Alessandro Pianese

Audio-Visual Deepfake Detection

Tutor: Prof. Giovanni Poggi

Cycle: XXXVIII

Year: First

Finanziato dall'Unione europea
NextGenerationEU

My background

- **MSc degree** in Computer Science, Intelligent systems and visual computing – University of Groningen
- **Research group**: GRIP (Image Processing Research Group)
- **PhD start date**: 03/01/2023
- **Scholarship type**: PNRR - Centro Nazionale CN1 HPC Digital & Smart cities
- **Cooperation**: N/A

Research field of interest

- **Multimedia Forensics**:
  - Analysis of forensic clues from audio and/or visual data
- **Audio Deepfake detection**:
  - Is this audio totally/partially fake? Has this person said such things?
- **Video Deepfake Detection**:
  - Has this video been manipulated?
    Audio manipulation is essential to video deepfake generation!

Which one is real and which is fake?

Fake Fake Fake Real

Summary of study activities

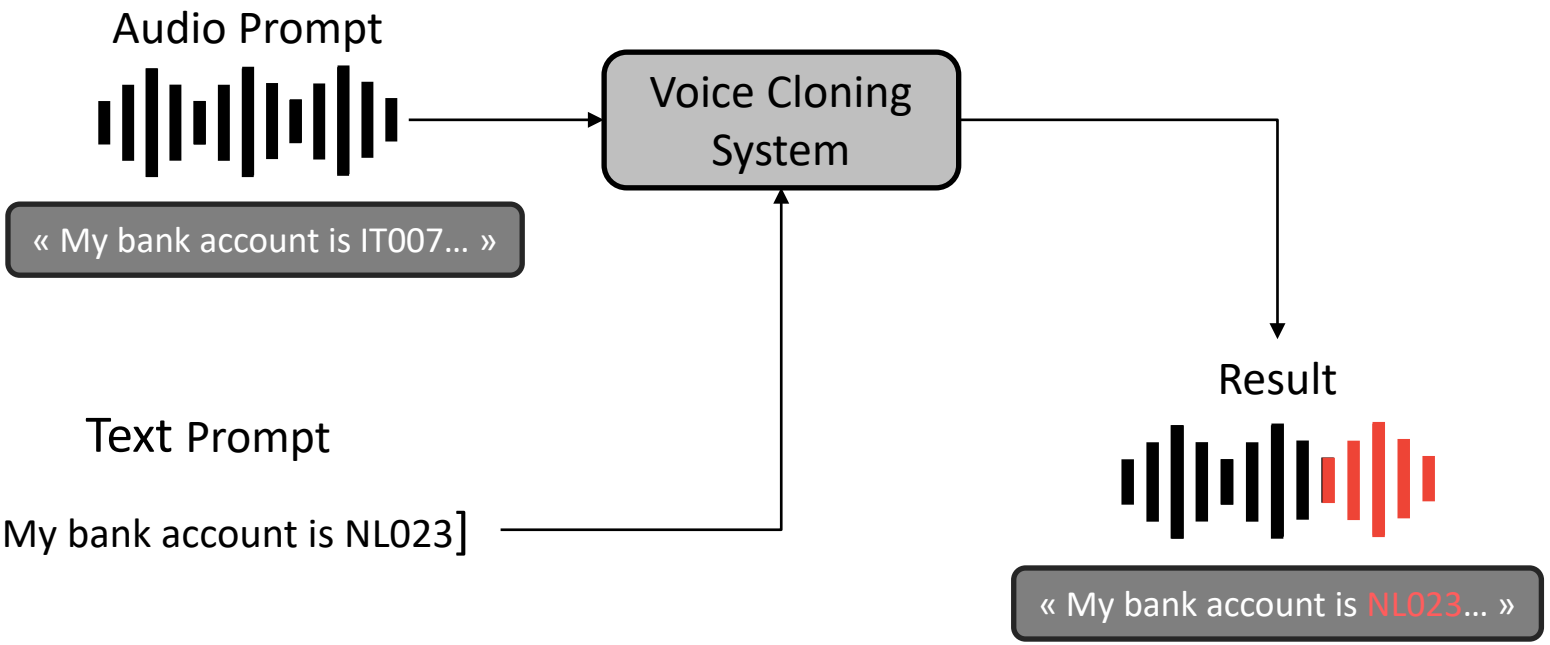
II year	Courses	Seminars	Research	Tutorship
Total	17	5.6	33.3	0
Expected	20 - 40	5 - 10	10 - 35	0 – 1.6

- Study of the state-of-the-art methods for single modality and multimodality audio/visual deepfake detection.

- **PhD School:**
  - 2023 IEEE SPS / EURASIP - Summer School on Metaverse Technologies, University of Cagliari
- **PhD courses:**
  - *Using Deep Learning Properly* – Dr. Andrea Apicella
  - *How to boost your PhD* - Prof. Antigone Marino
  - *Visione per Sistemi Robotici* – Dr. Davide Cozzolino
- **Conference:**
  - IEEE/CVF *Computer Vision and Pattern Recognition Conference* (**CVPR**), Vancouver, Jun 18-22, 2023

Research activity: Overview

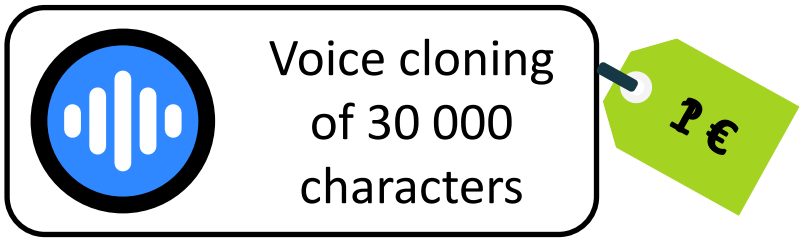
- Problem
  - **Voice cloning** has become very simple



Research activity: Overview

- Problem
  - **Voice cloning** has become very **simple**
  - Non skilled users can achieve good results for **extremely cheap**

Good quality entry-level cost for voice cloning?



Research activity: Overview

- Problem
  - **Voice cloning** has become very **simple**
  - Non skilled users can achieve good results for **extremely cheap**
  - It can be very **dangerous**


Forbes

FORBES > INNOVATION > CYBERSECURITY

EDITORS' PICK

Fraudsters Cloned Company Director's Voice In $35 Million Heist, Police Find


The Washington Post
Democracy Dies in Darkness

TECH Artificial Intelligence Help Desk Internet Culture Space Tech Policy

INNOVATIONS

They thought loved ones were calling for help. It was an AI scam.

Scammers are using artificial intelligence to sound more like family members in distress. People are falling for it and losing thousands of dollars.

Sources:
https://www.forbes.com/sites/thomasbrewster/2021/10/14/huge-bank-fraud-uses-deep-fake-voice-tech-to-steal-millions/?sh=cc78bb675591
https://www.washingtonpost.com/technology/2023/03/05/ai-voice-scam/

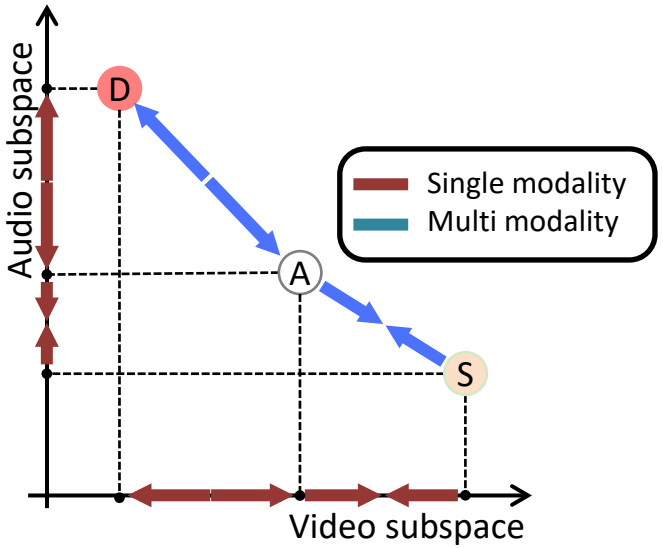
Research activity: Overview

- Problem
  - **Voice cloning** has become very **simple**
  - Non skilled users can achieve good results for **extremely cheap**
  - It can be very **dangerous**

- Objective
  - Develop techniques for **reliable** audio/visual deepfake detection
  - Methods need to be **robust** against social media compression and adversarial attacks

Research activity: Methodology

- POI Based detection
  - Training force embedded vectors of a reference video (A) to be close to vectors of the same subject (S) but far from those of different subjects (D)



Research activity: Results

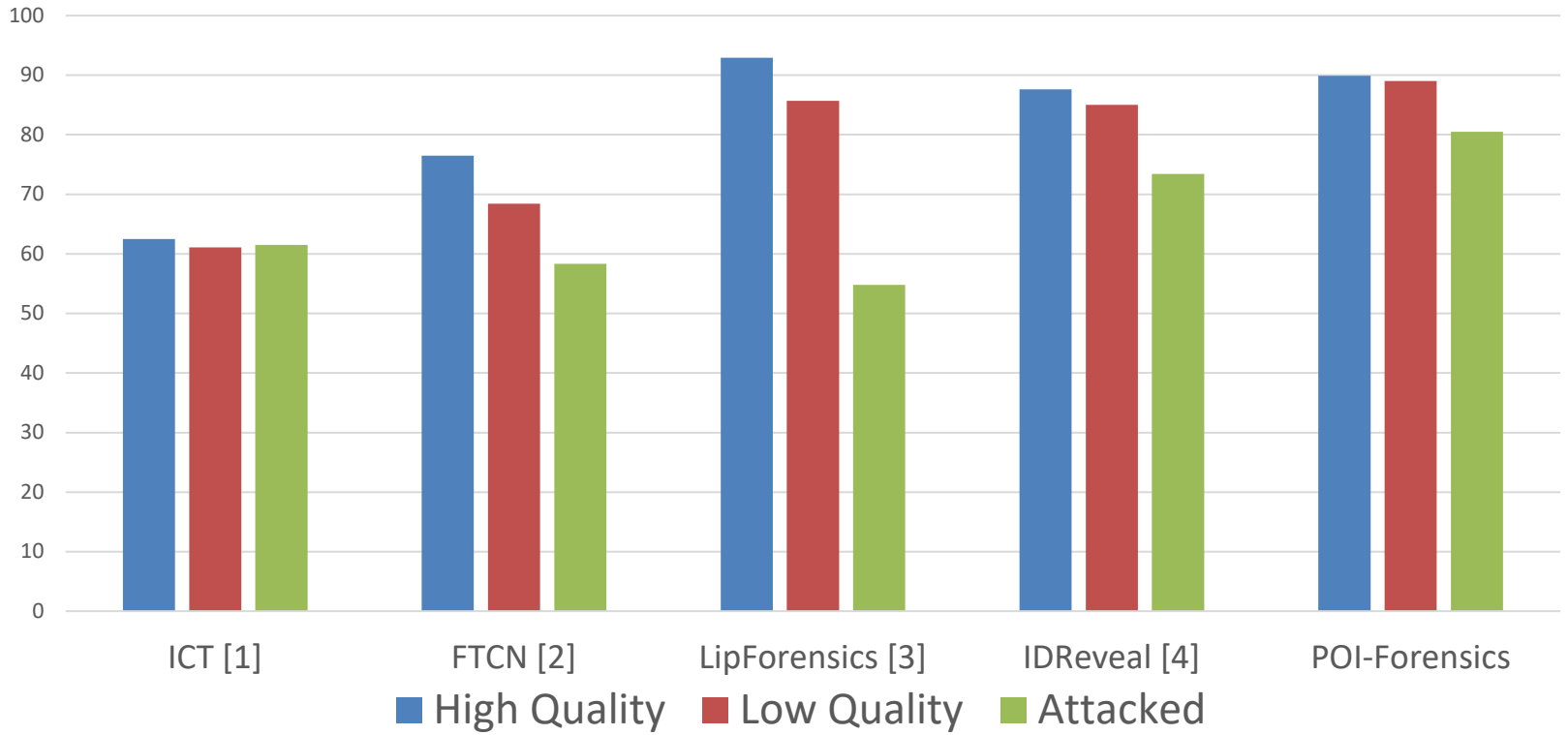
- Comparison with state-of-the-art (high quality)

AUC/ACC	pDFDC	DF-TIMIT	FakeAVCel.	KoDF	AVG
ICT [1]	77.1/70.7	87.8/77.1	68.2/63.9	62.5/58.9	73.9/67.7
FTCN [2]	72.3 / 63.9	**100.** / 87.4	84.0 / 64.9	76.5 / 63.0	83.2 / 69.8
LipForensics [3]	68.7 / 60.0	98.8 / 78.0	**97.6** / 83.3	**92.9** / 56.1	89.5 / 69.3
ID-Reveal [4]	91.3 / 80.4	99.0 / **92.8**	70.2 / 60.3	87.6 / 63.7	87.0 / 74.3
POI-Forensics	**95.2** / **86.7**	99.2 / 85.7	94.1 / **86.6**	89.9 / **81.1**	**94.6** / **85.0**

[1] Dong, X., et al. "Protecting celebrities from deepfake with identity consistency transformer." Proceedings of the IEEE/CVF CVPR. 2022.
[2] Zheng, Y., et al. "Exploring temporal coherence for more general video face forgery detection." Proceedings of the IEEE/CVF ICCV. 2021.
[3] Haliassos, A., et al. "Lips don't lie: A generalisable and robust approach to face forgery detection." Proceedings of the IEEE/CVF CVPR. 2021.
[4] Cozzolino, D., et al. "Id-reveal: Identity-aware deepfake video detection." Proceedings of the IEEE/CVF ICCV. 2021.

Research activity: Results

- Rosbustness analysis



[1] Dong, X., et al. "Protecting celebrities from deepfake with identity consistency transformer." Proceedings of the IEEE/CVF CVPR. 2022.
[2] Zheng, Y., et al. "Exploring temporal coherence for more general video face forgery detection." Proceedings of the IEEE/CVF ICCV. 2021.
[3] Haliassos, A., et al. "Lips don't lie: A generalisable and robust approach to face forgery detection." Proceedings of the IEEE/CVF CVPR. 2021.
[4] Cozzolino, D., et al. "Id-reveal: Identity-aware deepfake video detection." Proceedings of the IEEE/CVF ICCV. 2021.

Products

[P1]	**Conference Paper** D. Cozzolino, **A. Pianese**, M. Nießner, and L. Verdoliva "Audio-Visual Person-of-Interest Deepfake Detection" In *Workshop on Multimedia Forensics* (WMF) at *IEEE Conference on Computer Vision and Pattern Recognition* (CVPR), June 2023, Vancouver, pp. 943-952

Next Year

- Study the impact of text-based information during training for few/zero shot classification of audio signals
- Improve multimodal fusion strategies

Thank you for the attention!