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

PhD Student: Fabrizio Guillaro

Cycle: XXXVII

Training and Research Activities Report

Year: First

Tutor: prof. Luisa Verdoliva

Co-Tutor: Giovanni Poggi

Date: October 31, 2022

1. Information:

- **PhD student: Fabrizio Guillaro**
- **DR number: DR995863**
- **Date of birth: 06/01/1996**
- **Master Science degree: Computer Engineering**
- **University: Università degli Studi di Napoli Federico II**
- **Doctoral Cycle: XXXVII**
- **Scholarship type: UNINA - DII, DISCOVER project, funded by DARPA under the SEMAFOR program**
- **Tutor: Prof. Luisa Verdoliva**
- **Co-tutor: Prof. Giovanni Poggi**

2. Study and training activities:

Activity	Type[1]	Hours	Credits	Dates	Organizer	Certificate[2]
Single cell omics leverage Machine Learning to dissect tumor microenvironment and cancer immuno editing	Seminar	2	0.4	02/12/21	Prof. Anna Corazza	Y
Threat Hunting Essentials	Seminar	2	0.4	03/12/21	Dr. R. Natella, Prof. D. Cotroneo, Prof. S.P. Romano	Y
Physiological forensics	Seminar	3	0.6	07/12/21	Prof. William Puech (IEEE SPS-IFS)	Y
Is Machine Learning Security IFS-business as usual?	Seminar	1	0.2	08/12/21	Prof. Rémi Cogranne	Y
Multimedia data recovery and its related workflows in digital forensics	Seminar	1.5	0.3	09/12/21	Prof. Alessandro Piva	Y
GDPR basics for computer scientists	Seminar	2	0.4	14/12/21	Prof. Piero Bonatti	Y
The Deepfake Threat	Seminar	1	0.2	11/01/22	Matthew	Y

					Turek (DARPA)	
Intelligenza artificiale e sistemi d'arma autonomi	Seminar	1.5	0.3	19/01/22	Diego Latella (CNR-IBF. GI-STS)	Y
Federated Learning in Big Data over Networks	Seminar	1	0.2	17/02/22	Geert Leus (IEEE SPS)	Y
Towards Trustworthy AI - Integrating Reasoning and Learning	Seminar	1	0.2	22/02/22	M. Chetouani (International AI Doctoral Academy (AIDA))	Y
Visual Data Analysis: Why? When? How?	Seminar	1	0.2	10/03/22	Association for Computing Machinery (ACM)	Y
Apprendimento Automatico per la Visione Artificiale	Seminar	4	0.8	10/03/22	Prof. Vito Roberto (Università di Udine)	Y
Apprendimento Automatico per la Visione Artificiale	Seminar	4	0.8	17/03/22	Prof. Vito Roberto (Università di Udine)	Y
Apprendimento Automatico per la Visione Artificiale	Seminar	4	0.8	24/03/22	Prof. Vito Roberto (Università di Udine)	Y
Apprendimento Automatico per la Visione Artificiale	Seminar	4	0.8	31/03/22	Prof. Vito Roberto (Università di Udine)	Y
Apprendimento Automatico per la Visione Artificiale	Seminar	4	0.8	07/04/22	Prof. Vito Roberto (Università di Udine)	Y
Apprendimento Automatico per la Visione Artificiale	Seminar	4	0.8	21/04/22	Prof. Vito Roberto (Università di Udine)	Y
Apprendimento Automatico per la Visione Artificiale	Seminar	4	0.8	29/04/22	Prof. Vito Roberto (Università di Udine)	Y

IEEE Authorship and Open Access Symposium: Tips and Best Practices to Get Published from IEEE	Seminar	1.5	0.3	30/03/22	IEEE Association	Y
An Introduction to Deep Learning for Natural Language Processing	Seminar	1	0.2	13/04/22	Prof. Francesco Cutugno	Y
Explainable Natural Language Inference	Seminar	1.5	0.3	13/04/22	Prof. Francesco Cutugno	Y
Introduction to Deep Learning	Course	24	6	14/03/22 – 27/04/22	Giovanni Poggi, Diego Gragnaniello	Y
Elaborazione di Segnali Multimediali	Tutorship	16	0.64	01/03/22 – 30/04/22	Luisa Verdoliva	N
Deepfake Detection: Humans vs. Machines	Seminar	1	0.2	08/06/22	IEEE SPS Information Forensics and Security Technical Committee	Y
Visione per Sistemi Robotici	Course	72	9	07/03/22 – 07/06/22	Giovanni Poggi	Y
Image and Video Processing for Autonomous Driving	Course	48	6	07/03/22 – 07/06/22	Luisa Verdoliva	Y
Elaborazione di Segnali Multimediali	Tutorship	16	0.64	01/05/22 – 30/06/22	Luisa Verdoliva	N
Interactive Art and Pattern Recognition	Seminar	3	0.6	21/08/22	Larry O'Gorman - ICPR 2022	Y
DeepLearn 2022 Summer - 6th Internation Gran Canaria School on Deep Learning	Doctoral school	40	5	25/07/22 – 29/07/22	Prof. Marisol Izquierdo, Prof. Carlos Martìn-Vide	Y
Towards Actionable	Seminar	1	0.2	27/09/22	International	Y

XAI					AI Doctoral Academy (AIDA)	
Study of the state-of-the-art deep learning methods for image forgery detection and localization. Implementation of an improved version of the noiseprint based approach. Attendance to weekly technical meetings with Google.	Research		6	01/11/21 – 31/12/21		N
Study of the state-of-the-art deep learning methods for GAN generated image detection. Study of the state-of-the-art deep learning methods for image forgery detection and localization. Attendance to weekly technical meetings with Google. Attendance to weekly technical meetings for DARPA's SemaFor program.	Research		6	01/01/22 – 28/02/22		N
Study of the state-of-the-art deep learning methods for image forgery detection and localization, image generation with diffusion models and GANs.	Research		1	01/03/22 – 30/04/22		N

Attendance to weekly technical meetings with Google. Attendance to weekly technical meetings for DARPA's SemaFor program.						
Attendance to weekly technical meetings with Google. Attendance to weekly technical meetings for DARPA's SemaFor program. Preparation of the paper "Comprint: Image Forgery Detection and Localization using Compression Fingerprints" for the Workshop on MultiMedia FORensics in the WILD	Research		1	01/05/22 – 30/06/22		N
Participation to the IEEE International Conference on Pattern Recognition (ICPR 2022). Date: 21/08/2022 – 25/08/2022 Study of deep learning methods for image forgery detection and localization	Research		3	01/07/22 – 31/08/22		N
Study of deep learning methods for image forgery detection and localization. Study of deep learning	Research		6	01/09/22 – 31/10/22		N

methods for failure prediction and confidence estimation. Preparation of a conference paper about cross-modal fusion and confidence estimation for image forgery localization. Attendance to weekly technical meetings with Google. Attendance to weekly technical meetings for DARPA's SemaFor program.						

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	-	**2.3**	**6**	-	**8.3**
Bimonth 2	-	**0.9**	**6**	-	**6.9**
Bimonth 3	**6**	**6.6**	**1**	**0.64**	**14.2**
Bimonth 4	**15**	**0.2**	**1**	**0.64**	**16.8**
Bimonth 5	**5**	**0.6**	**3**	-	**8.6**
Bimonth 6	-	**0.2**	**6**	-	**6.2**
Total	**26**	**10.8**	**23**	**1.28**	**61.1**
Expected	**30 - 70**	**10 - 30**	**80 - 140**	**0 – 4.8**	

3. Research activity:

In the last years, manipulation of multimedia content is becoming easier and more realistic thanks to the growing popularity of smart and easy-to-use editing tools. The easy access to these tools, together with the power of social networks, makes it really simple to create and spread misinformation over the Internet. Hence it is important to develop tools able to verify the authenticity of shared media as well as localize potential forgeries.

Although misinformation can be carried out using different types of media, such as video, audio, text or image, in my first year of PhD I focused my research activities on digital image forensics tasks, especially **image forgery detection**, which consists in providing an image-level score that indicates the

probability of the image being manipulated, and **image forgery localization**, which provides a pixel-level localization map of the forgeries. These are strongly related tasks, since the mere detection of a manipulation is often not as satisfactory as locating specific forged areas .

Image manipulation can involve deepfake or cheapfake manipulations, depending on the tools used to accomplish the tampering:

- **Deepfake** is a term used to describe the synthesis of either a part or the entirety of an image (or other media) by means of sophisticated AI-based software. Through deep learning, extremely realistic results can be obtained, sometimes requiring the least amount of effort by the user, which can simply use text prompts to add something in an image or even just generate a fake, non-existing person's face. Generative Adversarial Networks [15, 16] and the more recent Diffusion Models [17, 18, 19] are the most common techniques used to generate synthetic imagery.
- Any manipulation that is not a deepfake, can be referred to as **cheapfake**, that is the traditional way to edit existing images. Photoshop, GIMP, Snapseed or even the built-in camera app of a smartphone are some of the programs that can be used to alter an image before distributing it on social media. Currently, this is the most common type of manipulation that can be found on the web, but the situation is probably going to reverse in the next decades, given the easier way to obtain realistic results with AI.

Manipulations can also be classified as:

- **Splicing**: it consists in copying a region from a source image and pasting it into a host image;
- **Copy-move**: similar to splicing, but the spliced area comes from the very same image;
- **Inpainting**: also known as removal, it consists in drawing on top of a region to hide some content, usually replacing it with the background. Lately, the term "inpainting" is also used to refer to splicing of objects by means of deep learning techniques, due to the fact that the object does not come from a real host image.

Image forgeries leave some imperceptible traces on the image, invisible to the naked eye but detectable by either conventional methods or more recent data-driven deep learning approaches. That is because every step in the camera acquisition process generates a specific noise, which is then disrupted after alterations of some sort. However, the uploading of images on social media makes the detection of manipulation more challenging, due to re-compression and other processing performed by the social network.

Existing techniques exploit these traces, such as analyzing JPEG artifacts introduced by double quantization [2, 3, 4] or disruption of the CFA pattern [5, 7]. While conventional methods [2, 5, 6] accomplish this by analyzing solely the image under test, deep learning techniques [3, 4, 7, 1, 8, 9, 10, 11, 12] use a large amount of data to learn how to detect specific or generic forensics clues.
In general, they work on the noise residuals [1, 4], obtained typically through high-pass filtering or denoising. In fact, traces and inconsistencies can be better highlighted by suppressing the scene content of the image. Some methods adopt a one-class learning approach [1, 12], which has the advantage of requiring only authentic images during training, treating manipulations as anomalies. While most of the deep learning methods rely on Convolutional Neural Networks, more recent techniques adopt the more recent architectures, such as the popular Transformers [8, 11].

My study through the first year of PhD mainly focused on the localization of cheapfakes, however it is important to note that the proposed approach can be also applied to the detection of local synthetic content.
We have developed a forgery localization method [P1] which aims at extracting a compression fingerprint, namely **Comprint**, from an image. Inconsistencies in the comprint are then found by means of an expectation-maximization algorithm. If two regions of the image underwent different compression, it is a hint of a forgery. This is a weaker assumption than the one made in most state-of-the-art works, which is that pristine regions underwent double compression in contrast to forged regions that underwent single compression. This should aid generalization in the wild.
The work is based on [1], which uses a siamese network in order to extract a noise residual, called **Noiseprint**, from pristine images using a contrastive learning approach to encourage the network to extract the same fingerprint from images coming from the same camera model. Therefore, the noiseprint acts like a fingerprint of the camera model that took the image.
However, our architecture is trained with the purpose of obtaining the fingerprint of compression artifacts, rather than the fingerprint of the camera model. Hence, the network is forced to extract the same fingerprint from images having the same type and quality of compression.
We also experimented that a fusion of Comprint and Noiseprint performs better than the two algorithms individually.

In addition, we are developing another forgery detection and localization technique which extends Noiseprint [1] by implementing a training strategy that includes the **editing histories** of an image and provides a better localization map. In particular, we refer to "editing histories" as the sequence of in-camera and out-camera processing that the image underwent. Images coming from the same device and that have the same editing history should have the same noiseprint.
Additionally, we experimented better ways to obtain the output heatmap, treating the localization problem as a binary segmentation task. To this end, we used a **cross-modal** transformer architecture [13] which uses both the RGB data and the noiseprint of an image as input, hence exploiting both the high-level forensic clues highlighted by the noiseprint and the low-level semantic features embedded in the RGB domain.
Eventually, we also implemented a **confidence estimator** in order to provide a confidence map as an additional output, based on [14]. This can be useful because it helps the user to understand where the network fails and which areas of the heatmap are less reliable. This can help to further improve the localization performance .

References:

[1] Noiseprint: D. Cozzolino and L. Verdoliva, "Noiseprint: A CNN-Based Camera Model Fingerprint," *IEEE Transactions on Information Forensics and Security*, vol. PP, pp. 1–1, 05 2019.

[2] T. Bianchi, A. De Rosa, and A. Piva, "Improved DCT coefficient analysis for forgery localization in jpeg images," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2011, pp. 2444–2447.

[3] Q. Wang and R. Zhang, "Double jpeg compression forensics based on a convolutional neural network," *EURASIP Journal on Information Security*, vol. 2016, pp. 1–12, 2016.

[4] M. Barni, L. Bondi, N. Bonettini, P. Bestagini, A. Costanzo, M. Maggini, B. Tondi, and S. Tubaro, "Aligned and non-aligned double jpeg detection using convolutional neural networks," *J. Vis. Commun. Image Represent.*, vol. 49, pp. 153–163, 2017.

[5] A. E. Dirik and N. D. Memon, "Image tamper detection based on demosaicing artifacts," *2009 16th IEEE International Conference on Image Processing (ICIP)*, pp. 1497–1500, 2009.

[6] D. Cozzolino, G. Poggi, and L. Verdoliva, "Splicebuster: A new blind image splicing detector," 2015 *IEEE International Workshop on Information Forensics and Security (WIFS)*, pp. 1–6, 2015.

[7] Q. Bammey, R. G. von Gioi, and J.-M. Morel, "An adaptive neural network for unsupervised mosaic consistency analysis in image forensics," *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 14 182–14 192, 2020.

[8] Hao, Jing, et al. "TransForensics: image forgery localization with dense self-attention." *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021.

[9] H. Li, C. Yang, F. Lin, B. Jiang, and H.-J. Zhao, "Constrained R-CNN: A general image manipulation detection model," *2020 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1–6, 2020.

[10] Kwon, Myung-Joon, et al. "Learning JPEG Compression Artifacts for Image Manipulation Detection and Localization." *International Journal of Computer Vision* (2022): 1-21.

[11] Wang, Junke, et al. "Objectformer for image manipulation detection and localization." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022.

[12] M. Huh, A. Liu, A. Owens, and A.A. Efros. "Fighting fake news: Image splice detection via learned self-consistency." In *European Conference on Computer Vision (ECCV)*, 2018.

[13] Liu, Huayao, et al. "CMX: Cross-Modal Fusion for RGB-X Semantic Segmentation with Transformers." *arXiv preprint arXiv:2203.04838 (2022)*.

[14] Corbière, Charles, et al. "Addressing failure prediction by learning model confidence." *Advances in Neural Information Processing Systems 32* (2019).

[15] Karras, Tero, et al. "Analyzing and improving the image quality of stylegan." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2020.

[16] Brock, Andrew, Jeff Donahue, and Karen Simonyan. "Large scale GAN training for high fidelity natural image synthesis." *arXiv preprint arXiv:1809.11096* (2018).

[17] Ramesh, Aditya, et al. "Hierarchical text-conditional image generation with clip latents." *arXiv preprint arXiv:2204.06125* (2022).

[18] Rombach, Robin, et al. "High-resolution image synthesis with latent diffusion models." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022.

[19] Nichol, Alex, et al. "Glide: Towards photorealistic image generation and editing with text-guided diffusion models." *arXiv preprint arXiv:2112.10741* (2021).

4. Research products:

Conference Papers:

[P1] H. Mareen, D. Vanden Bussche, **F. Guillaro**, D. Cozzolino, G. Van Wallendael, P. Lambert, L. Verdoliva, "Comprint: Image Forgery Detection and Localization using Compression Fingerprints," in Proceedings of the International Conference on Pattern Recognition (ICPR) 2022, Montréal; *Published, NOT yet indexed in Scopus*

5. Conferences and seminars attended

International Conference on Pattern Recognition (ICPR 2022)

- Dates: 21/08/2022 – 25/08/2022
- Place: Montréal, Québec, Canada
- Co-author of the paper “Comprint: Image Forgery Detection and Localization using Compression Fingerprints”, published in the **Workshop on MultiMedia FORensics in the WILD (MMForWILD)**

6. Activity abroad:

None

7. Tutorship

- Teaching assistance for the course “Elaborazione di Segnali Multimediali” (32 hours)