

# 基于 CiteSpace 的图像与视频中的深度伪造技术发展现状及检测方法研究

耿祥云 严留洋 丁颢航 周翔 殷弋晖

江苏警官学院, 江苏南京, 210031;

**摘要:**近年来, 深度伪造技术 (Deepfake) 基于生成对抗网络 (GAN)、扩散模型等技术快速发展, 在图像与视频领域的滥用已引发隐私泄露、虚假信息传播、金融欺诈等社会风险。在理论层面, 本文通过 CiteSpace 的文献计量与可视化分析功能来系统梳理图像与视频中的深度伪造技术及检测方法的研究脉络、核心热点与演化路径, 致力于进一步填补该领域缺乏系统性趋势预测的空白, 完善深度伪造技术的研究体系。在实际层面, 本文意在助力技术落地与风险防控, 为监管机构和互联网平台提供深度伪造技术的发展态势图谱, 辅助制定针对性的技术规范与治理政策, 降低深度伪造内容对社会信任、个人权益造成的影响。同时, 本文旨在支撑检测技术优化, 通过可视化图谱呈现的主流检测方法及其未来发展趋势, 为企业、科研机构等研发更高效的检测工具提供实践参考, 推动技术从实验室走向实际应用场景。

**关键词:** CiteSpace; 深度伪造技术; 图像伪造检测; 视频伪造检测; 发展现状

**DOI:** 10.69979/3041-0673.26.04.008

据世界经济论坛预测, AI 生成虚假信息将成为未来全球首要风险<sup>[1]</sup>。国外研究聚焦检测技术可靠性, 而国内团队如哈工大和南洋理工联合提出首个多模态检测模型 HAMMER, 支持图像-文本联合篡改定位<sup>[2]</sup>。在国内外研究持续探索技术突破与应用的背景下, 本项目立足既有成果, 致力于通过文献计量与可视化分析构建更全面的技术图谱, 可以整合现有的图像与视频中的深度伪造技术及识别方法, 将其直观可视化呈现, 使用更易于理解的方法展现给研究者, 助力识别检测技术更新迭代, 最终将该技术应用到公安实战中去。

## 1 研究方法数据来源

本项目将首先开展文献分析, 从数据库中检索涉及深度伪造技术发展状况及相应检测方法的文献, 进行数据清洗后导出为 CiteSpace 兼容的格式, 接着通过 CiteSpace 对文献数据进行聚类、共现分析和突发性检测; 本文在此基础上进行技术研究, 识别深度伪造技术相关高频关键词, 构建知识图谱, 从生成技术演进路径等角度深入分析, 进而从研究热点方向预测未来深度伪造技术发展趋势; 最后结合法学、伦理学与信息安全等多方面因素, 对预测的深度伪造技术发展方向提供政策上的预防建议。

中文数据库选择中国知网 (CNKI) 和万方数据知

识服务平台, 专业检索 (主题: 图像与视频中的深度伪造技术) AND (主题: (“深度伪造” OR “深度学习” OR “视频检测” OR “算法研究” OR “方法研究” OR “伪造检测” OR “特征融合” OR “人脸检测” OR “检测方法” OR “检测算法” OR “伪造人脸检测” OR “人脸图像” OR “人脸深度伪造防御” OR “视频检测算法” OR “人脸伪造检测” OR “特定人物” OR “溯源方法” OR “基于图像” OR “技术研究” OR “双流网络” OR “卷积神经网络” OR “数据集” OR “注意力机制” OR “检测模型” OR “卷积核” OR “特征提取”)) (两种数据库的检索语法有所差别, 以万方数据库为例), 发表时间限定在 2005-2025 年, 来源类型选择 “全部期刊”, 检索时间为 2025 年 10 月 21 日。在 CNKI 数据库检索出 146 篇中文文献, 其中 143 篇为有效文献 (删除综述类文献、学术报道、无关文献 3 篇); 在万方数据库检索出 235 篇中文文献, 其中 229 篇为有效文献 (删除科技成果、地方志条目 6 篇)。

因 CiteSpace 软件默认采用 WoS 格式开展数据分析, 需先对 CNKI 与万方数据库的文献数据进行格式转换。最终 CNKI 与万方数据库的 164 篇中文文献完成格式转换。

## 2 图像与视频中的深度伪造技术研究热点分析

### 2.1 关键词共现分析

一般而言，一篇文献里能反映文章核心内容或者主题的往往是关键词，通过对关键词的内容加以分析，能够找到领域内研究的发展脉络和热点方向。运行 CiteSpace 软件，调整节点类型为关键词，使用“Pathfinder”算法精简图谱，得到关键词共现图谱（如图 1）<sup>[4]</sup>。

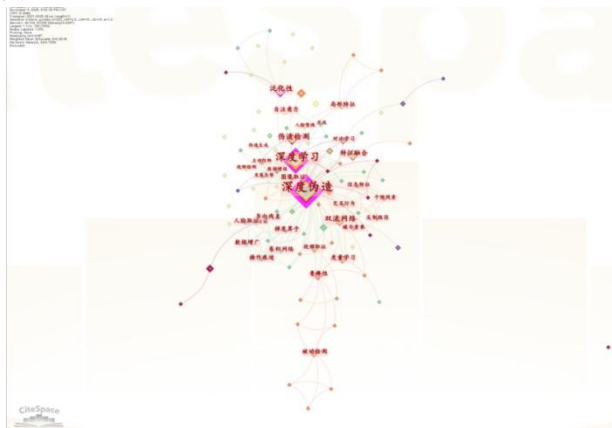


图 1 图像与视频中的深度伪造技术发展现状及检测方法（知网、万方）文献关键词共现图谱

在该图谱中，以深度伪造技术为中心发散的周围，有“深度学习”“伪造检测”“泛化性”“双流网络”“视频取证”“卷积网络”等多个相关性较大、出现频次较高的关键词，但其中的“鲁棒性”“信息特征”“局部特征”等关键词的表意过于宽泛，且“深度学习”“伪造检测”“伪造生成”等关键词受文献检索所用的关键词的影响较大，也不具有分析价值，综合来看有效的热点关键词包括“泛化性”“对比学习”“数据增广”“局部特征”“度量学习”“视频检测”等。

因此深度伪造技术在公安实战中的研究热点对象应该为“泛化性”“视频检测”“局部特征”，热点的研究方法为“对比学习”“数据增广度量学习”。

### 2.2 关键词聚类分析

CiteSpace 的关键词聚类分析可以把领域内高频关键词分组聚类，呈现研究热点主题、核心议题，以及各主题间的关联，快速摸清领域研究的“核心版图”<sup>[3]</sup>。设置显示最大的 8 个聚类，得到布局较为理想的关键词聚类图谱（如图 2）。已知模块聚类值  $Q > 0.3$  表明聚类结构显著，聚类平均轮廓值  $S > 0.7$  表明聚类结果令人信服<sup>[4][5]</sup>。观察本图可以得知图中  $Q=0.6187, S=0.9018$ 。因此，该聚类分析结果有效。

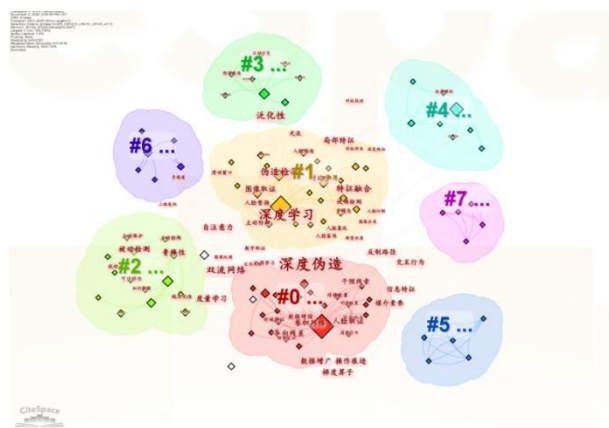


图 2 图像与视频中的深度伪造技术发展现状及检测方法（知网、万方）文献关键词聚类图谱

经过 CiteSpace 的关键词聚类分析，可以看到有 4 个聚类非常显目，分别是：“深度伪造”“深度学习”“被动检测”“泛化性”。其中，“深度伪造”与“深度学习”二者研究热点最高。不难发现，“深度学习”是领域的核心方向，涵盖“特征融合”“图像取证”等关键词，体现了深度伪造检测的核心技术与应用场景。本图中各聚类主题覆盖了深度伪造的“技术实现”“检测方法”“性能优化”“特征表征”等核心，清晰地呈现了该领域的研究热点与结构。

### 3 图像与视频中的深度伪造技术国内研究前沿分析

若对于研究前沿进行分析，就必须结合对施引文献的分析和对突现词、突现文献的分析<sup>[6]</sup>。因此，本文使用 CiteSpace 软件的“BrustDetection（突现检测）”功能，对图像与视频中的深度伪造技术及检测方法相关文献的突现词进行分析，对关键词的突现频率进行统计，并通过对突现关键词节点信息的二次检索，识别出我国该领域研究前沿的演进<sup>[7]</sup>（如图 3）

在中文文献的分析中，共得到 2 个突显词。“Year”表示该突现词最早出现的年份，“Strength”表示突现强度，突现强度越大，该关键词研究热度越高，“Begin”“End”分别表示该关键词作为研究前沿的起止年份，如图中红色线段所示。

#### Top 23 Keywords with the Strongest Citation Bursts

Keywords	Year	Strength	Begin	End
图像取证	2021	0.7	2021	2023
人脸篡改	2022	0.35	2022	2023

图 3 图像与视频中的深度伪造技术发展现状及检测方法（知网、万方）文献突现词

突现在 2021 年以后的关键词为近期研究前沿，包括“图像取证”“人脸篡改”。其中，“图像取证（0.7）”的表现强度最大，“人脸篡改（0.35）”的表现强度较弱。“图像取证（3年）”的突现时间较长，说明其在很长一段时间内成为研究者关注的焦点。“人脸篡改（2年）”的突现时间较短，说明其研究价值有限。“图像取证”成为前沿研究对象，一方面是由于深度伪造技术的迅猛发展，图像伪造逐渐成为最具破坏力的深度伪造形式；另一方面，图像取证具有研究的技术优势。图像包含空域、频域、纹理、光照等丰富信息，能为检测提供多样化线索。相比于视频整体分析，图像关键帧检测计算成本更低，适合大规模内容筛查。例如有学者提出频率感知判别特征学习，结合单中心损失监督，专注于捕捉伪造图像的频域异常<sup>[8]</sup>；也有学者设计了多注意力机制，同时捕捉图像空间特征和时序一致性，显著提升了检测精度<sup>[9]</sup>。

## 4 展望

### 4.1 构建“产学研用”一体化联防联控局面

应对深度伪造这一复杂挑战，单打独斗是不可行的，必须建立跨领域的协作关系。建议由网信、公安科技部门牵头，推动建立“高校-研究机构-互联网平台-公安基层”四位一体的联合实验室。这种合作不应停留在项目申报，而要形成常态化的数据共享、案例共研、技术共训机制。特别是在政策层面，应加快研究并出台深度伪造内容生成与传播的行业标准和管理办法，明确技术应用的边界，为监管执法提供清晰的法律依据。此外，也可以鼓励行业自行制定公约，从源头减少技术的滥用风险。

### 4.2 以公安实战需求进一步驱动技术迭代升级

一切技术的最终价值体现在解决实际问题。未来，技术研发需要更紧密地贴合公安实战的具体场景，建议建设覆盖各类深度伪造样本的数据库，并持续更新，为模型训练提供数据来源，将研发的检测技术与电子证据取证、案件研判流程深度融合，开发出与现有警务系统无缝对接的新办法，提升办案效率。最终目标是搭建一个技术检测、风险预警、溯源处置于一体的实战化平台，并为维护社会诚信体系和公共安全提供坚实支撑。

## 参考文献

[1]World Economic Forum. The Global Risks Repo

rt 2024[R]. Geneva: World Economic Forum, 2024.

[2]Zhang Y, Li J, Wang H, et al. HAMMER: A Multimodal Deepfake Detection Model for Image-Text Joint Tampering Localization[J]. IEEE Transactions on Information Forensics and Security, 2023, 18: 6221 - 6235.

[3]Chen C. Mapping Scientific Frontiers: The Quest for Knowledge Visualization[M]. London: Springer, 2003.

[4]陈超美. 引文空间分析原理与应用(CiteSpace 实用指南) [M]. 北京: 科学出版社, 2014.

[5]Chen C, Ibekwe-SanJuan F, Hou J. The structure and dynamics of co-citation clusters: A multiple-perspective analysis[J]. Journal of the American Society for Information Science and Technology, 2010, 61(7): 1386-1409.

[6]Li, J., Xie, H., Li, J., Wang, Z., & Zhang, Y. (2021). Frequency-aware discriminative feature learning supervised by single-center loss for face forgery detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 16264-16273).

[7]Zhao, H., Zhou, W., Chen, D., Wei, T., Zhang, W., & Yu, N. (2021). Multi-attentional deep fake detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 2185-2194).

[8]刘思雨. 基于 CiteSpace 的《语言教学与研究》20 年文献计量分析 [D]. 南京师范大学, 2021.

[9]侯剑华. 工商管理学科演进与前沿热点的可视化分析 [D]. 大连理工大学, 2009.

作者简介：耿祥云（2005—），男，汉，江苏南京人，本科在读，单位：江苏警官学院，研究方向：刑事科学技术。

基金项目：江苏警官学院大学生创新创业训练计划项目《基于 CiteSpace 的图像与视频中的深度伪造技术发展现状及检测方法研究》（项目编号：XJ202510329121）