APSIPA Transactions on Signal and Information Processing, 2025, 14, e206

Editorial

Editorial for the Special Issue on Deepfakes, Unrestricted Adversaries, and Synthetic Realities in Generative AI Era

Isao Echizen[1], Minoru Kuribayashi[2], Yuhong Liu[3] and Huy H. Nguyen[4]

[1] *National Institute of Informatics, Japan; iechizen@nii.ac.jp*
[2] *Tohoku University, Japan; kminoru@tohoku.ac.jp*
[3] *Santa Clara University, USA*
[4] *National Institute of Informatics, Japan*

The rise of generative AI has led to a notable increase in the use of deepfake technology, accompanied by a concerning surge in unrestricted adversarial examples. This trend has posed significant concerns in the signal and information processing community. Initially limited to visual media, deepfakes have evolved to include multimodal elements, seamlessly integrating audio, text, and imagery to create elaborate narratives. This advancement has given rise to synthetic realities where distinguishing between authentic and manipulated content poses an increasingly formidable challenge.

Moreover, the proliferation of deepfakes has coincided with unrestricted adversaries, which exploit inherent vulnerabilities in machine learning models and intensify existing challenges. While much research has concentrated on defending against norm-constrained unimodal attacks, it is crucial to address unrestricted adversaries, which pose significant risks to the reliability and safety of AI systems across various domains. Expanding research efforts to encompass uni- and multi-modal adversaries is vital for developing robust defenses and ensuring the integrity of AI-driven technologies in real-world applications.

This special issue focuses on the two sides of the coin – generation and defense against deepfakes, unrestricted adversaries, and synthetic realities. This special issue has collected five excellent articles reviewed and highly recommended by the editors and reviewers.

ISSN 2048-7703; DOI 10.1561/116.20250200

The 1st paper is "3D Morphable Master Face: Towards Controllable Wolf Attacks Against 2D and 3D Face Recognition Systems," authored by Siyun Liang, Huy H. Nguyen, Satoshi Ikehata, Junichi Yamagishi, and Isao Echizen. This paper proposes a systematic approach (3D Morphable Master Face) for generating master faces that can defeat both 2D and 3D face recognition systems by generating face samples that match multiple registered user templates in a database.

The 2nd paper is "How Much is the Source Mismatch an Important Problem for Deepfake Detection?" authored by Antoine Mallet, Rémi Cogranne, Minoru Kuribayashi, and Arthur Méreur. This paper addresses the fundamental problem of "source mismatch," in which a model is trained on a specific deepfake generation source and tested on a different source, investigates its causes and effects, and proposes solutions to this important issue.

The 3rd paper is "Navigating Real and Fake in the Era of Advanced Generative AI," authored by Huy H. Nguyen, Siyun Liang, Junichi Yamagishi, and Isao Echizen. This paper proposes a comprehensive countermeasure framework to address a wide range of attacks related to generative AI, considering that a perspective that transcends the dichotomy of "real or fake" is important in today's world, where machine-generated content and human-created content coexist.

The 4th paper is "InaSAS: Benchmarking Indonesian Speech Antispoofing Systems," authored by Candy Olivia Mawalim, Sarah Azka Arief, and Dessi Puji Lestari. This paper proposes a comprehensive benchmark dataset for spoofing detection in Indonesian and, through evaluation experiments, clarifies the challenges posed by the diversity of Indonesian and the influence of demographic factors.

The 5th paper is "Print and Scan Simulation for Adversarial Attacks on Printed Images," authored by Nischay Purnekar, Benedetta Tondi, Jana Dittmann, and Mauro Barni. This paper examines adversarial samples with print-and-scan resistance targeting an AI-based printer attribute system that identifies the printer used to create printed documents.

Through the five papers published in this special issue, we hope that readers will not only gain a broad perspective on the misuse of generative AI, but also acquire insights into advanced attack and defense methods in the Generative AI Era. These insights are expected to be beneficial in establishing defense strategies against new threats emerging in the rapidly evolving field of generative AI (e.g., uni- and multi-modal adversaries) and contribute to the development of trustworthy AI systems.

Guest Editors
Isao Echizen, Minoru Kuribayashi, Yuhong Liu and Huy H. Nguyen