UNIVERSITY of GUELPH

COLLEGE of ENGINEERING AND PHYSICAL SCIENCES

SCHOOL OF COMPUTER SCIENCE

PhD Defence

Tuesday May 14, 2024, at 1PM, online via Zoom (Remote)

Qi Li

Audio Forensic Analysis and Authentication Using Deep Learning

Chair: Dr. Stacey Scott
Advisor: Dr. Xiaodong Lin
Advisory: Dr. Lei Lei [School of Engineering]
Non-Advisory: Dr. Rozita Dara
External Examiner: Dr. Lingyu Wang [Concordia University]

Abstract:

In the swiftly evolving digital era, the importance of audio applications in communication, social networking, and entertainment sectors has dramatically increased, signaling immense opportunities and challenges. Audio encompasses a rich array of information, such as speech and background sounds, which play a crucial role in supporting investigations and legal processes. While the identification of speakers through their voices has been a direct method for establishing identities, the significance of background noise as a vital source of forensic evidence has been largely underestimated and remains underexplored. Moreover, in the legal and forensic fields, there is a paramount emphasis on maintaining the authenticity and integrity of audio evidence. Nevertheless, the advent of advanced deepfake technologies poses significant challenges to conventional methods of audio forensics, particularly those dependent on voice recognition. This development raises serious concerns about the reliability and acceptability of voice evidence within legal frameworks, necessitating a reevaluation of traditional approaches to audio authentication.

In addressing these challenges, this thesis turns its attention to the frequently overlooked non-speech elements within audio recordings. It introduces a pioneering framework that employs automation for the extraction, segregation, and classification of background noises, thereby paving the way for a new forensic analysis methodology. This approach significantly extends the boundaries of audio forensic capabilities beyond the conventional reliance on speaker identification, unveiling a novel pathway for evidence evaluation in legal contexts, particularly beneficial in settings characterized by a cacophony of mixed voices and a variety of background sounds.

Finally, to address the challenges of traceability in deepfake content and guarantee the integrity of audio, this thesis proposes a third solution specifically designed to counteract voice conversion and synthetic speech attacks. Leveraging cutting-edge deep learning technology, a revised training strategy with an interchangeable float encoder and ensemble learning of distortion layer, this approach not only overcomes the inherent limitations of existing forensic methods but also resolves the training issues associated with high-capacity watermarks. It achieves exceptionally high accuracy and imperceptibility across multiple speech datasets, various synthetic forgery methods, and numerous speech processing algorithms.