UNIVERSITÀ DEGLI STUDI DI NAPOLI FEDERICO II

itee PhD information technology electrical engineering

DIETI UNINA

PhD Riccardo Corvi

Synthetic Video Detection

Tutor: Luisa Verdoliva

Cycle: XXXVIII

Year: Second

itee PhD information technology electrical engineering

My background

- **MSc degree** in Computer Engineering – Università degli Studi di Napoli Federico II
- **Research group:** GRIP (Image Processing Research Group)
- **PhD start date:** 01/11/2022
- **Scholarship type:** UNINA - DII, DISCOVER project, funded by DARPA under the SEMAFOR program
- **Part-Time internship:** NVIDIA 11/03/2024 - 11/03/2025

Research field of interest

- **Multimedia Forensics**:
  - Development of techniques for the forensic analysis of images, audios and videos
- **Synthetic Video Detection**:
  - Identify if a video is AI generated or not

Synthetic videos downloaded from Social media
(Left: Twitter, Right: Reddit)

Summary of study activities

	Courses	Seminars	Research	Tutorship
Total	13	5.6	38.6	0
Expected	10 - 20	5 - 10	30 - 45	0 – 1.6

State-of-the-art analysis in synthetic image detection and attribution

- **PhD courses:**

  "Strategic Orientation for STEM Research & Writing" (Dr. Chie Shin Fraser)
  "Innovation and Entrepreneurship" (Prof. Pierluigi Rippa)

- **Summer schools**:

  2024 IEEE-EURASIP Summer School on Signal Processing (S3P-2024)
  "Understanding and modeling the world around us"
  Capri, Italy, from 23/09/2024 to 27/09/2024

- **Conference**:

  IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR), Seattle, USA, from 16/06/2023 to 20/06/2023

Research activity: Overview

- Problem
  - Easy access to **generative AI** allows to easily spread disinformation over the web
  - The advent of new types of synthetic generators has led to **new ways** to spread fake news
- Objective
  - Analyze the **artifacts** present in synthetic videos to gain insight into the most discriminative features
  - Develop a synthetic video detector that can **generalize** across different AI-based models

Research activity: Dataset creation

- We created a dataset of synthetic videos generated with 4 text-to-video generators
- We used captions of the real videos from the Panda70M dataset. Real and the synthetic data have the same content

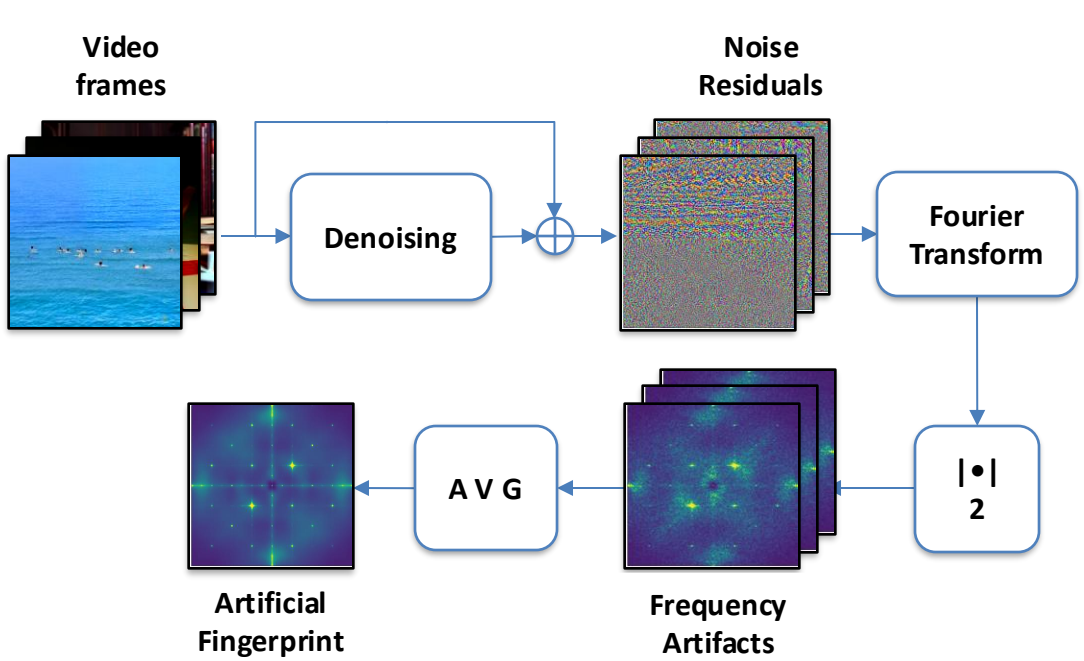
Original

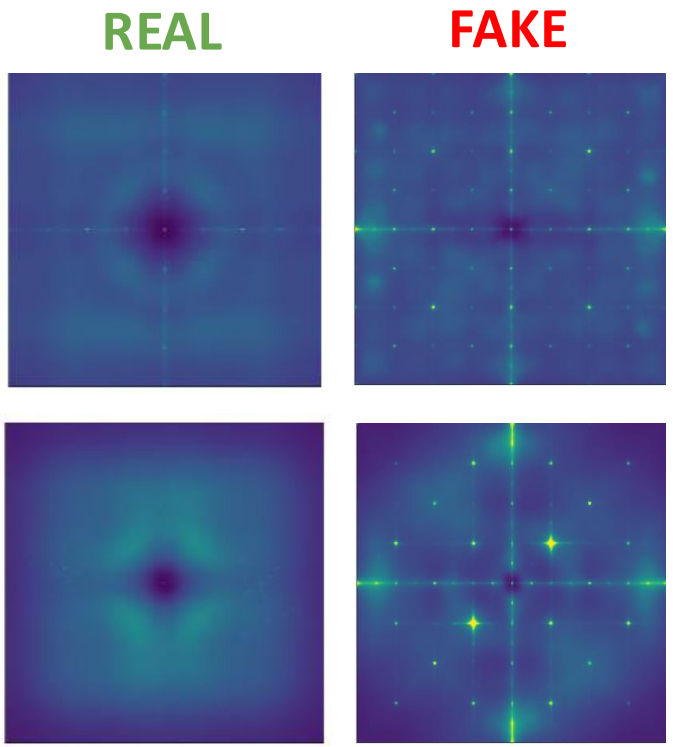
Synthetic video

Caption: A woman wearing a blue shirt is interviewed by a news reporter

Research activity: Analysis

- We extracted artifacts in the spatial and frequency domain by computing the power spectra of noise residuals aggregated though the time domain

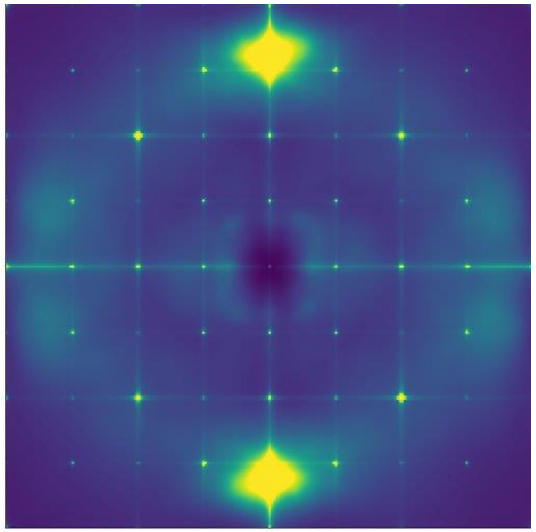




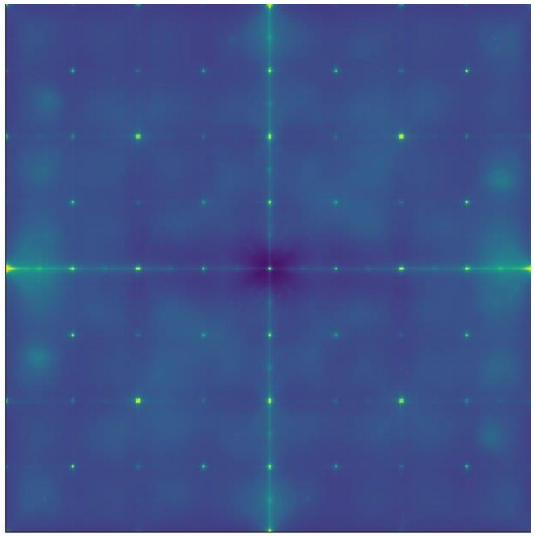
Research activity: Analysis

- Such artifacts can be very different based on the type of generator

- It is difficult then that one detector can generalize well to all the possible generators without re-training



Generator 1

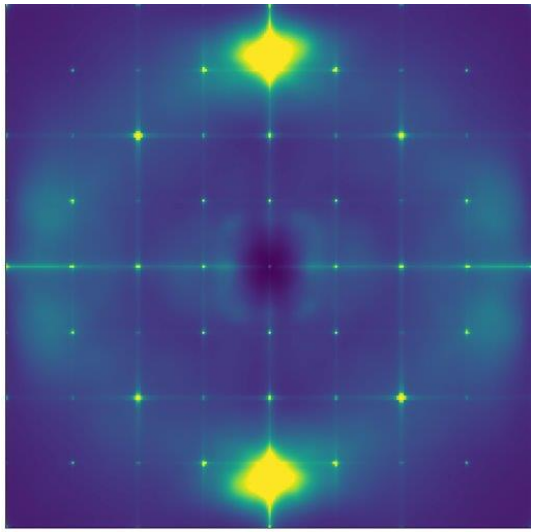


Generator 2

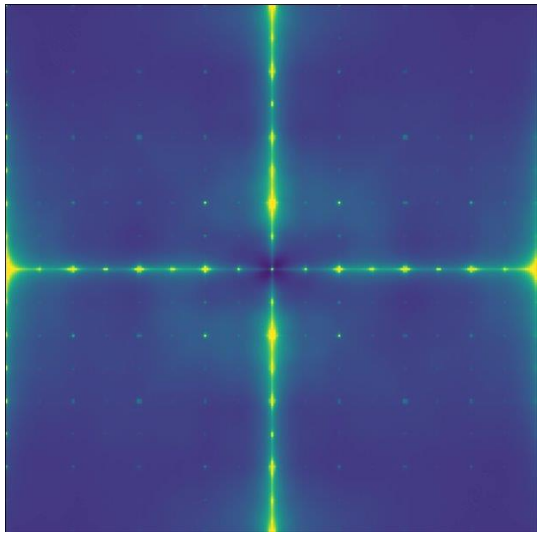
Research activity: Analysis

- Another major issue comes from perturbations commonly applied when uploading on social networks

- Compression reduces the strength of the artifacts and introduces peaks caused by the coding process



Before compression



After compression

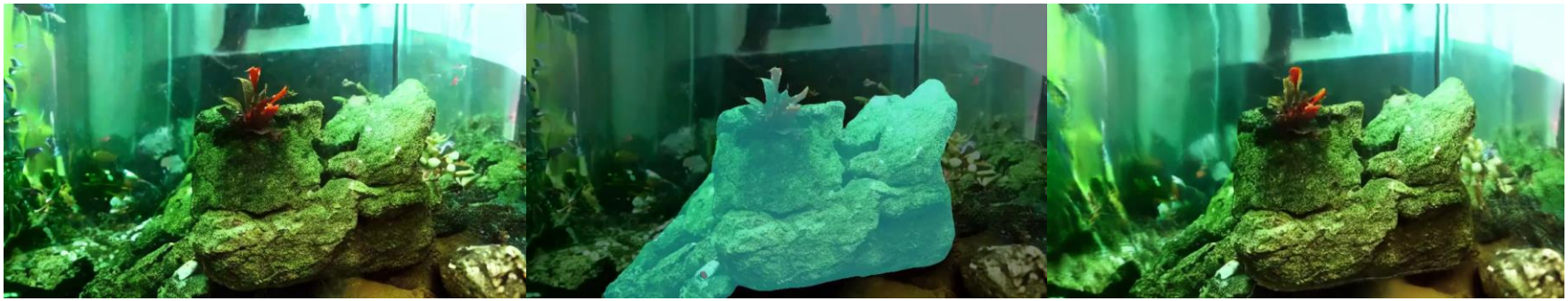
Research activity: Methodology

- Our approach: rely on high-level cues, which are less affected by post-processing operations
- We leverage 4D representations from a monocular video using 3D Gaussians

Research activity: Methodology

- We can observe inconsistencies in the reconstruction of the synthetic video
- The rock, despite being static in the synthetic video, morphs and changes shape in the reconstruction

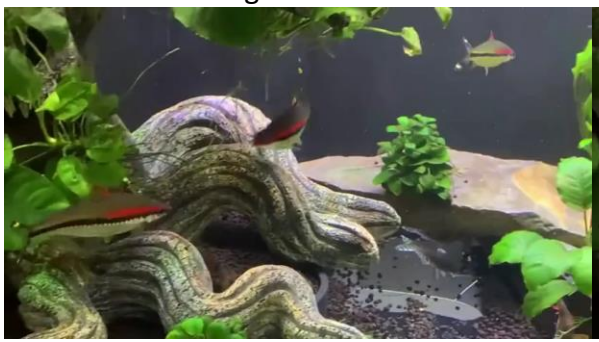
Synthetic Video | Segmented Video | Reconstructed Video



Research activity: Methodology

- Real videos are correctly reconstructed
- The tracked fishes are well reconstructed, even in their more complex movements

Original Video



Segmented Video

Reconstructed Video

Research products

[P1]	**Conference Paper** G. Zingarini, D. Cozzolino, **R. Corvi**, G. Poggi, L. Verdoliva, "M3Dsynth: A dataset of medical 3D images with AI-generated local manipulations", IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), *Seul, June 2024*
[P2]	**Workshop Paper** D. Cozzolino, G. Poggi, **R. Corvi**, M. Nießner, L. Verdoliva “Raising the Bar of AI-generated Image Detection with CLIP”, IEEE Workshop on Media Forensics at CVPR, *Seattle, June 2024*
[P3]	**Journal Paper** D. Tariang, **R. Corvi**, D. Cozzolino, G. Poggi, K. Nagano, L. Verdoliva, "Synthetic Image Verification in the Era of Generative Artificial Intelligence: What Works and What Isn’t There yet" IEEE Security & Privacy, May-June 2024

Next Year

- Carry out a large-scale quantitative analysis on the semantic clues related to 4D reconstruction
- Analyze the performance when videos are compressed/resampled
- Compare with state-of-the-art synthetic video detection approaches

Thank you for the attention!