UNIVERSITÀ DEGLI STUDI DI NAPOLI
FEDERICO II

itee PhD
information technology
electrical engineering

DIETI. UNINA

PhD Riccardo Corvi

Synthetic Image Detection

Tutor: Luisa Verdoliva

Cycle: XXXVIII

Year: First

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
- **Cooperation:** NVIDIA

Research field of interest

- **Multimedia Forensics**:
  - Development of techniques for the forensic analysis of images, audios and videos
- **Synthetic Image Detection**:
  - Identify if an image is AI generated or not
- **Source attribution**:
  - Trace the generative model that synthetized the image



Images generated using Midjourney (downloaded from Twitter)

Summary of study activities

	Courses	Seminars	Research	Tutorship
Total	21	5.6	31	0
Expected	20 - 40	5 - 10	10 - 35	0 – 1.6

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

- **PhD courses:**
  "Using Deep Learning Properly" (Dr. Andrea Apicella)
  "How to boost your PhD" (Prof. Antigone Marino)
  "Statistical Multimedia Security and Forensics"- University of Trento (Prof. Fernando Perez-Gonzalez)
- **MSc courses**:
  "Visione per Sistemi Robotici" (Dr. Davide Cozzolino)
- **Conference**:
  IEEE International Workshop on Information Forensics 2022 (Online), 13/12/2022 -16/12/2022
  IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Rhodes, 06/06/2023 - 10/06/2023

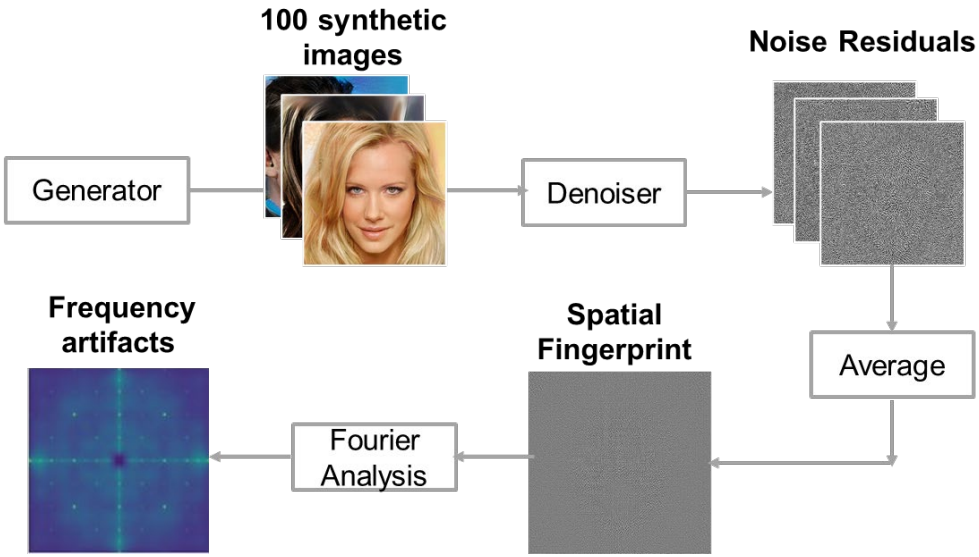
Research activity: Overview

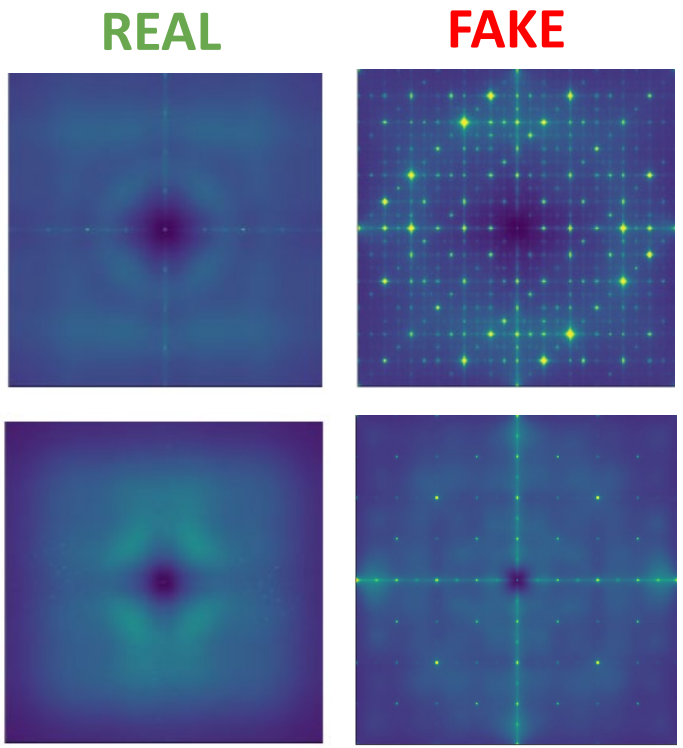
- Problem
  - Easy access to **generative AI** allows to easily spread disinformation over the web
  - The advent of new types of synthetic generators has led to **generalization** being one of the main challenge

- Objective
  - Analyze low level **forensic artifacts** hidden in synthetic images to gain insight into the most discriminative features
  - Develop a synthetic image detector that can **generalize** across different AI-based models

Research activity: Methodology

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





Research activity: Methodology

- We designed a deep learning-based detector that
  - aims at preserving such artifacts as much as possible
  - is robust to post-processing operations (compression, resizing)

- We trained on a single model to check for generalization

AUC	Train: Generative Adversarial Network (ProGAN)	Train: Diffusion Model (Latent DM)
Model 1 (GAN)	97.1	84.4
Model 2 (GAN)	94.9	95.6
Model 1 (DM)	72.8	85.9
Model 2 (DM)	91.6	93.1

Code available at: https://github.com/grip-unina/DMimageDetection (130 stars, 12 forks)

Products

[P1]	**Conference Paper** **R. Corvi**, D. Cozzolino, G. Zingarini, G. Poggi, K. Nagano, L. Verdoliva, "On the detection of synthetic images generated by diffusion models", *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),* Rhodes, June 2023 **Award:** Top 3% Paper Recognition
[P2]	**Workshop Paper** **R. Corvi**, D. Cozzolino, G. Poggi, K. Nagano, L. Verdoliva. "Intriguing properties of synthetic images: from generative adversarial networks to diffusion models", *IEEE Workshop on Media Forensics at CVPR,* Vancouver, June 2023

Next Year

- Develop an attribution method to identify the specific generative architecture (from a binary to N-ary classification task)
- Expand such analysis to an open set scenario where the number of models are not known in advance

Thank you for the attention!