# 前　言

韩冀中[1], 陶建华[2], 陆哲明[3], 陈　恺[1,4]

[1]中国科学院信息工程研究所 北京 中国 100093
[2]中国科学院自动化研究所 北京 中国 100190
[3]浙江大学 杭州 中国 310058
[4]中国科学院大学 网络空间安全学院 北京 中国 100049

随着生成对抗网络等人工智能技术的发展, 深度伪造技术及其应用不断成熟, 可以自动生成高度逼真且难以甄别的虚假多媒体信息。近两年, 为解决该技术可能带来的安全风险, 深度伪造的生成、鉴别以及防御技术已经引起工业界及研究人员的广泛关注。本期专题旨在总结当前国内外研究趋势, 并展示国内研究人员在深度伪造鉴别与防御技术方向的最新研究成果。

本期专题征稿历时 4 个月, 从 16 篇稿件中遴选出 7 篇收录。每篇稿件均经过多次审稿与复审。专题收录 4 篇综述性文章, 它们分别从网络图像与视频鉴伪、生物特征检测、语音鉴伪以及对抗样本生成的角度梳理与分析深度伪造鉴伪与防御技术的发展趋势; 并收录 3 篇深度伪造检测技术文章, 它们分别通过观测帧间差异, 融合时序与空间特征, 以及双流网络等手段提出了视频鉴伪的解决方案。

《深度伪造检测技术研究综述》围绕图像、视频和音频中的深度伪造, 从伪造和反伪造两个角度阐述了深度伪造技术的基本原理、发展脉络以及未来趋势, 并完整梳理了当前的专用数据集, 最后从构建深度伪造检测与防御体系的角度, 提供了可行的解决思路。

《人脸反欺诈活体检测综述》从生物特征的角度揭示人脸活体检测与深度人脸伪造的关系, 分析了近年来的人脸反欺诈活体检测技术, 总结了可用的人脸活体检测的数据集, 并分类归纳了基于传统方法和深度学习方法的人脸活体检测算法, 最后分析了技术的发展趋势。

《深度语音伪造与鉴伪的发展与挑战》围绕基于深度学习技术的语音伪造与鉴伪展开论述, 不仅揭示了现有语音伪造和鉴伪中的问题, 梳理和分析了三类深度语音伪造关键技术的发展历程、优势与不足, 并指出未来语音鉴伪的三个重点研究方向。

《对抗样本生成技术概述》从视觉对抗样本生成的角度分析深度伪造对信息内容安全体系的影响, 在信号层、内容层以及语义层三个层面剖析了对抗样本生成技术, 并详细分析了生物识别系统中人脸识别模型的安全性, 最后指出对抗样本生成技术中仍需探索的关键问题。

《基于帧间差异的人脸篡改视频检测方法》利用伪造视频中相邻视频帧存在显著差异的特点, 提出一种基于帧间差异的人脸篡改视频检测框架。先通过 LBP/HOG 等特征的检测方法验证检测框架的有效性, 再设计了基于孪生网络来增强人脸图像特征表示的检测方法, 有效提高了检测效果。

《基于全局时序和局部空间特征的伪造人脸视频检测方法》利用视频中的全局时序特征和局部空间特征进行综合分析, 不仅发掘出连续多帧图像中的伪造痕迹, 而且重点关注了眼睛、嘴巴和鼻子等五官区域的局部信息, 从而提高了检测的准确率。

《一种基于双流网络的 Deepfakes 检测技术》利用了视频生成时图像压缩对噪声域影响较小的发现, 设计了基于 EfficientNet 的双流网络检测框架, 通过 RGB 和噪声两个维度学习图像的特征, 并提出在判断结果上进行融合的方法, 有效提高了虚假视频检测方法对抗视频压缩的能力。

我们要特别感谢《信息安全学报》编委会对本期专题工作的信任和指导, 感谢编辑部各位工作人员从征稿启事发布、审稿专家邀请至评审意见汇总、论文定稿、修改、校对和出版所付出的辛勤工作和汗水, 非常感谢专题评审专家及时、专业、细致的评审。我们还要感谢向专题踊跃投稿的各位作者。

最后, 感谢本期专题的读者们, 希望专题能够有助于你们的技术研究工作。