frontier for causality[EB]. arXiv preprint, 2023, arXiv: 2305.00050.
- [53] SU J W, VARGAS D V, SAKURAI K. One pixel attack for fooling deep neural networks[J]. IEEE Transactions on Evolutionary Computation, 2019, 23(5): 828-841.
- [54] MIRSKY Y, LEE W. The creation and detection of deepfakes: a survey[J]. ACM Computing Surveys, 2020, 54: 1-41.
- [55] GOODFELLOW I J, SHLENS J, SZEGEDY C. Explaining and harnessing adversarial examples[EB]. arXiv preprint, 2014, arXiv: 1412.6572.
- [56] KURAKIN A, GOODFELLOW I J, BENGIO S. Adversarial examples in the physical world[M]//Artificial Intelligence Safety and Security. Florida: Chapman and Hall/CRC, 2018: 99-112.
- [57] MOOSAVI-DEZFOOLI S M, FAWZI A, FROSSARD P. DeepFool: a simple and accurate method to fool deep neural networks[C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Piscataway: IEEE Press, 2016: 2574-2582.
- [58] PAPERNOT N, MCDANIEL P, JHA S, et al. The limitations of deep learning in adversarial settings[C]//Proceedings of the 2016 IEEE European Symposium on Security and Privacy (EuroS&P). Piscataway: IEEE Press, 2016: 372-387.
- [59] CHEN P Y, ZHANG H, SHARMA Y, et al. ZOO: zeroth order optimization based black-box attacks to deep neural networks without training substitute models[C]//Proceedings of the 10th ACM Workshop on Artificial Intelligence and Security. New York: ACM, 2017: 15-26.
- [60] CHEN S Z, HE Z B, SUN C J, et al. Universal adversarial attack on attention and the resulting dataset DAmageNet[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(4): 2188-2197.
- [61] MOOSAVI-DEZFOOLI S M, FAWZI A, FAWZI O, et al. Universal adversarial perturbations[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Piscataway: IEEE Press, 2017: 86-94.
- [62] CHEN G K, CHEN B S, FAN L L, et al. Who is real bob? adversarial attacks on speaker recognition systems[C]//Proceedings of the 2021 IEEE Symposium on Security and Privacy (SP). Piscataway: IEEE Press, 2021: 694-711.
- [63] CISSE M, ADI Y, NEVEROVA N, et al. Houdini: Fooling deep structured prediction models[EB]. arXiv preprint, 2017, arXiv: 1707.05373.
- [64] NEWELL A, YANG K, DENG J. Stacked hourglass networks for human pose estimation[C]//Proceedings of the 14th European Conference on Computer Vision. Cham: Springer, 2016: 483-499.
- [65] THYS S, VAN RANST W, GOEDEME T. Fooling automated surveillance cameras: adversarial patches to attack person detection[C]//Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Piscataway: IEEE Press, 2019: 49-55.
- [66] LIU A S, LIU X L, FAN J X, et al. Perceptual-sensitive GAN for generating adversarial patches[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33(1): 1028-1035.
- [67] WEI X X, RUAN S W, DONG Y P, et al. Distributional modeling for location-aware adversarial patches[EB]. arXiv preprint, 2023, arXiv: 2023.16131.
- [68] DUAN R J, MA X J, WANG Y S, et al. Adversarial camouflage: hiding physical-world attacks with natural styles[C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2020: 997-1005.
- [69] WEI X X, HUANG Y, SUN Y T, et al. Unified adversarial patch for cross-modal attacks in the physical world[C]//Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2023: 4422-4431.
- [70] AKHTAR N, MIAN A, KARDAN N, et al. Advances in adversarial attacks and defenses in computer vision: a survey[J]. IEEE Access, 2021, 9: 155161-155196.
- [71] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial networks[EB]. arXiv preprint, 2014, arXiv: 1406.2661.
- [72] ISOLA P, ZHU J Y, ZHOU T H, et al. Image-to-image translation with conditional adversarial networks[C]//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Pis-

- cataway: IEEE Press, 2017: 5967-5976.
- [73] ZHU J Y, PARK T, ISOLA P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]//Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2017: 2242-2251.
- [74] KINGMA D P, WELING M. Auto-encoding variational bayes[EB]. arXiv preprint, 2013, arXiv: 1312.6114.
- [75] REZENDE D J, MOHAMED S. Variational inference with normalizing flows[EB]. arXiv preprint, 2015, arXiv: 1505.05770.
- [76] HO J, JAIN A, ABBEEL P. Denoising diffusion probabilistic models[J]. Advances in Neural Information Processing Systems, 2020, 33: 6840-6851.
- [77] POPOV V, VOVK I, GOGORYAN V, et al. Grad-TTS: a diffusion probabilistic model for text-to-speech[EB]. arXiv preprint, 2021, arXiv: 2105.06337.
- [78] SAHARIA C, CHAN W, SAXENA S, et al. Photorealistic text-to-image diffusion models with deep language understanding[J]. Advances in Neural Information Processing Systems, 2022, 35: 36479-36494.
- [79] MOLAD E, HORWITZ E, VALEVSKI D, et al. Dreamix: video diffusion models are general video editors[EB]. arXiv preprint, 2023, arXiv: 2302.01329.
- [80] WU J Z, GE Y X, WANG X T, et al. Tune-A-video: one-shot tuning of image diffusion models for text-to-video generation[C]//Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2023: 7589-7599.
- [81] WU X, HE R, SUN Z N, et al. A light CNN for deep face representation with noisy labels[J]. IEEE Transactions on Information Forensics and Security, 2018, 13(11): 2884-2896.
- [82] LIU H X, SIMONYAN K, YANG Y M. DARTS: differentiable architecture search[EB]. arXiv preprint, 2018, arXiv: 1806.09055.
- [83] LI X, LI N, WENG C, et al. Replay and synthetic speech detection with Res2Net architecture[C]//Proceedings of the ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway: IEEE Press, 2021: 6354-6358.
- [84] 周文柏, 张卫明, 俞能海, 赵汉卿, 刘泓谷, 韦天一. 人脸视频深度伪造与防御技术综述[J]. 信号处理, 2021, 37(12): 2338-2355.
- ZHOU W B, ZHANG W M, YU N H, et al. An overview of deepfake forgery and defense techniques[J]. Journal Of Signal Processing, 2021, 37(12): 2338-2355.
- [85] DOLHANSKY B, BITTON J, PFLAUM B, et al. The DeepFake detection challenge (DFDC) dataset[EB]. arXiv preprint, 2020, arXiv: 2006.07397.
- [86] JIANG L M, LI R, WU W, et al. DeeperForensics-1.0: a large-scale dataset for real-world face forgery detection[C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2020: 2886-2895.
- [87] 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. New York: ACM, 2017: 795-816.
- [88] LI Y Z, LYU S W. Exposing deepfake videos by detecting face warping artifacts[EB]. arXiv preprint, 2019, arXiv: 1811.00656.
- [89] YANG X, LI Y Z, QI H G, et al. Exposing GAN-synthesized faces using landmark locations[C]//Proceedings of the ACM Workshop on Information Hiding and Multimedia Security. New York: ACM, 2019: 113-118.
- [90] HE P S, LI H L, WANG H X. Detection of fake images via the ensemble of deep representations from multi color spaces[C]//Proceedings of the 2019 IEEE International Conference on Image Processing (ICIP). Piscataway: IEEE Press, 2019: 2299-2303.
- [91] SEGALIS E, GALILI E. OGAN: disrupting deepfakes with an adversarial attack that survives training[EB]. arXiv preprint, 2020, arXiv: 2006.12247.
- [92] BALLESTEROS D M, RODRIGUEZ-ORTEGA Y, RENZA D, et al. Deep4SNet: deep learning for fake speech classification[J]. Expert Systems with Applications, 2021, 184: 115465.
- [93] ANEJA S, BREGLER C, NIESSNER M. COSMOS: catching out-of-context misinformation with self-supervised learning [EB]. arXiv preprint, 2021, arXiv: 2101.06278.
- [94] WANG R, JUEFEI-XU F, LUO M, et al. FakeTagger: robust safeguards against DeepFake dissemination via provenance tracking[C]//Proceedings of the 29th ACM International Conference on Multimedia. New York: ACM, 2021: 3546-3555.
- [95] ZHANG Y X, WANG W C, ZHANG P Y. The effect of silence and dual-band fusion in anti-spoofing system[C]//Proceedings of the Interspeech 2021. ISCA: ISCA, 2021: 4279-4283.
- [96] WODAJO D, ATNAFU S. Deepfake video detection using convolutional vision transformer[EB]. arXiv preprint, 2021, arXiv: 2102.11126.
- [97] ZHAO Y, YI J Y, TAO J H, et al. EmoFake: an initial dataset for emotion fake audio detection[EB]. arXiv preprint, 2022, arXiv: 2211.05363.
- [98] WANG Y Q, MA F L, JIN Z W, 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. New York: ACM, 2018: 849-857.
- [99] ABDELNABI S, HASAN R, FRITZ M. Open-domain, content-based, multi-modal fact-checking of out-of-context images via online resources[C]//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2022: 14920-14929.
- [100] SHAO R, WU T X, LIU Z W. Detecting and grounding multi-modal media manipulation[C]//Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2023: 6904-6913.
- [101] SHA Z Y, LI Z, YU N, et al. DE-FAKE: detection and attribution of fake images generated by text-to-image generation models[C]//Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM, 2023: 3418-3432.
- [102] YANG X, LI Y Z, LYU S W. Exposing deep fakes using inconsistent head poses[C]//Proceedings of the ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway: IEEE Press, 2019: 8261-8265.
- [103] LI H D, LI B, TAN S Q, et al. Detection of deep network generated images using disparities in color components[EB]. arXiv preprint, 2018, arXiv: 1808.07276.
- [104] WANG J, LI Z, ZHANG C, et al. Fighting malicious media data: a survey on tampering detection and deepfake detection[EB]. arXiv preprint, 2022, arXiv: 2212.05667.
- [105] FUNG S, LU X Q, ZHANG C, et al. DeepfakeUCL: deepfake detection via unsupervised contrastive learning[C]//Proceedings of the 2021 International Joint Conference on Neural Networks (IJCNN). Piscataway: IEEE Press, 2021: 1-8.
- [106] CAO J Y, MA C, YAO T P, et al. End-to-end reconstruction-classification learning for face forgery detection[C]//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Rec-

- ognition (CVPR). Piscataway: IEEE Press, 2022: 4103-4112.
- [107] WANG C R, DENG W H. Representative forgery mining for fake face detection[C]//Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2021: 14918-14927.
- [108] LIU H G, LI X D, ZHOU W B, et al. Spatial-phase shallow learning: rethinking face forgery detection in frequency domain[C]//Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2021: 772-781.
- [109] BINH L M, WOO S. ADD: frequency attention and multi-view based knowledge distillation to detect low-quality compressed deepfake images[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2022, 36(1): 122-130.
- [110] YI J Y, WANG C L, TAO J H, et al. Audio deepfake detection: a survey[EB]. arXiv preprint, 2023, arXiv: 2308.14970.
- [111] ALZANTOT M, WANG Z Q, SRIVASTAVA M B. Deep residual neural networks for audio spoofing detection[EB]. arXiv preprint, 2019, arXiv: 1907.00501.
- [112] CHEN T X, KHOURY E, PHATAK K, et al. Pindrop labs' submission to the ASVspoof 2021 challenge[C]//Proceedings of the 2021 Edition of the Automatic Speaker Verification and Spoofing Countermeasures Challenge. ISCA: ISCA, 2021: 89-93.
- [113] TODISCO M, DELGADO H, EVANS N. Constant Q cepstral coefficients: a spoofing countermeasure for automatic speaker verification[J]. Computer Speech & Language, 2017, 45: 516-535.
- [114] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2016: 770-778.
- [115] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 7132-7141.
- [116] AL-QURISHI M, AL-RAKHAMI M, ALRUBAIAN M, et al. Selecting the best open source tools for collecting and visualizing social media content[C]//Proceedings of the 2015 2nd World Symposium on Web Applications and Networking (WSWAN). Piscataway: IEEE Press, 2015: 1-6.
- [117] ZUBIAGA A, AKER A, BONTCHEVA K, et al. Detection and resolution of rumours in social media[J]. ACM Computing Surveys, 2019, 51(2): 1-36.
- [118] CASTILLO C, MENDOZA M, POBLETE B. Information credibility on twitter[C]//Proceedings of the 20th International Conference on World Wide Web. New York: ACM, 2011: 675-684.
- [119] VOSOUGHI S, MOHSENVAND M N, ROY D. Rumor gauge[J]. ACM Transactions on Knowledge Discovery from Data, 2017, 11(4): 1-36.
- [120] ZUBIAGA A, LIAKATA M, PROCTER R, et al. Analysing how people orient to and spread rumours in social media by looking at conversational threads[J]. PLoS One, 2016, 11(3): e0150989.
- [121] ZUBIAGA A, LIAKATA M, PROCTER R. Learning reporting dynamics during breaking news for rumour detection in social media[EB]. arXiv preprint, 2016, arXiv: 1610.07363.
- [122] BONDIELLI A, MARCELLONI F. A survey on fake news and rumour detection techniques[J]. Information Sciences, 2019, 497: 38-55.
- [123] RUIZ N, BARGAL S A, SCLAROFF S. Disrupting deepfakes: adversarial attacks against conditional image translation networks and facial manipulation systems[M]//Lecture Notes in Computer Science. Cham: Springer International Publishing, 2020: 236-251.
- [124] YANG C F, DING L, CHEN Y R, et al. Defending against GAN-based DeepFake attacks via transformation-aware adversarial faces[C]//Proceedings of the 2021 International Joint Conference on Neural Networks (IJCNN). Piscataway: IEEE Press, 2021: 1-8.
- [125] 白如江, 张玉洁, 赵梦梦, 等. 面向关联推理的智慧情报感知: 内涵、组织与路径[J]. 情报理论与实践, 2022, 45(8): 31.
- BAI R J, ZHANG Y J, ZHAO M M, et al. Intelligence perception for association and reasoning: connotation, organization and path[J]. Information studies: Theory & Application. 2022, 45(8): 31.
- [126] WILLCOX S, GOLDBERG D, VAGANAY J, et al. Multi-vehicle cooperative navigation and autonomy with the bluefin cadre system[C]//Proceedings of IFAC(International Federation of Automatic Control) Conference. Oxford: Elsevier, 2006: 20-22.
- [127] BIRK A, ANTONELLI G, PASCOAL A, et al. Cognitive cooperative control for autonomous underwater vehicles[C]//Proceedings of 8th International Conference on Computer and IT Applications in the Maritime Industries. Oxford: Elsevier, 2009: 136-149.
- [128] 龚胜佳, 张琳琳, 赵楷, 等. 基于区块链技术的虚假新闻检测方法[J]. 计算机应用, 2022, 42(11): 3458-3464.
- GONG S J, ZHANG L L, ZHAO K, et al. Fake news detection method based on blockchain technology[J]. Journal of Computer Applications, 2022, 42(11): 3458-3464.
- [129] 陈晋音, 章燕, 王雪柯, 等. 深度强化学习的攻防与安全性分析综述[J]. 自动化学报, 2022, 48(1): 21-39.
- CHEN J Z, ZHANG Y, WANG X K, et al. A survey of attack, defense and related security analysis for deep reinforcement learning[J]. Acta Automatica Sinica, 2022, 48(1): 21-39.
- [130] MA F, LI Y, LI Q, et al. Faitcrowd: fine grained truth discovery for crowdsourced data aggregation[C]//Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: Association for Machinery, 2015: 745-754.
- [131] 易侃, 徐欣. 面向深度伪造的情报可信分析机理[J]. 指挥信息系统与技术, 2023, 14(1): 48-56.
- YI K, XU X. Intelligence trustworthiness analysis mechanism for deep falsification[J]. Command Information System And Technology, 2023, 14(1): 48-56.
- [132] 颜克冬, 徐琳, 宫小泽, 等. 开源情报可信分析系统的关键模型与技术[J]. 指挥信息系统与技术, 2023, 14(1): 57-61.
- YAN K D, XU L, GONG X Z, et al. Key models and technologies of open source intelligence trustworthy analysis system[J]. Command Information System And Technology, 2023, 14(1): 57-61.
- [133] DU M, LIU N, HU X. Techniques for interpretable machine learning[J]. Communications of the ACM, 2019, 63(1): 68-77.
- [134] 吴明曦. 智能化战争[M]. 北京: 国防工业出版社, 2020.
- WU M X. Intelligent warfare[M]. Beijing: National Defense Industry Press, 2020.

[作者简介]



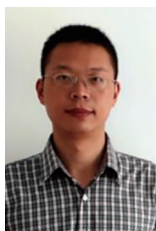
袁唯淋 (1994-), 男, 博士, 国防科技大学信息通信学院助理研究员, 主要研究方向为智能决策、多智能体强化学习、开源情报融合分析。



赵卫伟 (1978-), 男, 国防科技大学信息通信学院研究员, 主要研究方向为通信工程、人工智能、指控决策。



董绍进 (1984-), 男, 国防科技大学信息通信学院助理研究员, 主要研究方向为通信工程、网云融合。



胡振震 (1984-), 男, 国防科技大学智能科学学院博士生, 主要研究方向为对手建模、智能决策、网络安全。



王程远 (1988-), 男, 国防科技大学信息通信学院助理研究员, 主要研究方向为通信工程、无人机通信。



曹巍 (1981-), 男, 国防科技大学信息通信学院助理研究员, 主要研究方向为通信工程、无线通信。



王盛青 (1986-), 男, 国防科技大学信息通信学院副研究员, 主要研究方向为通信工程、无人机通信。



何俊 (1983-), 男, 博士, 国防科技大学信息通信学院副研究员, 主要研究方向为人工智能、自然语言处理、运筹优化。