Ph.D. in Information Technology
Thesis Defense

July 12, 2022
at 10:00
Room Seminari "Alessandra Alario" and online by Webex

Clara BORRELLI – XXXIV Cycle

Data Driven and Signal Processing Techniques for Audio Forensics

Supervisor: Prof. **Augusto Sarti**

Abstract:

The recent developments and diffusion of audio recording devices, audio editing tools and speech synthesis techniques have opened questions about how to verify the authenticity and integrity of audio assets. On one side, audio recordings are frequently used as fundamental assets in trials and audio analysis methods are needed to asses their admissibility in court. On the other side, falsification of digital media represents nowadays a menace for modern communication and information ecosystems. Fake news, distributed through social media platforms, are frequently distributed together with forged media content, to acquire credibility at the eyes of deceived users and to increase the engagement. The development of detection methods able to expose fake speech signals is therefore paramount.

In this thesis we propose a set of methods for both authenticity and integrity assessment in audio forensics scenarios. Depending on the context, the analysis aims at retrieving information on the recording acoustic scenario or on the speech signal origin. Authenticity is evaluated by matching the extracted cues with a preliminary hypothesis while manipulations are detected by looking at cue's inconsistencies over time.

In the last years, the audio forensic research community has frequently addressed these two problems, proposing solutions based on digital signal processing techniques or, more recently, the combination of hand-crafted features with supervised classic machine learning method. In this work we present new methods that expand this approach with the use of recent neural-network-based architectures and, by combining all these different strategies, able to successfully address various different scenarios. If large training audio corpora are available, leveraging deep neural networks allows to extract high-level semantic information and to achieve higher generalisation ability and robustness. On the contrary, if either available data or computational power is reduced, methods based on signal model and low-level descriptors are more suitable and still successful, even if less robust to possible small modifications of the input audio.

With this paradigm in mind, we first focus on the definition of two indicators of the acoustic recording environment and present how to blindly estimate them from single-channel noisy audio signal. Then, we focus on synthetic speech detection and attribution for authenticity assessment, presenting solutions that analyse speech signals at various abstraction levels. Finally, two integrity verification methods are presented, focusing in particular on splicing identification and localisation.

All methods are validated trough a set of experiments designed to test at the same time detection performance and robustness in real-world conditions.
This thesis represents a preliminary investigation, which we hope will help widening the perspectives of audio forensic research.

PhD Committee
Prof. **Marco Marcon**, DEIB-Politecnico di Milano
Prof. **Tiziano Bianchi,** Politecnico di Torino
Prof. **Simone Milan**i, Universita' di Padova