



UNIVERSITÀ DEGLI STUDI DI NAPOLI  
**FEDERICO II**

**itee**<sup>PhD</sup>  
information technology  
electrical engineering



**DIE  
TI**

**UNI  
NA**

**Giada Zingarini**

# AI-generated local manipulation detection in images

Tutor: Luisa Verdoliva

Cycle: XXXVIII

Year: First

# My background

- **MSc degree** in Biomedical Engineering, curriculum in Biorobotic and Bionic – 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

# Research field of interest

- **Multimedia Forensics:**
  - Develop methods for the analysis of multimedia content for forensic applications such as fake image detection
- **AI-based local generation:**
  - Generation of manipulated data with AI methods such as text-to-image techniques
- **Image forgery localization:**
  - Identification of manipulated areas by providing a pixel-level localization map

«pizza»



«pink yarn ball»



# Summary of study activities

	Courses	Seminars	Research	Tutorship
Total	26	5.3	29.8	0
Expected	20 - 40	5 - 10	10 - 35	0 – 1.6

- Study of state-of-the-art text-driven technologies for image generation and detection
- **PhD School:**
  - “2023 IEEE SPS / EURASIP - Summer School on Metaverse Technologies, University of Cagliari ” - Award (First Prize Winner in the Team Competition)
- **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 Pérez-González)
- **MSc courses:**
  - “Elaborazione di segnali multimediali” (Prof. Luisa Verdoliva)
- **Conferences:**
  - International Workshop on Information Forensics 2022 (Online)

# Research activity: Overview

- Problem

- AI tools that generate local manipulations are very easy to use and can be maliciously used to spread **disinformation**
- It is even possible to **modify the content** of medical images, altering the resulting diagnoses

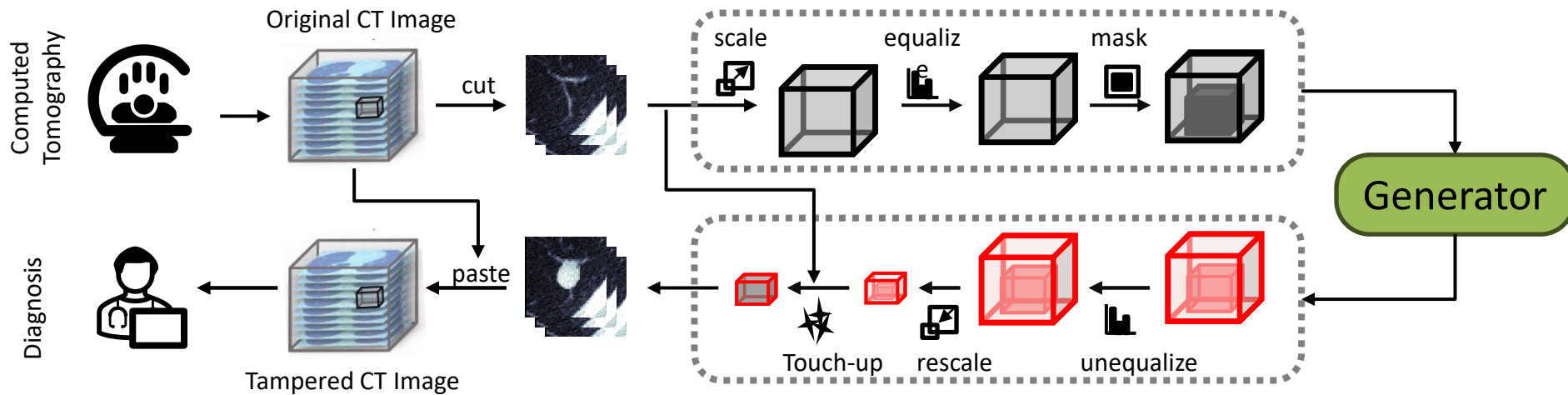
- Objective

- **Create a large dataset** of manipulated data to be used for training the detection methods
- Develop techniques for the **detection and localization** of local synthetic content for natural/biomedical images

# Research activity: Overview

- **Methodology**

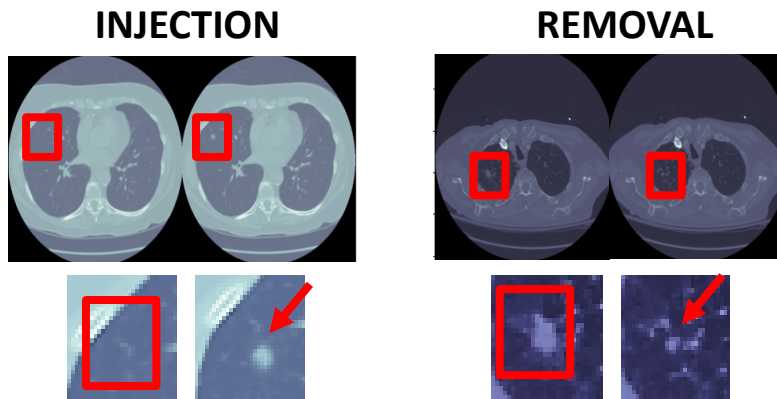
- We generated biomedical manipulated images by injecting or removing lung cancer nodules in real CT scans (around 8,577 samples)
- We used different generative architectures and adapted them for 3D biomedical images, e.g. Pix2Pix, CycleGAN, Diffusion Model (DM)



# Research activity: Overview

- **Methodology**

- We used a diagnostic tool to evaluate the quality of synthetic nodules
- Fake injected nodules have the same histogram as malignant pristine nodules and vice-versa
- The dataset, called **M3Dsynth**, helped to train effective approaches to detect and localize such manipulations



## Detection Accuracy

	Training Set	Test Set		
		Pix2Pix	CycleGAN	DM
G.P. images	ProGAN	50.0	47.1	48.8
	StyleGAN2	50.4	49.6	52.0
	LDM	44.6	44.5	46.2
M3Dsynth	Pix2Pix	99.5	96.6	95.8
	CycleGAN	97.7	98.5	91.6
	DM	96.1	92.8	97.3

# Products

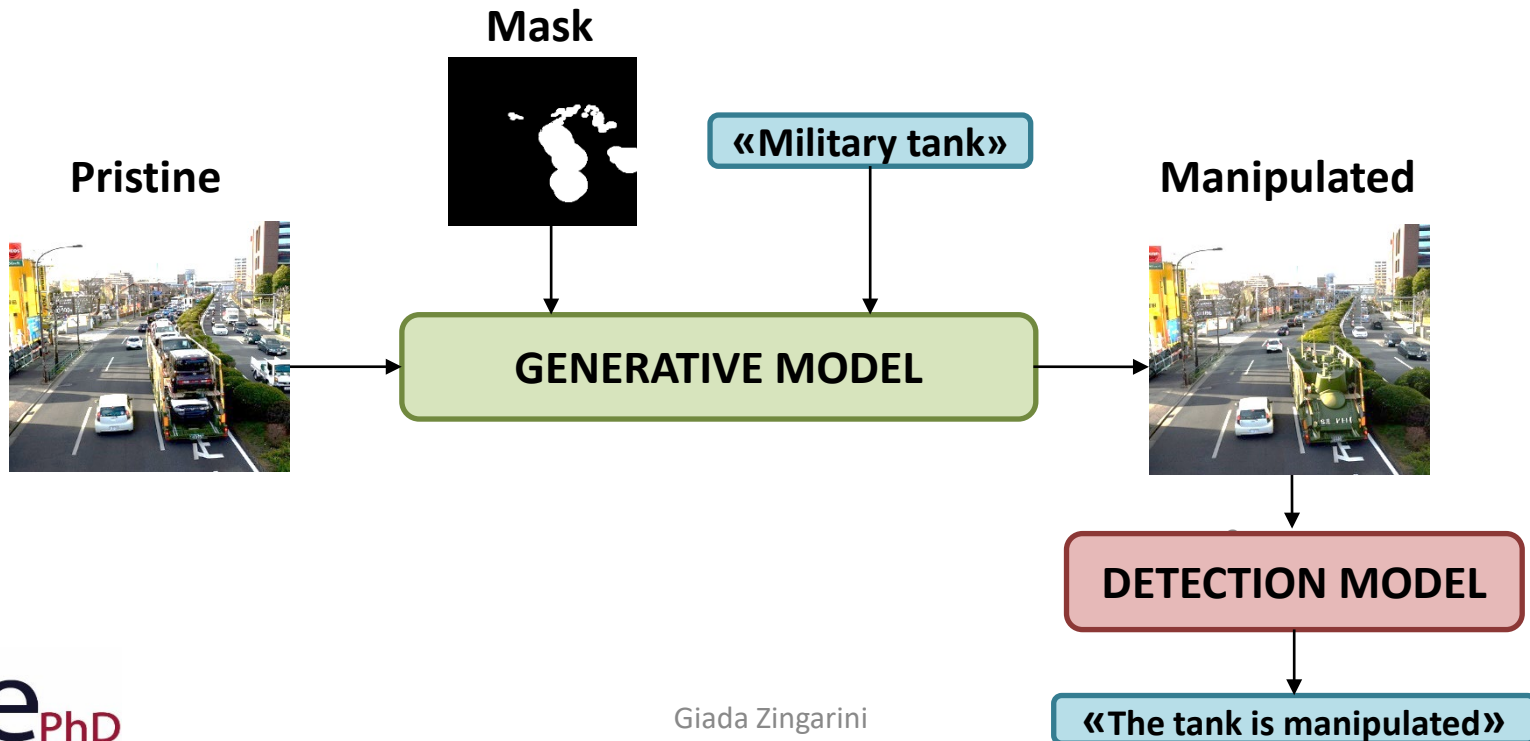
[P1]	<p><b>Conference Paper</b></p> <p>R. Corvi, D. Cozzolino, <b>G. Zingarini</b>, G. Poggi, K. Nagano, and L. Verdoliva, “On the detection of synthetic images generated by diffusion models”, in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2023, Rhodes.</p> <p><b>Award:</b> Top 3% Paper Recognition</p>
[P2]	<p><b>Conference Paper Submission</b></p> <p><b>G. Zingarini</b>, D. Cozzolino, R. Corvi, G. Poggi, L. Verdoliva: "M3Dsynth: A dataset of medical 3D images with AI-generated local manipulations", submitted to the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), arXiv preprint arXiv:2309.07973</p>



# Next year

- **Natural images**

- We plan to work on images locally manipulated using recent text-to-image approaches
- We want to develop a new method that works with paired image and text to spot the manipulations



Thank you for the attention!