PhD in Information Technology and Electrical Engineering
Università degli Studi di Napoli Federico II

PhD Student: Giada Zingarini

Cycle: XXXVIII

Training and Research Activities Report

Year: First

Giada Zingarini

Tutor: prof. Luisa Verdoliva Luisa Verdoliva

Date: October 31, 2023

1. Information:

- **PhD student: Giada Zingarini**
- **DR number: DR996629**
- **Date of birth: 20/02/1995**
- **Master Science degree: Biomedical Engineering University: Federico II**
- **Doctoral Cycle: XXXVIII**
- **Scholarship type:** ***UNINA - DII, DISCOVER project, funded by DARPA under the SEMAFOR program***
- **Tutor: Prof. Luisa Verdoliva**

2. Study and training activities:

Activity	Type[1]	Hours	Credits	Dates	Organizer	Certificate[2]
From Handcrafted to End-to-End Learning, and Back: a Journey for Multi-Object Tracking	Seminar	2	0.4	02/12/22	University of Modena and Reggio Emilia	Y
Digital Forensics	Seminar	2	0.4	06/12/22	DIETI - UNINA	Y
Face Representation Attack Detection	Seminar	1.5	0.3	07/12/22	IEEE Biometrics Council	Y
Advances on Multimodal Machine Learning Solutions for Speech Processing Tasks and Emotion Recognition	Seminar	1	0.2	19/01/23	IEEE Xplore	Y
Using Deep Learning Properly	Course	10	4	10/01/23 - 24/01/23	DIETI - UNINA	Y
The Super Neuron Model - A new generation of ANN-based Machine Learning and Applications	Seminar	1	0.2	09/02/23	EURASIP Journal on Image and Video Processing	Y
SPS Webinar: Human Centric Visual Analysis - Hand, Gesture, Pose, Action, and Beyond	Seminar	1	0.2	13/02/23	IEEE/SPS	Y

Algorithm Unrolling: Efficient, Interpretable Deep Learning for Signal and Image Processing	Seminar	1	0.2	14/02/23	DIETI - UNINA	Y
How to boost your PhD	Course	16	4	11/01/23-01/03/23	DIETI ITEE - ICTH - CQB PhD programs	Y
What's up with image & video forensics and security?	Seminar	1	0.2	02/03/23	EURASIP Journal on Image and Video Processing	Y
Unleashing the Power of LLMs: a Historical perspective on Generative AI	Seminar	1	0.2	02/03/23	DIETI - UNINA	Y
Nanoneuro: the power of nanoscience to explore the frontiers of neuroscience	Seminar	1	0.2	03/05/23	DIETI - UNINA	Y
Statistical Multimedia Security and Forensics	Course	20	4	08/05/23-12/05/23	University of Trento - IECS Doctoral School	Y
Computational Disinformation Symposium	Seminar	7	1.4	06/06/23	NYU Tandon School of Engineering	Y
Elaborazione di segnali multimediali	Course	72	9	06/03/23-09/06/23	DIETI - UNINA	Y
Scientific Integrity Verification Through Image Forensics	Seminar	1	0.2	06/07/23	SPS-IFS	Y
The Digital World: Artificial Intelligence	Seminar	6	1.2	13/07/23	British Standards Institution	Y
Summer School on Metaverse Technologies	Course	24	5	18/09/23-22/09/23	2023 IEEE SPS / EURASIP University of Cagliari	Y
Preparation of the conference paper “On	Research		9	01/11/22-31/12/22		N

the detection of synthetic images generated by diffusion models" for ICASSP 2023 **Attendance to weekly technical meetings** **Participation to International Workshop on Information Forensics (WIFS) 2022. Date: 13/12/2022 – 16/12/2022**						
Study of segmentation model based on Transformer architecture **Study of Deepfake video technique for video detection** **Preparation of the camera-ready for the accepted paper "On the detection of synthetic images generated by diffusion models" for ICASSP 2023** **Attendance to weekly technical meetings**	**Research**		**5.2**	**01/01/23-28/02/23**		**N**
Study of Language Models **Study of image-to-text architectures for captioning task** **Experiments using text-guided methods for image forgery detection**	**Research**		**5.6**	**01/03/23-30/04/23**		**N**

Experiments using Transformers architecture for forgery detection in medical field **Attendance to weekly technical meetings**						
Experiments using text-guided algorithms for captioning task **Attendance to weekly technical meetings**	**Research**		**2**	**01/05/23-30/06/23**		**N**
Experiments for the detection of synthetic manipulations in medical images **Experiments using text-guided algorithms for captioning task**	**Research**		**4**	**01/07/23-31/08/23**		**N**
Preparation of the conference paper “M3Dsynth: A dataset of medical 3D images with AI-generated local manipulations” for ICASSP 2024	**Research**		**5**	**01/09/23-31/10/23**		**N**

1) Courses, Seminar, Doctoral School, Research, Tutorship
2) Choose: Y or N

2.1. Study and training activities - credits earned

	Courses	Seminars	Research	Tutorship	Total
Bimonth 1	-	1.1	9	-	10.1
Bimonth 2	4	0.8	5.2	-	10.0
Bimonth 3	4	0.4	5.6	-	10.0
Bimonth 4	13	1.6	2	-	16.6
Bimonth 5	-	1.4	4	-	5.4
Bimonth 6	5	-	5	-	10.0
Total	26	5.3	30.8	0	62.1
Expected	30 – 70	10 - 30	80 - 140	0 – 4.8	

3. Research activity:

The ability to detect manipulated visual content is becoming increasingly important in many application fields, given the rapid advances in image synthesis methods. Powered by large language models (LLMs), text-to-image synthesis tools allow the user to create from scratch and modify images at will by means of simple text instructions. My research activity is focused on the development of methods for the detection of synthetic local manipulations in images. In my first year of PhD I worked on the detection of fully generated data from diffusion models [P1] and on the detection of AI-generated local manipulation in 3D medical images [P2]. In particular, most of my efforts this year were devoted to the latter task, which is described in the following.

Nowadays, the diagnosis of diseases relies heavily on noninvasive medical imaging techniques, such as Magnetic Resonance Imaging (MRI) and Computed Tomography (CT), which can produce high-resolution images of the body’s internal organs. 3D medical images are typically stored in secure Picture and Archive Communication System (PACS) servers. In [1], however, it was shown that an attacker could enter the system and use deep learning to modify medical CT scans, injecting or removing lung cancer nodules. Such actions may have the purpose of committing insurance fraud, falsifying scientific research data or even have political or terrorism-related purposes. Unfortunately, such manipulated images can easily fool automated cancer detectors and even medical experts. This motivates the research to develop methods that can localize and detect these manipulations. Despite that, very limited attention has been paid to detect tampering in medical images, and the question has been explored only focusing on the detection task and without examining the generalization ability of the detectors.

In this scenario, we worked to create M3Dsynth [P2], a large dataset of manipulated medical images, and then we conducted a benchmark to test the ability of state-of-the-art detectors to spot such manipulations. We employ three different methods to generate over 8,000 manipulated pulmonary CT-scans with the injection or removal of lung cancer nodules. We show that the manipulated images easily fool automated diagnosis tools. Moreover, we demonstrate in our preliminary study with several state-of-the-art detectors, that it is possible to detect and also localize both injections and removals even in a cross-generator scenario.

The pristine CT scans are from the dataset proposed in [2], comprising 1018 CT scans of 1010 patients, fully annotated with position and size of all detected nodules. In the manipulation pipeline, the same for all the methods, we modify only a local 3D cube of the CT scan. We use three different generative architectures, two based on GANs and one on DMs:

- **Pix2Pix**: We use the CT-GAN network proposed in [1], a 3D version of the Pix2Pix GAN proposed originally [3] for 2D images. We train two models of the same architecture, one for the injection and one for the removal task.
- **CycleGAN**: Inspired by [4,5], we adapted the networks to operate on 3D cubes, considering two translation tasks. For injection, the translation is between real cancerous tissues and the corresponding masked cubes, while for removal the translation is between samples without cancer and the corresponding masked cubes.

- **DM**: Finally, considering the strong impact diffusion models have had in the recent period, we use a 3D adaptation of DDPM [6,7] by replacing the original denoiser based on a 2D U-Net architecture with an analogous denoiser based on a 3D U-Net. We adopted the 3D model for our inpainting task, where the generated cube has to be coherent with the available masked cube.

To evaluate the quality of generated data, we used the computer-aided diagnostic tool proposed in [8]. This is a deep learning-based tool that comprises a detection network to localize the nodules and a classification network working on each of them. We applied only the classification network at the position where the nodule was injected or removed. The results show that before the manipulations, the diagnostic tool separates relatively well benign from malignant nodules, while after manipulation the removed malignant nodules have the same histogram as benign nodules had before manipulation and vice-versa.

The proposed dataset may be precious to assess new detectors, but it is even more important in the training phase. To demonstrate this point and provide further evidence of the role-specific dataset like M3Dsynth has to evaluate the ability of forensics instruments, we test the synthetic image detector proposed in [P1]. This method had shown brilliant results in terms of detecting fake images generated both with GANs and DMs, but it evidences poor results with the specific medical images of M3Dsynth dataset. Then, we repeated the analysis after fine tuning on the proposed dataset obtaining very different results. In fact, now the detection accuracy is always over 90% even when training and test images are generated with different methods.

For the benchmark evaluation, we selected six architectures [9,10,11,12,13,14] that could perform both localization and detection, and could be easily fine-tuned on our dataset. In fact, not all image forensics tools are appropriate, several classical approaches look for compression artifacts or traces of internal camera processing [15]. However, compression is not customary for CT images, and medical imaging sensors have very different properties than smartphones or general purpose cameras. To evaluate their generalization ability, methods are trained on images manipulated by a single synthetic generator and tested against all the others. Both detection and localization performances are very good on average, especially for Tru-For [14] and ManTra-Net [12], and the best results are for aligned data, but only a limited impairment is observed on non-aligned data, testifying of a good generalization ability.

References:

[1] Y. Mirsky et al., “CT-GAN: Malicious tampering of 3d medical imagery using deep learning,” in 28th USENIX Security Symposium, 2019.
[2] S. A. III et al., “The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans,” Medical physics, 2011.
[3] P. Isola et al., “Image-to-image translation with conditional adversarial networks,” in IEEE CVPR, 2017.
[4] D.Iommi,3D-CycleGan-Pytorch-Medical-Imaging-Translation, https://github.com/davidiommi/3D-CycleGan-Pytorch-MedImaging.

[5] N. Mangaokar et al., “Jekyll: Attacking medical image diagnostics using deep generative models,” in IEEE European Symposium on Security and Privacy (EuroS&P), 2020.

[6] Z. Dorjsembe et al., “Three dimensional medical image synthesis with denoising diffusion probabilistic models,” in Medical Imaging with Deep Learning, 2022.

[7] J. Ho et al., “Denoising diffusion probabilistic models,” Advances in neural information processing systems, vol. 33, pp. 6840–6851, 2020.

[8] F. Liao et al., “Evaluate the Malignancy of Pulmonary Nodules Using the 3-D Deep Leaky Noisy-OR Network,” IEEE Transactions on Neural Networks and Learning Systems, vol. 30, no. 11, pp. 3484–3495, 2019.

[9] F. Chollet, “Xception: Deep learning with depthwise separable convolutions,” in IEEE CVPR 2017.

[10] O. Ronneberger, et al., “U-net: Convolutional networks for biomedical image segmentation,” in MICCAI 2015.

[11] H. Li , et al., “Localization of deep inpainting using high-pass fully convolutional network,” in IEEE/CVF ICCV 2019.

[12] Y. Wu et al., “ManTra-Net: manipulation tracing network for detection and localization of image forgeries with anomalous features,” in IEEE/CVF CVPR 2019.

[13] X. Chen , et al. “Image Manipulation Detection by Multi-View Multi-Scale Supervision,” in IEEE/CVF ICCV 2021.

[14] F. Guillaro et al., “Trufor: Leveraging all-round clues for trustworthy image forgery detection and localization,” in IEEE CVPR 2023.

[15] L. Verdoliva, “Media Forensics and DeepFakes: an overview,” IEEE Journal of Selected Topics in Signal Processing, vol. 14, no. 5, pp. 910 – 932, 2020.

4. Research products

Publications:

*[P1] R. Corvi, D. Cozzolino, **G. Zingarini**, 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; Published, NOT yet indexed in Scopus.*

*[P2] **G. Zingarini**, D. Cozzolino, R. Corvi, G. Poggi, and L. Verdoliva, “M3Dsynth: A dataset of medical 3D images with AI-generated local manipulations”, arXiv preprint arXiv:2309.07973, 2023, Submitted to the IEEE International Conference on Acoustics, Speech and Signal Processing.*

Awards:

Top 3% Paper Recognition for the paper: "On the detection of synthetic images generated by diffusion models" , R. Corvi, D. Cozzolino, G. Zingarini, G. Poggi, K. Nagano, and L. Verdoliva at the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Rhodes, June 2023.

First Prize Winner in Team Competition at "2023 IEEE SPS / EURASIP - Summer School on Metaverse Technologies, University of Cagliari".

5. Conferences and seminars attended

Title: International Workshop on Information Forensics (WIFS) 2022
Date: 13-16/12/2022
Place: Online

6. Activity abroad:

None

7. Tutorship

None