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

PhD Student: Riccardo Corvi

Cycle: XXXVIII

Training and Research Activities Report

Year: First

Riccardo Corvi

Tutor: prof. Luisa Verdoliva

Luisa Verdoliva

Date: October 31, 2023

1. Information:

- **PhD student: Riccardo Corvi**
- **DR number: DR996631**
- **Date of birth: 03/02/1998**
- **Master Science degree: Computer Engineering University: Università Federico II di Napoli**
- **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]
Privacy and Data Protection	Seminar	2	0.4	22/11/22	DIETI-UNINA	Y
From Handcrafted to End-to-End Learning, and Back: a Journey for Multi-Object Tracking	Seminar	2	0.4	02/12/22	International AI Doctoral Academy and University of Modena and Reggio Emilia	Y
Digital Forensics	Seminar	2	0.4	06/12/22	DIETI-UNINA	Y
Face Presentation 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 Signal Processing Society	Y
Using Deep Learning Properly	Course	10	4	10/01/23 -24/01/23	DIETI-UNINA	Y
The Super Neuron Model - A new	Seminar	1	0.2	09/02/23	EURASIP Journal on	Y

generation of ANN-based Machine Learning and Applications					Image and Video Processing	
Human Centric Visual Analysis - Hand, Gesture, Pose, Action, and Beyond	Seminar	1	0.2	13/02/23	IEEE Signal Processing Society	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?	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
Statistical Multimedia Security and Forensics	Course	20	4	08/05/23-12/05/23	IECS Doctoral School-University of Trento	Y
Visione per Sistemi Robotici	Course	72	9	07/03/23 -09/06/23	DIETI-UNINA	Y
Self-supervised learning for robotic bin-picking	Seminar	1	0.2	06/07/23	EURASIP Journal on Image and Video Processing	Y
Scientific Integrity Verification Through Image Forensics	Seminar	1	0.2	06/07/23	IEEE SPS-IFS (Informati	Y

					on Forensics and Security Technical Committee)	
The Digital World: Artificial Intelligence	**Seminar**	**6**	**1.2**	**13/07/23**	**British Standards Institution**	**Y**
IEEE Authorship and Open Access Symposium: Tips and Best Practices to Get Published from IEEE Editors	**Seminar**	**1.5**	**0.3**	**20/09/23**	**IEEE Xplore**	**Y**
RICERCA E FORMAZIONE NELLA SOCIETÀ DELLA TRANSIZIONE DIGITALE	**Seminar**	**5**	**1**	**22/09/23**	**CINI - Consorzio Interuniversitario Nazionale per l'Informatica**	**Y**
Study of current literature in generative models and detectors of synthetic images **Preparation of the conference paper "On the detection of synthetic images generated by diffusion models" submitted at ICASSP 2023** **Attendance of weekly technical meetings** **Participation to the International**	**Research**		**8.6**	**01/11/22-31/12/22**		**N**

Workshop on Information Forensics (WIFS) 2022. Date: 13/12/2022 – 16/12/2022						
Study of the current literature in the analysis of the artifacts of synthetic Images **Preparation of the camera-ready version of the accepted conference 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 the state of the art methods for Large Language Models **Study of the state of the art methods for Synthetic Images Attribution** **Preparation of the workshop paper "Intriguing properties of synthetic images: from generative adversarial networks to diffusion models" for the Workshop on Media Forensics at CVPR 2023**	**Research**		**5.6**	**01/03/23-30/04/23**		**N**

Attendance to weekly technical meetings						
Participation to the Conference ICASSP 2023 in Rhodes, Greece from 06/06/2023 - 10/06/2023 **Presentation of the paper "On the detection of synthetic images generated by diffusion models" to the Conference ICASSP 2023 on the 08/06/2023** **Study of the state of the art methods for Synthetic Images Detection**	**Research**		**2**	**01/05/23-30/06/23**		**N**
Study of the state-of-the-art methods for Synthetic Image Attribution **Attendance to weekly technical meetings**	**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 **Experiments on synthetic image attribution both in closed-set and open-set scenario**	**Research**		**5.6**	**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	**-**	**1.5**	**8.6**	**-**	**10.1**
Bimonth 2	**4**	**0.8**	**5.2**	**-**	**10**
Bimonth 3	**4**	**0.4**	**5.6**	**-**	**10**
Bimonth 4	**13**	**0.0**	**2**	**-**	**15**
Bimonth 5	**-**	**1.6**	**4**	**-**	**5.6**
Bimonth 6	**-**	**1.3**	**5.6**	**-**	**6.9**
Total	**21**	**5.6**	**31**	**-**	**57.6**
Expected	**30 – 70**	**10 - 30**	**80 - 140**	**0 – 4.8**	

3. Research activity:

Synthetic media generation has seen tremendous progress in the span of just a few years, especially thanks to generative adversarial networks (GANs) and, more recently, diffusion models (DMs). These models are not only capable of generating images with extraordinary levels of photorealism, but they are also able to create images starting from textual descriptions. These powerful technologies offer numerous opportunities for many industries, from entertainment, to healthcare, to finance and manufacturing. At the same time, they can be used for all kinds of illicit purposes, especially to strengthen disinformation campaigns and political propaganda. This threat justifies the growing focus on automated tools that distinguish synthetic images from natural ones in the multimedia forensics community. During my first year of PhD, I focused my work on the detection of synthetic images and on the analysis of the forensic artifacts that are exploited by the approaches designed for their detection.

There are several types of architectures that generate high-quality images. The most important are:

- **Generative Adversarial Networks (GANs):** these techniques [1,2,3] focus on an adversarial approach training strategy based on a min-max game, where a network, the discriminator, has the objective of detecting synthetic images, while the other network, the generator, must produce synthetic images that evade detection by the discriminator.
- **Vector Quantized Generative Networks:** first introduced as Vector Quantized Variational Autoencoders (VQ-VAE), they rely on an encoder network to compress data in a low-dimensional latent space and a decoder to recover the original data from it. VQ-VAEs quantize the latent space using a discrete codebook, and use an autoregressive model based on a transformer architecture to model the images. VQ-GANs [4,5] are an extension of the model which introduces an adversarial loss.
- **Diffusion Models (DMs):** these recently proposed techniques [6,7] learn to generate synthetic samples by inverting an additive noise process. During training, clean images are converted to white noise through many small steps. A neural network is then trained to invert each step, reducing the amount of noise observed in input. After training, to generate an image, the model is fed a white noise field which is gradually denoised in successive steps until the desired clean output is obtained.
- **Latent Diffusion Models (LDMS):** these techniques [8,9] try to make the process of Diffusion Models more efficient by not operating directly in the pixel space but by first moving to a more compact representation using a Variational Auto Encoder. This last component learns a representation which is compact but also representative of the original image.

For synthetic image detection some works focus on the inability of the generators to replicate the high-level semantic features of natural images [10,11,12]. They rely on visible artifacts, such as chromatic anomalies or lack of natural symmetries. However, since the quality of the generators is rapidly improving, these types of artifacts will be less and less present. On the other hand, there are approaches which focus on low-level statistical discrepancies [13,14], which, while invisible to the naked eye, are related to the generative architectures. Many supervised detectors have been proposed [15,16,17,18] to identify GAN-generated images based on such traces. However, most state-of-the-art detectors fail to generalize to architectures not seen during training and suffer a drastic loss in performance when post-processing operations are applied to the images, like those used in social media.

My study in the first year of the PhD has focused on the detection of synthetic images and on the analysis of the generation artifacts.

In [P1] we studied how state-of-the-art detection models developed for GANs perform on the newest state-of-the-art generative architectures DMs. Firstly, we observed that the new DMs architectures generate their unique fingerprints and forensic traces just like GANs do. We also noted that, due to differences in the observed traces, detectors trained on GAN images perform poorly on DM images, and the opposite happens if architectures are retrained on DM images. We then investigated a solution to this problem by performing a fusion of detectors for GANs and DMs which also includes a calibration process.

In [P2] we proceeded to further explore the artifacts left on the images by the generative model. We perform analyses based on second-order statistics, more specifically we use the autocorrelation function in the spatial domain and the image power spectrum in the frequency domain. Firstly, we remove the high-level semantic content of the images, using a denoising filter. Then, to extract the source statistics, we average 1000 such residual images. Besides the expected forensic traces, we observed other interesting phenomena. In fact, we noted that a model can insert “alien” traces in the generated images that depend only on biases in the training dataset. Furthermore, we noted many post-processing operations may lead to a significant deterioration of the artifacts present in the images, which explains the significant loss in performance when post-processing is applied. Finally, we proved that current generators are not able to properly reproduce the spectral distribution of real images at mid-high frequencies, which could be exploited to detect synthetic images.

In [P3] we investigated the possibility of modifying the local content of 3D medical images. For this paper my contribution is limited to the experimental section, where I helped to test baseline synthetic detectors used in the benchmark analysis.

References

[1] T. Karras, M. Aittala, S. Laine, E. Härkönen, J. Hellsten, J. Lehtinen, and T. Aila. “Alias-free generative adversarial networks” NeurIPS, 34:852–863, 2021. 3, 4, 7

[2] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila. Analyzing and improving the image quality of StyleGAN. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2023, pages 8110–8119, 2020. 2, 3, 4

[3] M.Kang, J.-Y. Zhu, R. Zhang, J. Park, E. Shechtman, S. Paris and T. Park “Scaling up gans for text-to-image synthesis!” Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2023, pages 10124--10134

[4] P. Esser, R. Rombach, and B. Ommer. “Taming transformers for high-resolution image synthesis” In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2023, pages 12873-12883, 2021.

[5] B. Dayma, S. Patil, P. Cuenca, K. Saifullah, T. Abraham, P.Lê Khàc, L. Melas, and R. Ghosh. “DALL-E Mini” 7 2021.

[6] A. Ramesh, P. Dhariwal, A. Nichol, C. Chu, and M. Chen. "Hierarchical text-conditional image generation with clip latents" arXiv preprint arXiv:2204.06125v1, 2022.

[7] Y. Balaji, S. Nah, X. Huang, A. Vahdat, J. Song, K. Kreis, M. Aittala, T. Aila, S. Laine, B. Catanzaro, T. Karras, and M.Y. Liu. "ediffi: Text-to-image diffusion models with an ensemble of expert denoisers". arXiv preprint arXiv:2211.01324, 2022

[8] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer. "High-resolution image synthesis with latent diffusion models" In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2022 pages 10684–10695, .

[9] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B.Ommer. "Stable diffusion." https://github.com/CompVis/stable-diffusion, 2022.

[10] H. Guo, S. Hu, X. Wang, M.C. Chang, and S. Lyu. "Eyes tell all: Irregular pupil shapes reveal GAN-generated faces" In ICASSP, 2022. 2

[11] F. Matern, C. Riess, , and M. Stamminger. "Exploiting visual artifacts to expose deepfakes and face manipulations". In IEEE WACV Workshops, 2019.

[12] H. Farid. "Perspective (in)consistency of paint by text". arXiv preprint arXiv:2206.14617v1, 2022.

[13] F. Marra, D. Gragnaniello, L. Verdoliva, and G. Poggi, "Do GANs Leave Artificial Fingerprints?" in IEEE MIPR, 2019, pp. 506–511.

[14] N. Yu, L. Davis, and M. Fritz, "Attributing Fake Images to GANs: Learning and Analyzing GAN Fingerprints," in ICCV, 2019.

[15] S.-Y. Wang, O. Wang, R. Zhang, A. Owens, and A. Efros, "CNN-generated images are surprisingly easy to spot... for now," in CVPR, 2020.

[16] D. Gragnaniello, D. Cozzolino, F. Marra, G. Poggi, and L. Verdoliva, "Are GAN generated images easy to detect? A critical analysis of the state-of-the-art," in IEEE ICME, 2021.

[17] L. Chai, D. Bau, S.-N. Lim, and P. Isola, "What makes fake images detectable? Understanding properties that generalize," in ECCV, 2020.

[18] Y. Ju, S. Jia, L. Ke, H. Xue, K. Nagano, and S. Lyu, "Fusing Global and Local Features for Generalized AI-Synthesized Image Detection," in IEEE ICIP, 2022.

4. Research products

Publications

[P1] ***R. Corvi****, D. Cozzolino, G. Zingarini, G. Poggi, K. Nagano, 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 indexed in Scopus yet

[P2] ***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 2023, Vancouver, Published

[P3] G. Zingarini, D. Cozzolino, ***R. Corvi****, G. Poggi, L. Verdoliva*
"M3Dsynth: A dataset of medical 3D images with AI-generated local manipulations", Submitted
arXiv preprint arXiv:2309.07973

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.

5. Conferences and seminars attended

IEEE International Workshop on Information Forensics and Security (WIFS) 2022
- Dates: 13/12/2022 -16/12/2022
- Place: Online

IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 2023
- Dates: 06/06/2023 - 10/06/2023
- Place: Rhodes, South Aegean, Greece
- poster presentation of the paper "On the detection of synthetic images generated by diffusion models" on the 08/06/2023

6. Activity abroad:

None

7. Tutorship

None