Deep Fake
La nuova frontiera del Falso d'autore

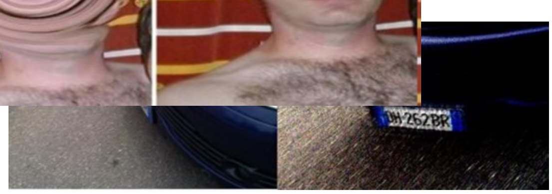
Prof. Sebastiano Battiato
www.dmi.unict.it/~battiato
battiato@dmi.unict.it

Dipartimento di Matematica e Informatica
University of Catania, Italy

1

Two words on myself

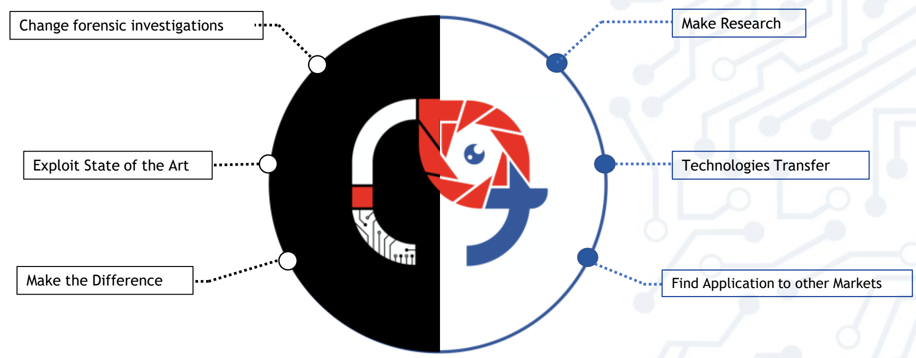
- Full Professor of Computer Science
  - Deputy Rector for Strategic Planning and Information Systems
  - Scientific Coord. PhD Program in Computer Science
  - Research and directorship of the IPLab research lab
- KeyWords: Multimedia, Computer Vision and….



2

The two-fold reality of iCTLab

Digital Forensics



Digital Forensics & CV R&D

UNIVERSITÀ degli STUDI di CATANIA

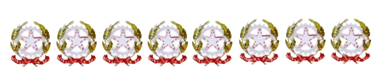
iCtLab Digital Forensics

3

Number in Forensics

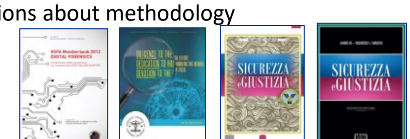
Digital Forensics

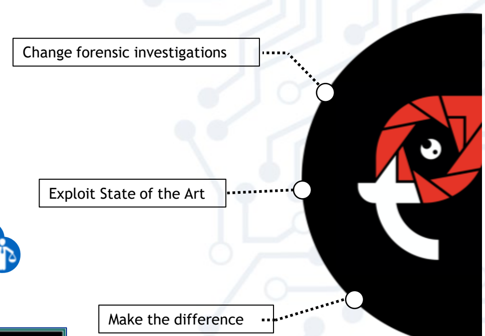
Continuous collaboration with 8 LEAs



About 30-35 cases per year

Publications about methodology





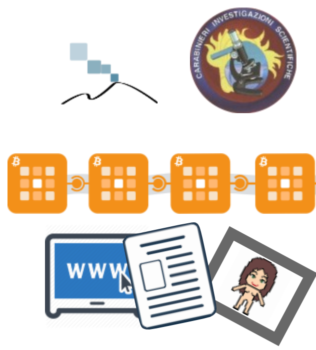
UNIVERSITÀ degli STUDI di CATANIA

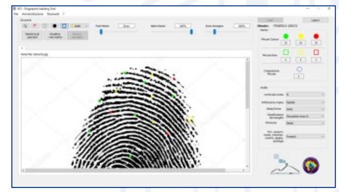
iCtLab Digital Forensics

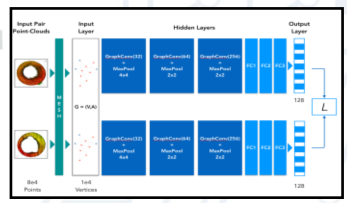
4

in the meanwhile…

Make Research







UNIVERSITÀ degli STUDI di CATANIA

ICtLab Digital Forensics

5

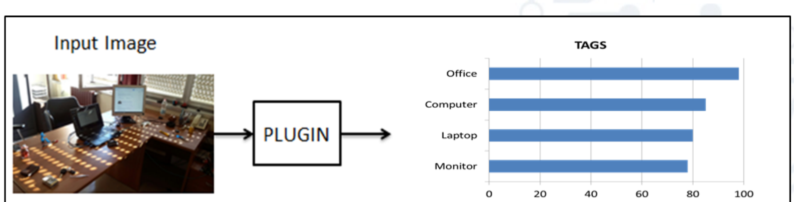
in the meanwhile…

Make Research

Tech. Transfer

DSINT ARTIFICIAL INTELLIGENCE AND DATA FUSION

CY4GATE



UNIVERSITÀ degli STUDI di CATANIA

ICtLab Digital Forensics

6

Other

- Education/Training (Advanced)
  - CyberSecurity
  - GDPR
  - Video forensics
  - Digital Forensics
  - Computer Vision applied on Forensics problems
  - Computer Vision solution for Developer
- DPO Consultancy
- POC Analysis
- Support for R&D projects

UNIVERSITÀ degli STUDI di CATANIA

iCtLab Digital Forensics

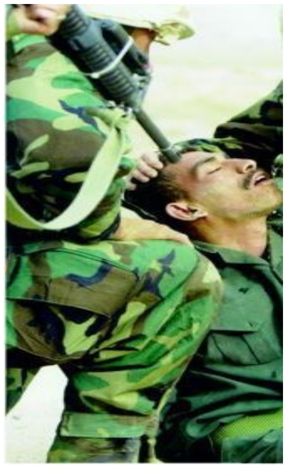
7

9

Seeing isn't believing

10

11



12

13

14

Multimedia Forensics is based on the idea that inherent **traces** (like digital fingerprints) are left behind in a digital media during both the **creation** phase and any other successively process.

15

Camera Ballistics

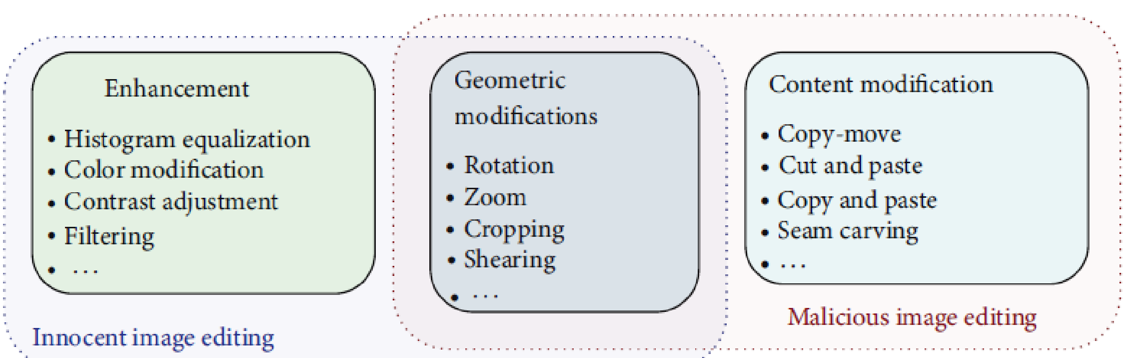
Which Device Has Created This Picture?

- Example:
  - Forensic analysis of a smartphone: which pictures have been generated on the device and which ones have been generated by other devices and sent by messaging application or saved from the internet

- We can identify:
  - Type of device
  - Maker and model
  - Specific exemplar

16

Image Editing



Piva 2013

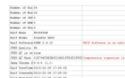
- Malicious image editing alters the image semantic content:
  - **Adding** information
  - **Removing** information

17

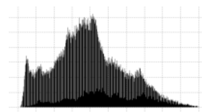
How To Authenticate An Image?

- Visual Inspection

- File Analysis
  - File Format and Structures
  - Metadata (EXIF)
  - Compression Parameters (Quantization Tables)



- Global Analysis
  - Pixel and compressed data statistics

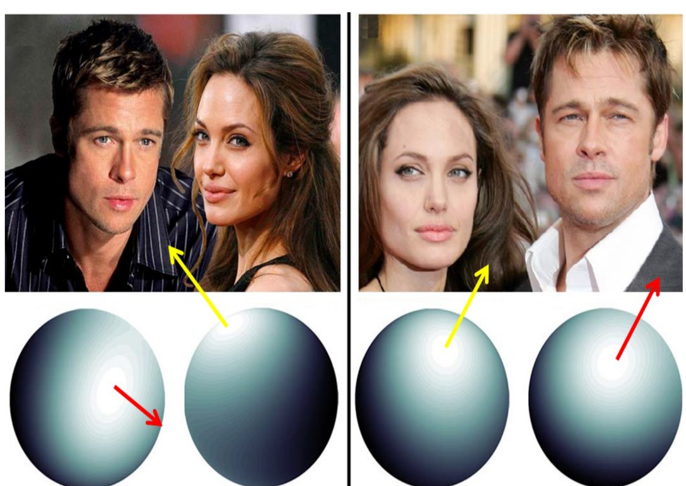


- Local Analysis
  - Finding inconsistencies of pixel statistics across the image

18

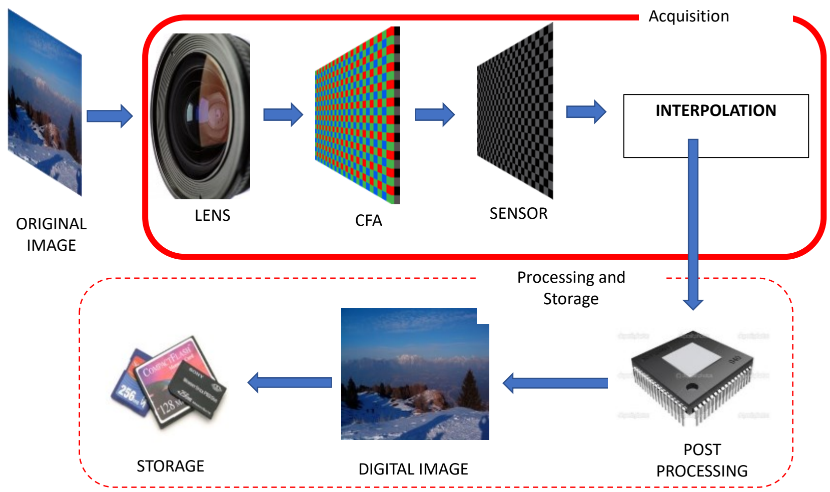
Physics-Based

Lighting inconsistencies can used for revealing traces of digital tampering.



19

Camera-Based



20

Social (Multimedia) Forensics

- Uploading an image on a Social Network
  - The process alters images

- Resize
- Rename
- Meta-Data deletion/editing
- Re-Compression
- NEW JPEG file Structure

M. Moltisanti, A. Paratore, S. Battiato, L. Saravo - ***Image Manipulation on Facebook for Forensics Evidence*** – ICIAP 2015, LNCS 2015;
O. Giudice, A. Paratore, M. Moltisanti, S. Battiato - ***A Classification Engine for Image Ballistics of Social Data*** – (Arxiv 2016 No. 1699257) http://arxiv.org/abs/1610.06347

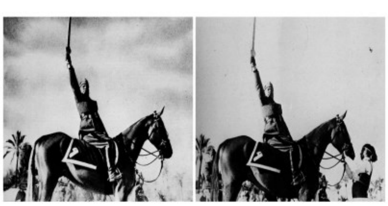
21

Huawei (2016)

22

Who Cares?

geopolitics…



…and political propaganda

23

I cacciatori di bufale digitali: «Così staniamo i falsi» - CorriereTV (2017)

24

25

These people may look familiar, like ones you've seen on Facebook or Twitter.

26

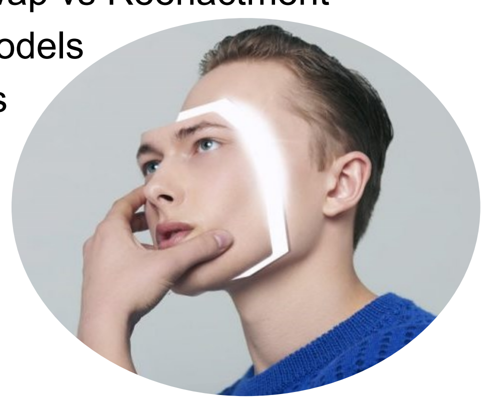
27

English AI Anchor

28

Outline

- Introduction
- Taxonomy: Face Swap vs Reenactment
- Deep Generative Models
- Detection Strategies
- Challenges
- Future Evolution
- References



29

What are DeepFake?

DeepFakes refers to all those multimedia contents synthetically altered or created by exploiting machine learning generative models.

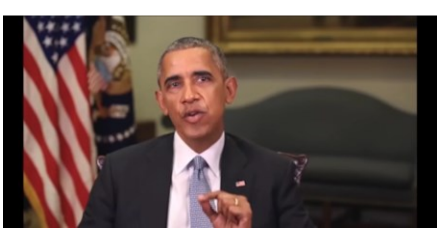
DeepFakes are image, audio or video contents that appear extremely realistic to humans specifically when they are used to generate and/or alter/swap image of faces.

30

Deepfake in the world

Other more worrying examples are the video of Obama **(a)**, created by **Buzzfeed** in collaboration with ***Monkeypaw Studios***, or the video in which Mark Zuckerberg **(b)** claims a series of statements about the platform's ability to steal its users' data.

Striscia la Notizia (September 2019), showed a video of the ex-premier Matteo Renzi **(c)** saying a series of ironic and not very "respectful" statements against his colleagues



(a)

(b)

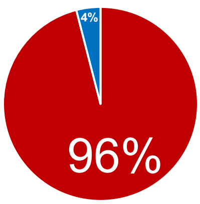
(c)

31

DeepFake

DeepFake have serious repercussions on the truthfulness of the information **spread through mass media** and represent a new threat to the world of **politics, companies** and **personal privacy**.

As many as **96% of DeepFake videos are porn** (**deep porn**) while only the remaining **4% are of another kind**. Deep fakes are evolving quickly and are becoming dangerous, not just for the reputation of the victims but also for security.



The State of Deepfakes: Landscape, Threats, and Impact, Henry Ajder, Giorgio Patrini, Francesco Cavalli, and Laurence Cullen, September 2019.

ICtLab Digital Forensics

32

DeepNude App

ICtLab Digital Forensics

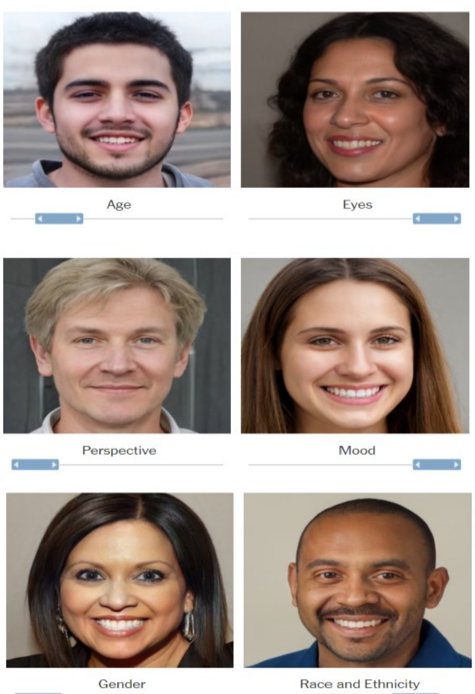
33

Sell Fake People

On the website **Generated.Photos**, you can buy a “unique, worry-free” fake person for $2.99, or 1,000 people for $1,000. If you just need a couple of fake people — for characters in a video game, or to make your company website appear more diverse — you can get their photos for free on **ThisPersonDoesNotExist.com**. Adjust their likeness as needed; make them old or young or the ethnicity of your choosing. If you want your fake person animated, a company called **Rosebud.AI** can do that and can even make them talk.

Designed to Deceive: Do These People Look Real to You? - NYTimes - Nov. 2020

34



Where

APP:
Impressions
Zao
Reface
Faceapp
DoubliCat

OTHER:

https://generated.photos/faces

https://www.rosebud.ai/

35

Generative Adversarial Nets

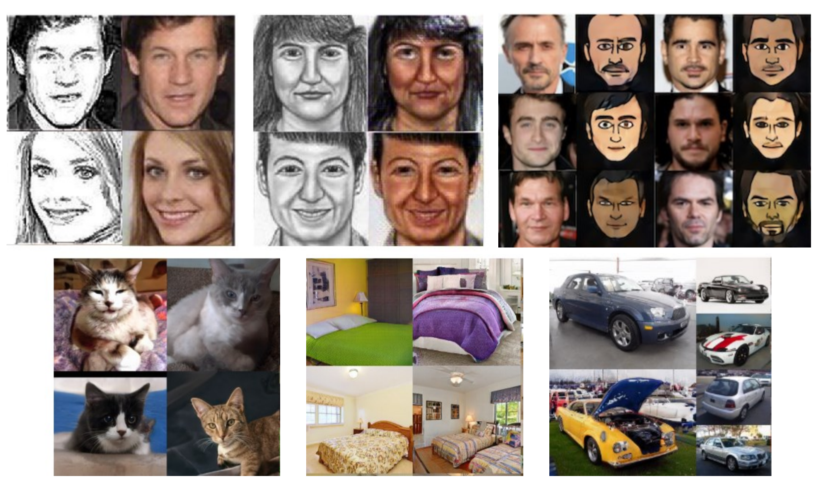
Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. Advances in Neural Information Processing Systems, 2014, pp. 2672–2680.

iCtLab Digital Forensics

36

Impressive Applications of Generative Adversarial Networks (GANs)

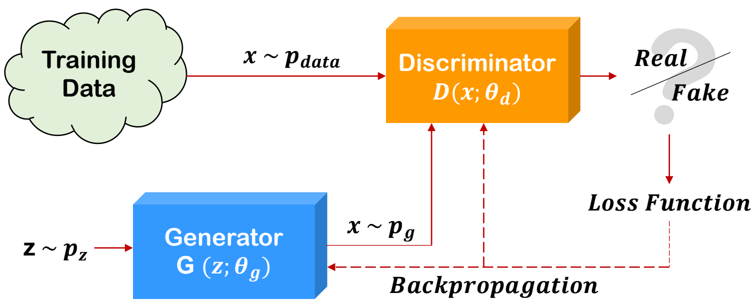
- Generate Human Faces
- Generate Cartoon Characters
- Image-to-Image Translation
- Text-to-Image Translation
- Semantic-Image-to-Photo Translation
- Generate New Human Poses
- Face Aging
- Super Resolution
- Photo Inpainting
- …



iCtLab Digital Forensics

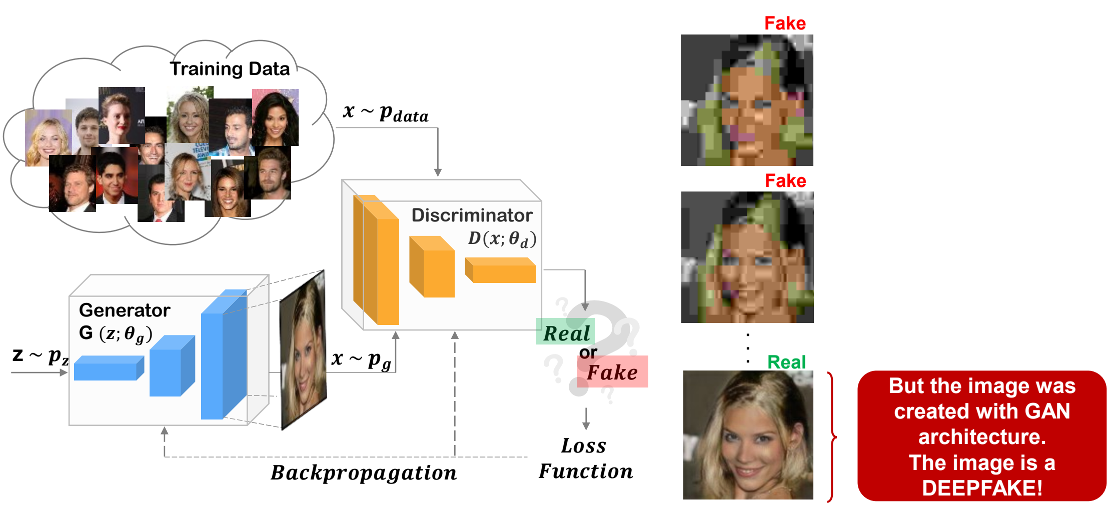
37

Generative Adversarial Networks



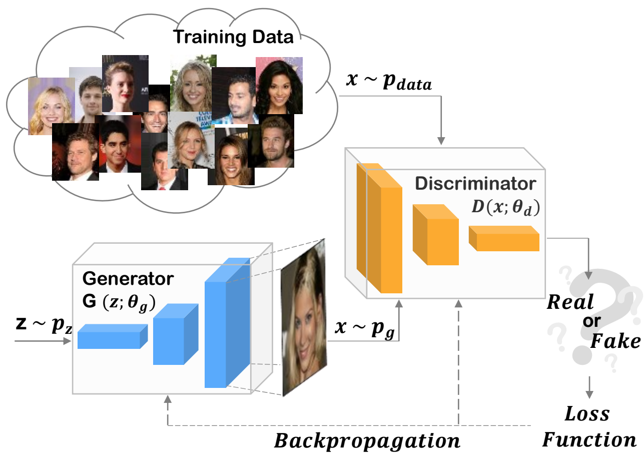
38

GAN example for human faces generation



39

GAN example for human faces generation



The evolution of GAN models for the creation of synthetic multimedia contents

Image source: Salehi, Pegah, Abdolah Chalechale, and Maryam Taghizadeh. **"Generative Adversarial Networks (GANs): An Overview of Theoretical Model, Evaluation Metrics, and Recent Developments."** *arXiv preprint arXiv:2005.13178* (2020).

ICtLab Digital Forensics

40

Deepfake: technologies for image creation of human faces

ICtLab Digital Forensics

FORENSIC VIRTUAL SUMMIT

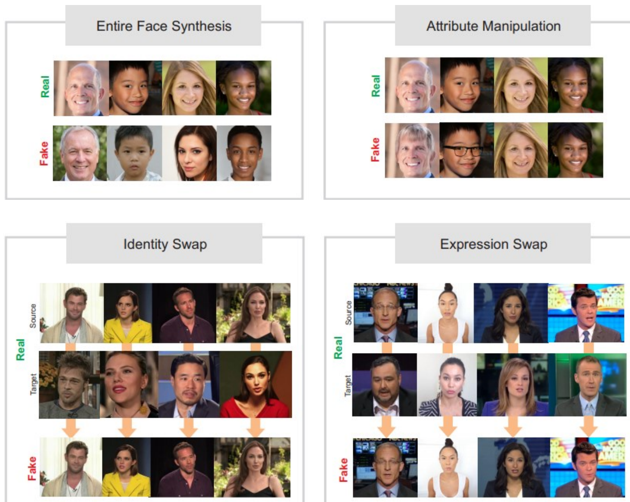
41

Tolosana, Ruben, et al. **"Deepfakes and beyond: A survey of face manipulation and fake detection."** *arXiv preprint arXiv:2001.00179* (2020).

Facial manipulation group

Entire Face Synthesis
This manipulation creates entire non-existent face images, usually through powerful GAN, e.g. **StyleGAN**.

Identity Swap
This manipulation consists of replacing the face of one person in a video with the face of another person.



Attribute Manipulation
This manipulation, also known as **face editing** or **face retouching**, consists of modifying some attributes of the face such as the hair color, the gender, the age, adding glasses, etc.

Expression Swap
This manipulation, also known as **face reenactment**, consists of modifying the facial expression of the person

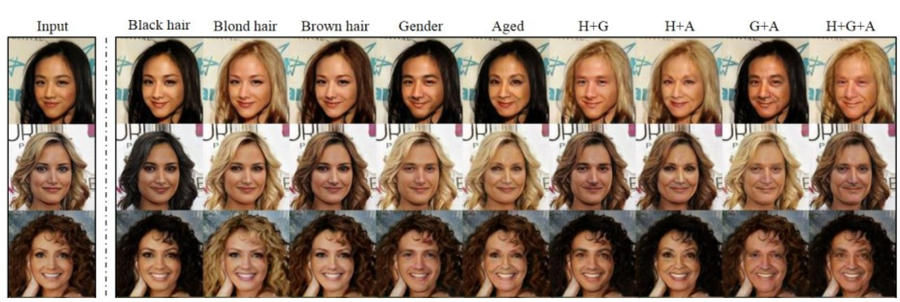
ICtLab Digital Forensics

42

StarGAN

Several techniques present at the state of the art based on DeepFake are limited in the management of **more than two domains** (for example, to change **hair color, gender, age, and many others features in a face**), since they should be generated different models for each pair of image domains.

Attribute Manipulation
This manipulation, also known as **face editing** or **face retouching**, consists of modifying some attributes of the face such as the hair color, the gender, the age, adding glasses, etc.



Yunjey Choi, Minje Choi, Munyoung Kim, Jung-Woo Ha, Sunghun Kim, and Jaegul Choo. **Stargan: Unified generative adversarial networks for multi-domain image-to-image translation**. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 8789–8797, 2018

ICtLab Digital Forensics

43

Entire Face Synthesis

This manipulation creates entire non-existent face images, usually through powerful GAN, e.g. **StyleGAN**.

StyleGAN

The **Style Generative Adversarial Network**, or **StyleGAN**, is an extension to the GAN architecture that proposes **large changes to the generator model**, including the use of a mapping network to map points in latent space to an intermediate latent space, the use of the intermediate latent space to **control style at each point in the generator model**, and the introduction to noise as a source of variation at each point in the generator model.

https://github.com/NVlabs/stylegan

Karras, Tero, Samuli Laine, and Timo Aila. **A style-based generator architecture for generative adversarial networks.** *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2019.

iCtLab Digital Forensics

44

StyleGAN artifacts

In this example the teeth do not follow the pose but stay aligned to the camera, as indicated by the blue line

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.

iCtLab Digital Forensics

45

StyleGAN2

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.

iCtLab Digital Forensics

46

Deepfake detection methods

Preliminary Forensics Analysis of DeepFake Images

Guarnera, Luca, et al. "Preliminary Forensics Analysis of DeepFake Images." arXiv preprint arXiv:2004.12626 (2020).

iCtLab Digital Forensics

FORENSIC VIRTUAL SUMMIT

47

Guarnera, Luca, et al. "**Preliminary Forensics Analysis of DeepFake Images**." *arXiv preprint arXiv:2004.12626* (2020).

DeepFake Forensics Analysis

A forensics analysis was carried out on sample DeepFake Images by means of one of the most famous image forensics software **Amped Authenticate.**

Amped authenticate.
https://ampedsoftware.com/it/authenticate/

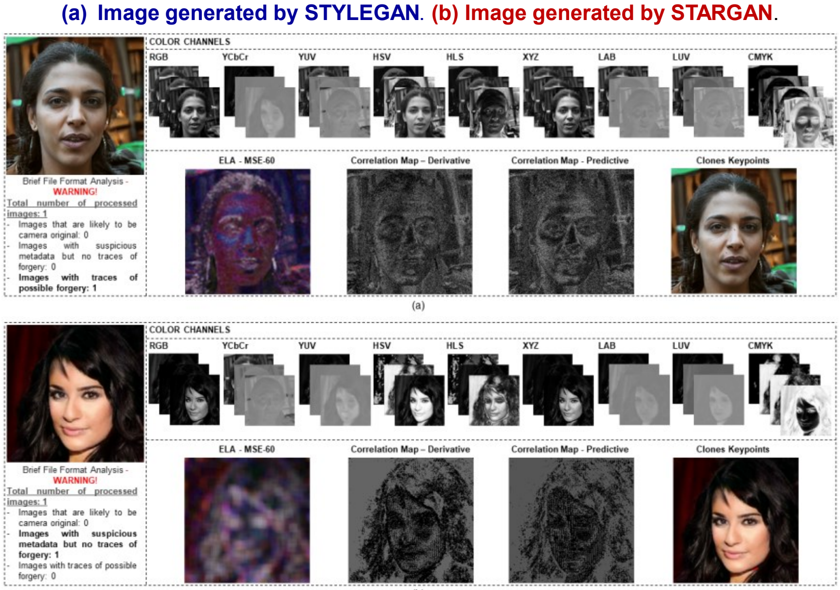
In particular we analyzed the images in:

- **different color spaces** (**RGB, YCbCr, YUV, HSV, HLS, XYZ, LAB, LUV, CMYK**);
- **domains** (**ELA, DCT Map, JPEG Dimples Map, Blocking Artifacts, JPEG Ghosts Map, Fusion Map, Correlation Map, PRNU Map, PRNU Tampering, LGA**)
- and by means of many **forgery detection techniques** (**Clones Blocks, Clones Keypoints** (Orb and Brisk)).

48

Example of Analysis carried out with Amped Authenticate software.
(a) Image generated by STYLEGAN. **(b) Image generated by STARGAN**.

Guarnera, Luca, et al. "**Preliminary Forensics Analysis of DeepFake Images**." *arXiv preprint arXiv:2004.12626* (2020).



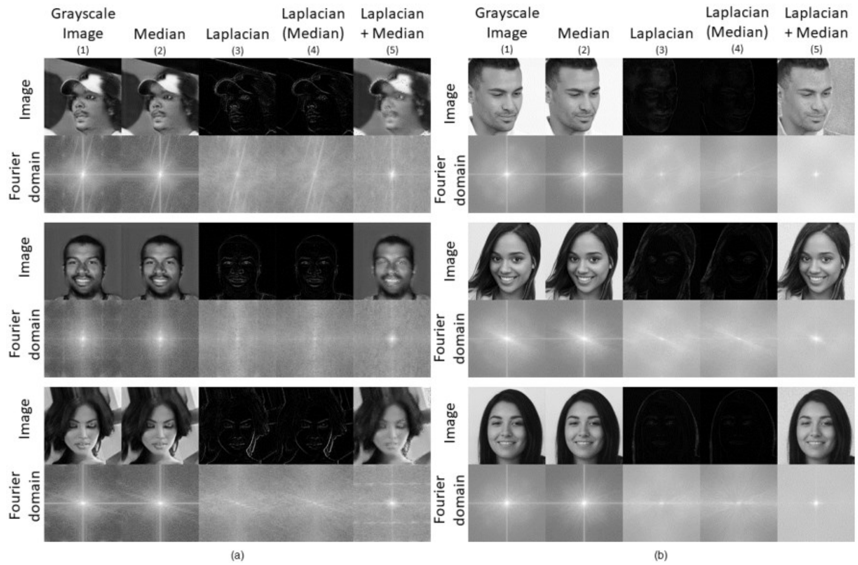
49

Fourier Transform

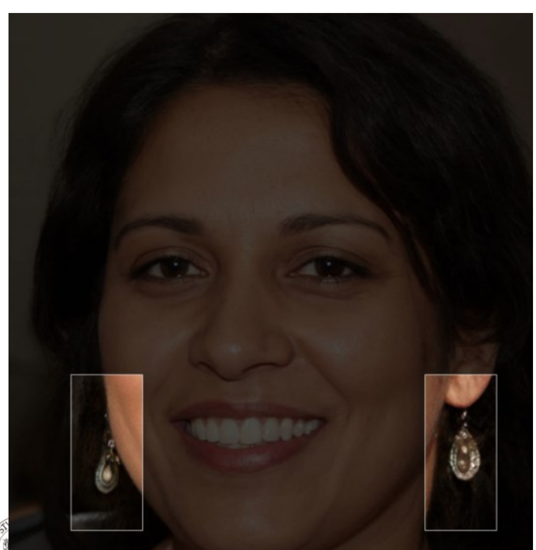
Example of Analysis carried out with Amped Authenticate software.
(a) **Image generated by STARGAN**.
(b) **Image generated by STYLEGAN**.

Each image of both datasets was converted to **grayscale (1)** and applied progressively: the **Median filter (2)**, the **Laplacian filter (3)**, the **Laplacian filter (4)** applied to the result of (2), the **sum of the results between the Median and Laplacian filters (5).**
For each operation performed, we show the **Fourier transform**.

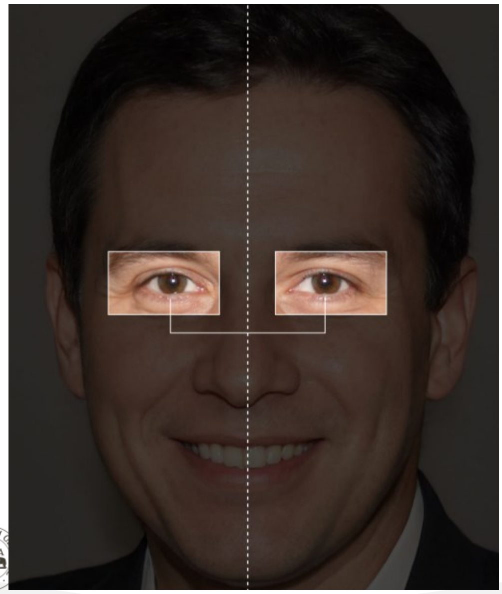
Guarnera, Luca, et al. "**Preliminary Forensics Analysis of DeepFake Images**." *arXiv preprint arXiv:2004.12626* (2020).



Other Anomalies



Other Anomalies



52

Other Anomalies

53

Other Anomalies

54

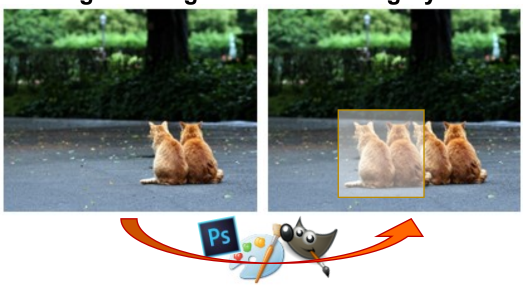
Deepfake detection methods

Guarnera, Luca, Oliver Giudice, and Sebastiano Battiato. "DeepFake Detection by Analyzing Convolutional Traces." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2020.

Guarnera, Luca, Oliver Giudice, and Sebastiano Battiato. "Fighting Deepfake by Exposing the Convolutional Traces on Images." IEEE Access 8 (2020): 165085-165098.

55

Original Image **Forgery**



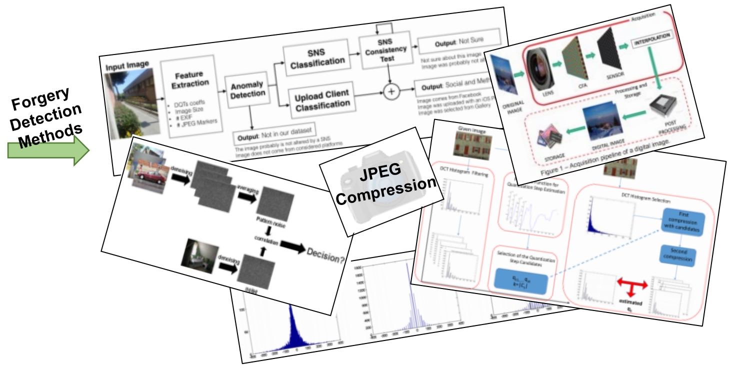
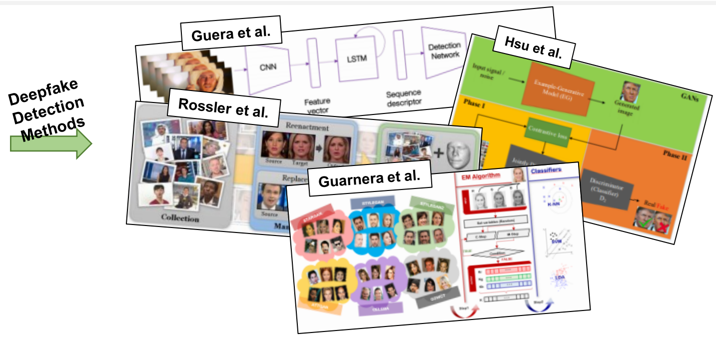


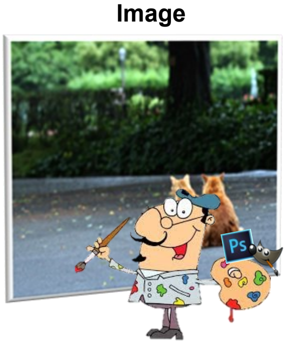
Image Forgery
Vs
DeepFake



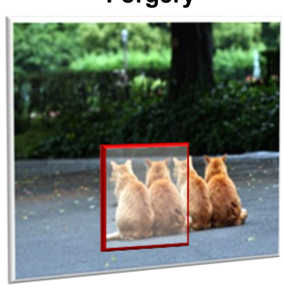
56

Image forgery

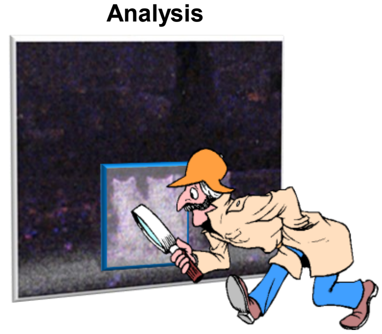
Image



Forgery



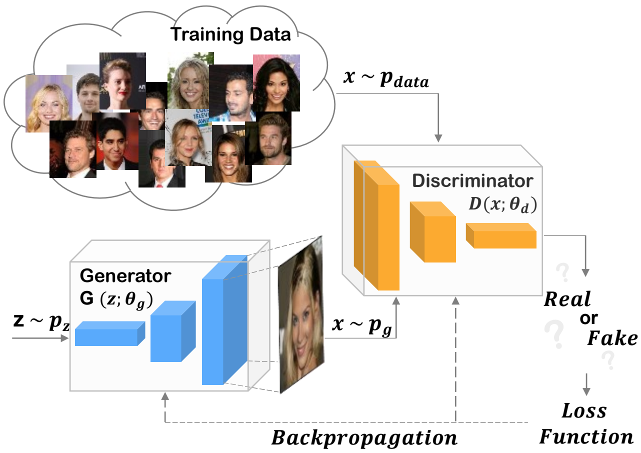
Analysis



iCtLab Digital Forensics

57

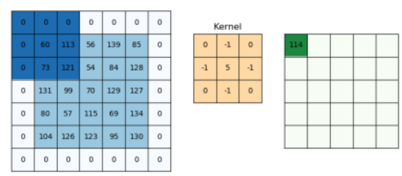
GAN – Convolution layer



Convolution operation

$$I[x,y] = \sum_{s,t=-\alpha}^{\alpha} k_{s,t} * I[x+s, y+t]$$

KERNEL

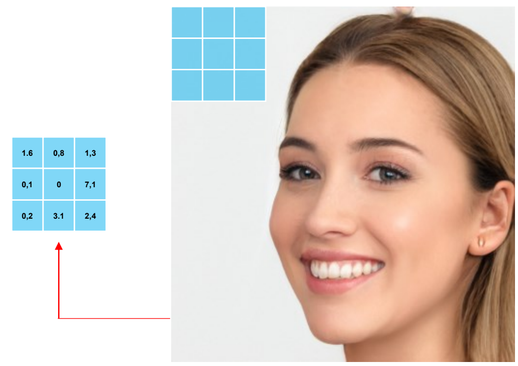


GAN Architectures

Method	Kernel size of the latest Convolution Layer
GDWCT	4x4
STARGAN	7x7
ATTGAN	4x4
STYLEGAN	3x3
STYLEGAN2	3x3

Expectation Maximization Algorithm

The goal is to **extract the fingerprint** from **input image I** able to numerically represent the **correlations** between each pixel.



$$I[x,y] = \sum_{s,t=-\alpha}^{\alpha} k_{s,t} * I[x+s, y+t]$$

Guarnera, Luca, Oliver Giudice, and Sebastiano Battiato. "**DeepFake Detection by Analyzing Convolutional Traces**." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2020.

iCtLab Digital Forensics

60

Expectation Maximization Algorithm

Assume that the element **I[x,y]** belongs to one of the following models:

- **M1**: when the element **I[x,y]** satisfies

$$I[x,y] = \sum_{s,t=-\alpha}^{\alpha} k_{s,t} * I[x+s, y+t]$$

- **M2:** otherwise.

Expectation Maximization (EM) algorithm:

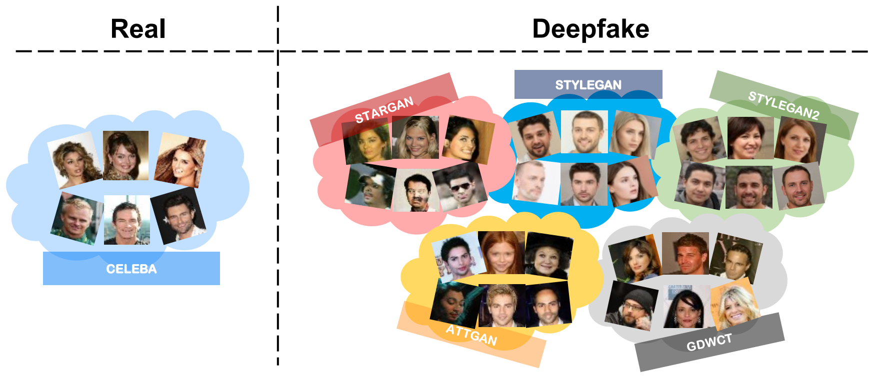
- **Expectation step**: calculates the (density of) probability that each element belongs to a model (M1 or M2);
- **Maximization step**: estimates the (weighted) parameters based on the probabilities of belonging to instances of (M1 or M2).

Guarnera, Luca, Oliver Giudice, and Sebastiano Battiato. "**DeepFake Detection by Analyzing Convolutional Traces**." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2020.

iCtLab Digital Forensics

61

Dataset

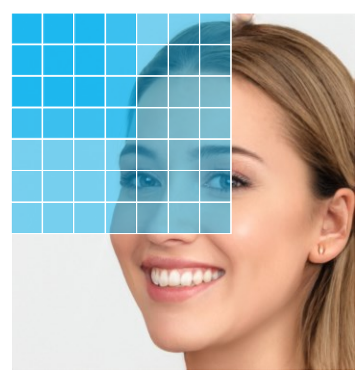


Guarnera, Luca, Oliver Giudice, and Sebastiano Battiato. "**DeepFake Detection by Analyzing Convolutional Traces**." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2020.

62

Dataset

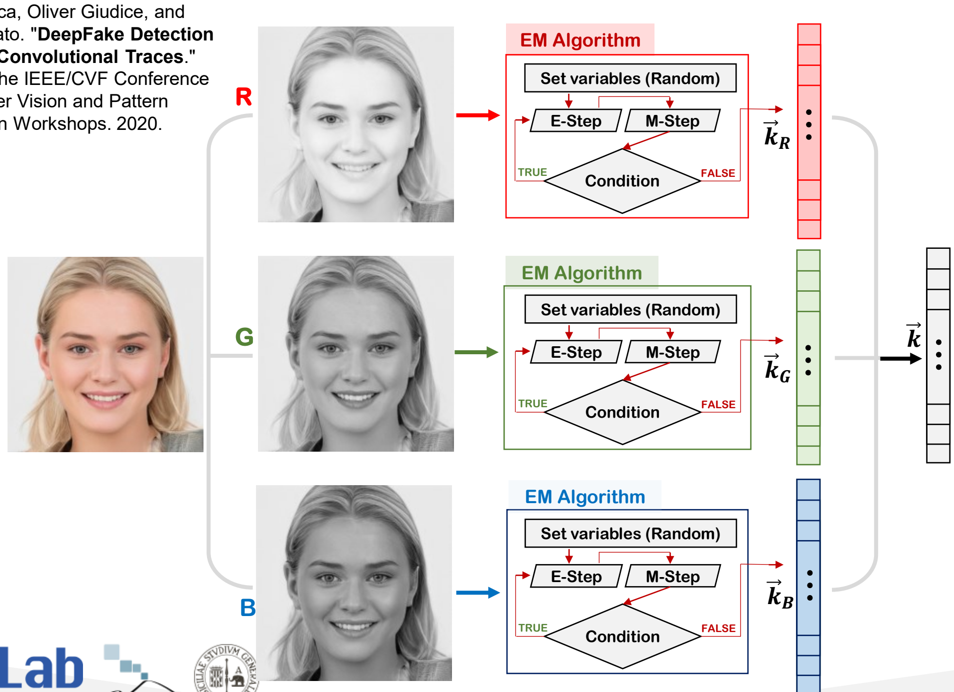
Method	Number of images generated	Size	Data input to the network	Goal of the network	Kernel size of the latest Convolution Layer
GDWCT	3369	216x216	CELEBA	Improves the stylization capability	4x4
STARGAN	5648	256x256	CELEBA	Image-to-image translations on multiple domains using a single model	7x7
ATTGAN	6005	256x256	CELEBA	Transfer of face attributes with classification constraints	4x4
STYLEGAN	9999	1024x1024	CELEBA-HQ FFHQ	Transfer semantic content from a source domain to a target domain characterized by a different style	3x3
STYLEGAN2	3000	1024x1024	FFHQ	Transfer semantic content from a source domain to a target domain characterized by a different style	3x3



Guarnera, Luca, Oliver Giudice, and Sebastiano Battiato. "**DeepFake Detection by Analyzing Convolutional Traces**." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2020.

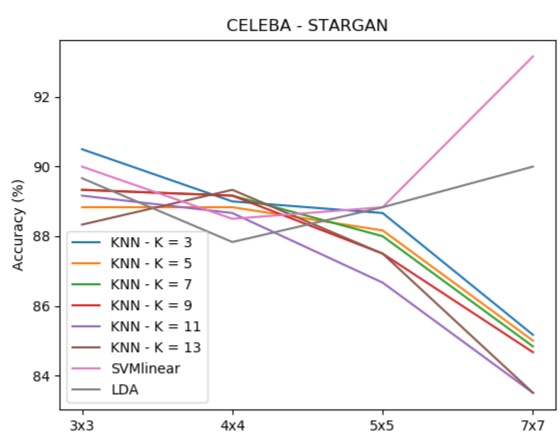
63

Guarnera, Luca, Oliver Giudice, and Sebastiano Battiato. "**DeepFake Detection by Analyzing Convolutional Traces**." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2020.



CELEBA Vs STARGAN

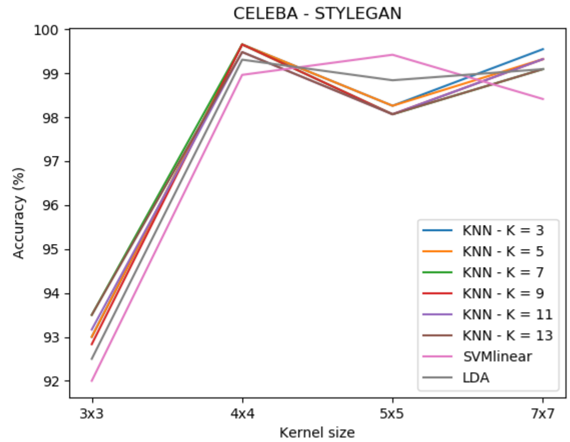
	Kernel 3x3	Kernel 4x4	Kernel 5x5	Kernel 7x7
KNN, k = 3	90.50	89.00	88.67	85.17
KNN, k = 5	88.83	88.83	88.17	85.00
KNN, k = 7	89.33	89.17	88.00	84.83
KNN, k = 9	89.33	89.17	87.50	84.67
KNN, k = 11	89.17	88.67	86.67	83.50
KNN, k = 13	88.33	89.33	87.50	83.50
SVMLinear	90.00	88.50	88.83	93.17
LDA	89.67	87.83	88.83	90.00



Guarnera, Luca, Oliver Giudice, and Sebastiano Battiato. "**DeepFake Detection by Analyzing Convolutional Traces**." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2020.

CELEBA - STYLEGAN

	Kernel 3x3	Kernel 4x4	Kernel 5x5	Kernel 7x7
KNN, k = 3	93.00	99.65	98.26	99.55
KNN, k = 5	93.00	99.65	98.26	99.32
KNN, k = 7	93.50	99.65	98.07	99.09
KNN, k = 9	92.83	99.65	98.07	99.32
KNN, k = 11	93.17	99.48	98.07	99.32
KNN, k = 13	93.50	99.48	98.07	99.09
SVMLinear	92.00	98.96	99.42	98.41
LDA	92.50	99.31	98.84	99.09



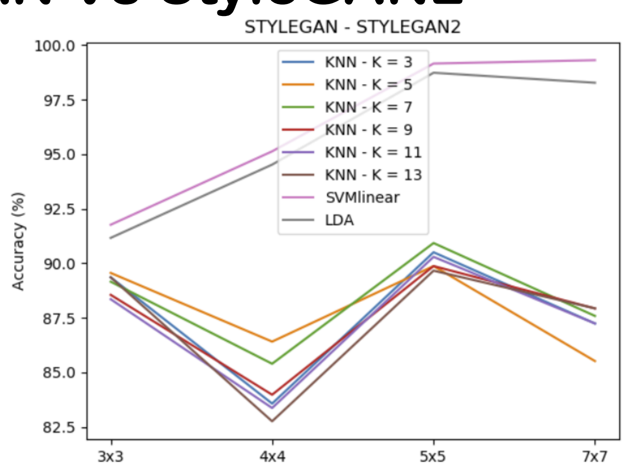
Guarnera, Luca, Oliver Giudice, and Sebastiano Battiato. "**DeepFake Detection by Analyzing Convolutional Traces**." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2020.

iCtLab Digital Forensics

66

StyleGAN Vs StyleGAN2

	Kernel 3x3	Kernel 4x4	Kernel 5x5	Kernel 7x7
KNN, k = 3	89.36	83.57	90.51	87.24
KNN, k = 5	89.56	86.41	89.87	85.52
KNN, k = 7	89.16	85.40	90.93	87.59
KNN, k = 9	88.55	83.98	89.87	87.93
KNN, k = 11	88.35	83.37	90.30	87.24
KNN, k = 13	89.36	82.76	89.66	87.93
SVMLinear	91.77	95.13	99.16	99.31
LDA	91.16	94.52	98.73	98.28



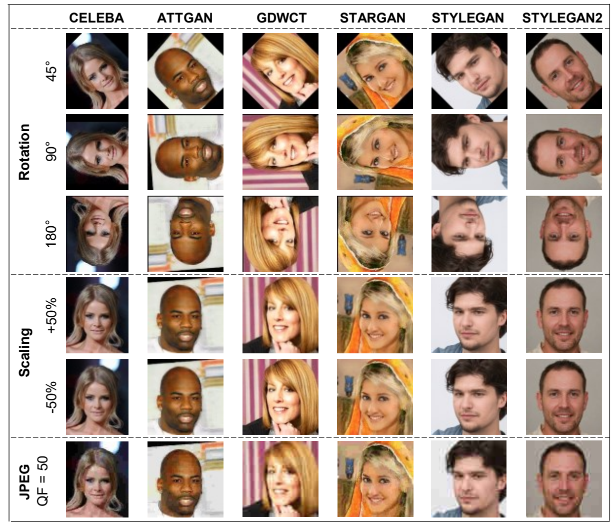
Guarnera, Luca, Oliver Giudice, and Sebastiano Battiato. "**DeepFake Detection by Analyzing Convolutional Traces**." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2020.

iCtLab Digital Forensics

67

Guarnera, Luca, Oliver Giudice, and Sebastiano Battiato. "**Fighting Deepfake by Exposing the Convolutional Traces on Images**." *IEEE Access* 8 (2020): 165085-165098.

Robustness Experiments



iCtLab Digital Forensics

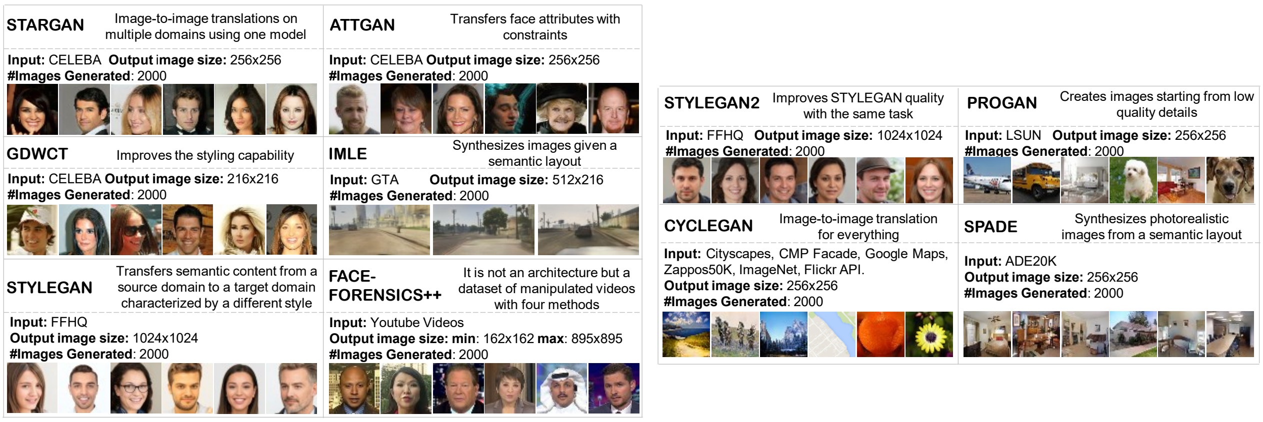
Robustness Experiments

	ATTGAN Kernel Size			GDWCT Kernel Size			STARGAN Kernel Size			STYLEGAN Kernel Size			STYLEGAN2 Kernel Size		
	3x3	5x5	7x7	3x3	5x5	7x7	3x3	5x5	7x7	3x3	5x5	7x7	3x3	5x5	7x7
Raw Images	92.99	88.51	87.11	91.58	77.41	82.01	88.54	84.43	90.55	95.29	99.48	99.30	96.91	99.64	99.32
Random Square	82.54	75.47	75	62.03	61.54	63.27	81.16	78.95	76.19	97.26	100	97.37	99.02	100	100
Gaussian Blur. kernel size = 3x3	77.78	73.58	72.22	56.96	59.38	61.22	73.91	80.7	61.9	93.15	98.33	92.11	96.08	98.81	96.08
Gaussian Blur. kernel size = 9x9	76.19	76.92	68.57	56.96	67.19	61.22	72.46	77.19	64.29	97.26	100	94.59	96.08	97.62	94.12
Gaussian Blur. kernel size = 15x15	80.95	76.92	77.14	64.56	67.69	57.14	82.61	80.7	75.61	97.26	98.33	94.59	100	97.59	98.04
Rotation 45°	90	84.31	85.29	67.53	73.02	66.67	85.29	82.14	87.8	89.04	91.67	91.89	97.4	94.2	97.62
Rotation 90°	100	94.23	100	93.59	92.19	93.75	92.75	92.98	97.56	100	100	97.3	100	100	100
Rotation 180°	83.87	86.54	82.86	74.36	67.19	59.18	84.06	91.23	78.57	100	100	91.89	97.03	98.8	98.04
Scaling +50%	88.71	78.43	91.18	78.21	71.88	68.09	89.71	83.93	90	97.22	100	97.3	99	98.78	100
Scaling -50%	75.81	78.85	77.78	71.79	57.81	68.09	79.71	64.91	64.29	95.83	96.67	100	99.01	97.59	94.23
JPEG Compression	86.69	91.67	91.18	85.17	89.33	84.66	89.17	92.69	92.01	99.5	99.33	97.57	99.49	98.96	98.55

Guarnera, Luca, Oliver Giudice, and Sebastiano Battiato. "**Fighting Deepfake by Exposing the Convolutional Traces on Images**." *IEEE Access* 8 (2020): 165085-165098.

iCtLab Digital Forensics

Real Vs Deepfake

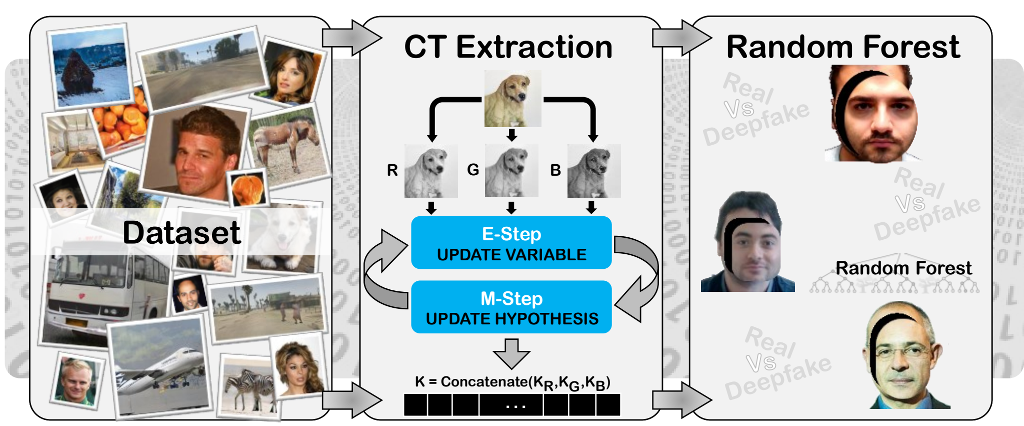


Guarnera, Luca, Oliver Giudice, and Sebastiano Battiato. "**Fighting Deepfake by Exposing the Convolutional Traces on Images**." *IEEE Access* 8 (2020): 165085-165098.

iCtLab Digital Forensics

70

Real Vs Deepfake

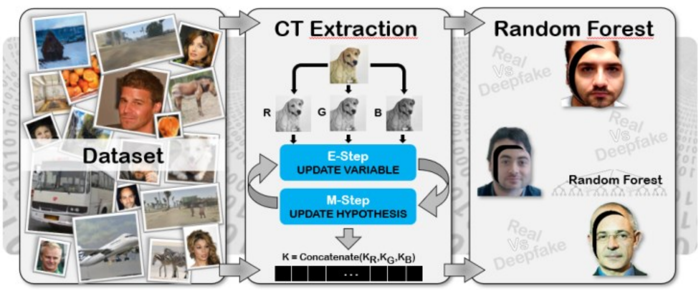


Guarnera, Luca, Oliver Giudice, and Sebastiano Battiato. "**Fighting Deepfake by Exposing the Convolutional Traces on Images**." *IEEE Access* 8 (2020): 165085-165098.

iCtLab Digital Forensics

71

Real Vs Deepfake (Binary classification)



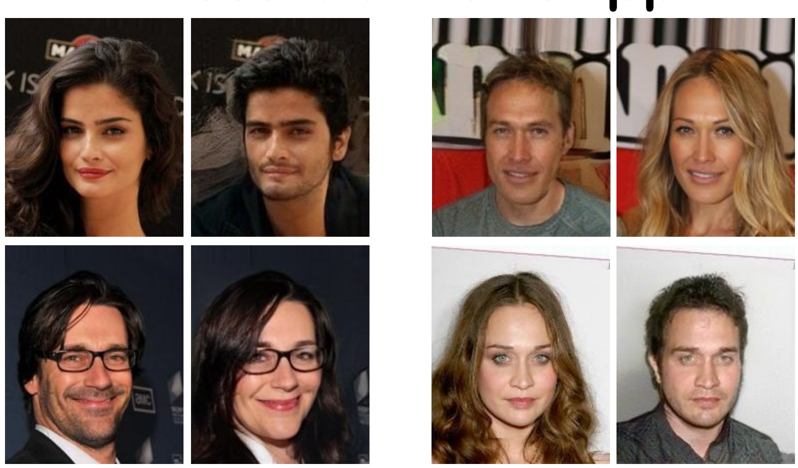
	CELEBA Vs DeepNetworks Kernel Size		
	3x3	5x5	7x7
3-NN	**89.80**	77.38	78.63
5-NN	**90.79**	77.20	77.80
7-NN	**90.44**	76.47	78.39
9-NN	**90.30**	77.20	78.28
11-NN	**89.80**	77.29	77.45
13-NN	**89.73**	77.66	77.69
SVMLinear	**84.14**	76.28	80.28
SVMSigmoid	58.57	61.36	**63.52**
SVMrbf	**91.22**	80.04	80.87
SVMPoly	**88.74**	78.66	78.87
LDA	**83.50**	77.38	78.98
Random Forest	**98.07**	93.81	91.22

Guarnera, Luca, Oliver Giudice, and Sebastiano Battiato. "**Fighting Deepfake by Exposing the Convolutional Traces on Images**." *IEEE Access* 8 (2020): 165085-165098.

iCtLab Digital Forensics

73

Test with FaceApp



CELEBA FACEAPP CELEBA FACEAPP

Accuracy

> 93%

Guarnera, Luca, Oliver Giudice, and Sebastiano Battiato. "**Fighting Deepfake by Exposing the Convolutional Traces on Images**." *IEEE Access* 8 (2020): 165085-165098.

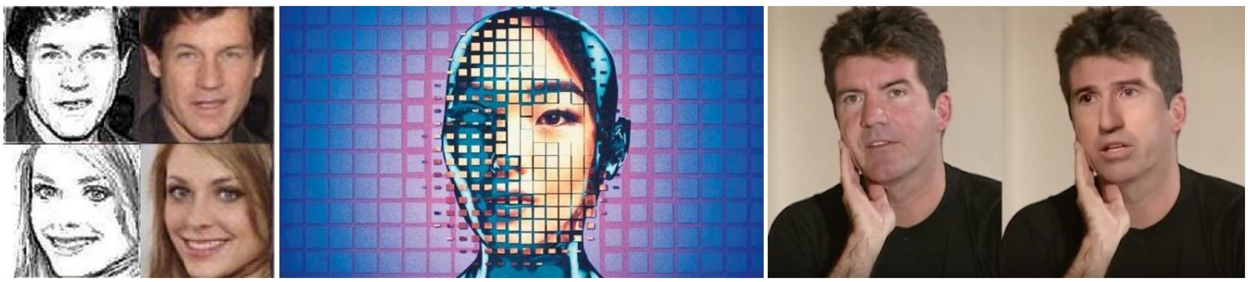
iCtLab Digital Forensics

74

Conclusion and Future Works

It turns out to be very important to be able to create new methods that can counter this phenomenon.

This could be done by analyzing details and traces of underlying generation process of the image (e.g. in the Fourier domain).



75

Conclusion so far …

Generalization to new datasets and methods is **extremely hard**!!!

- Training on more datasets helps
- Training on more methods helps
- Domain adaptation / transfer learning methods could be the key issue
- Larger question: how many generalizable features?
- More emphasis on combination of multiple cues (audio and video)

- Testing on the “wild” is fundamental

76

References/links

- Media forensics and deepfakes: an overview - L. Verdoliva - IEEE Journal of Selected Topics in Signal Processing - 2020
- DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection – Tolosana et al. (2020) –
- Celeb-DF: A Large-Scale Challenging Dataset for DeepFake Forensics – Li et al. CVPR 2020
- Workshop on Media Forensics:

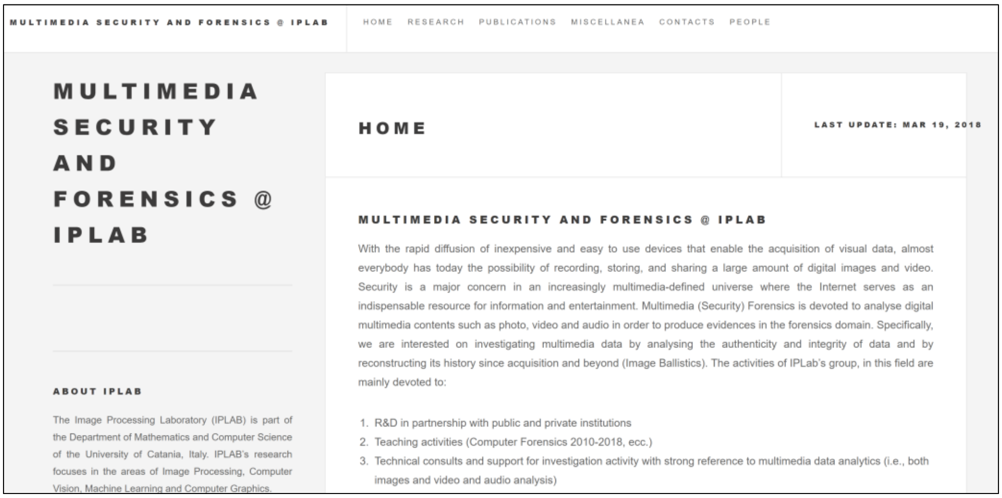
77

Reference

- Guarnera, Luca, Oliver Giudice, and Sebastiano Battiato. "**DeepFake Detection by Analyzing Convolutional Traces**." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2020.
- Guarnera, Luca, Oliver Giudice, and Sebastiano Battiato. "**Fighting Deepfake by Exposing the Convolutional Traces on Images**." IEEE Access 8 (2020): 165085-165098.
- Guarnera, Luca, Oliver Giudice, Cristina Nastasi, and Sebastiano Battiato. "**Preliminary Forensics Analysis of DeepFake Images.**" AEIT proceedings (2020).
- Oliver Giudice, Luca Guarnera, Sebastiano Battiato - «**Fighting Deepfakes by Detecting GAN DCT Anomalies**” https://arxiv.org/abs/2101.09781 (2021)

78

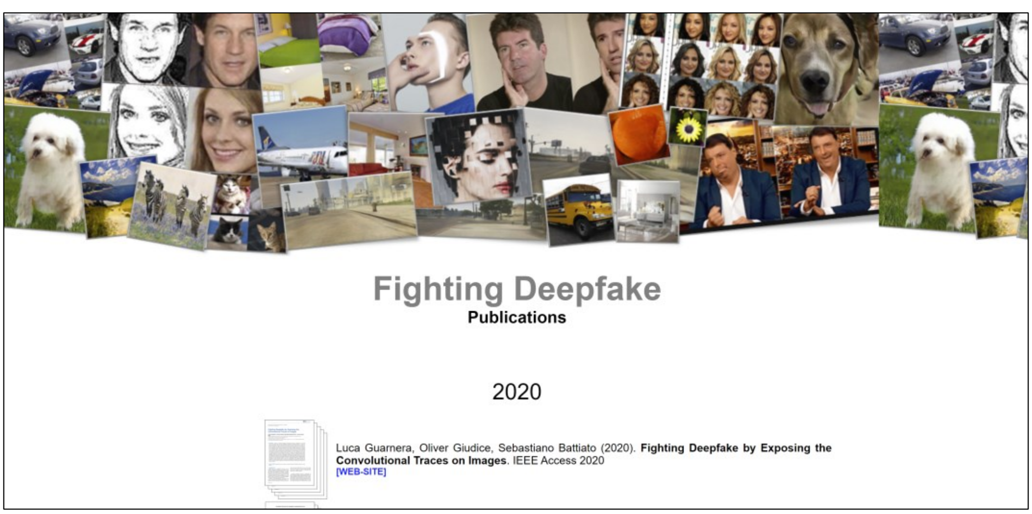
https://iplab.dmi.unict.it/mfs/



iCtLab Digital Forensics

79

https://iplab.dmi.unict.it/mfs/Deepfakes/



iCtLab Digital Forensics

80

Further references

- W. Cho, S. Choi, D. K. Park, I. Shin, and J. Choo. ***Image-to-image translation via group-wise deep whitening-and-coloring transformation***. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 10639–10647, 2019
- Y. Choi, M. Choi, M. Kim, J. Ha, S. Kim, and J. Choo. ***Stargan: Unified generative adversarial networks for multi-domain image-to-image translation***. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 8789–8797, 2018.
- Z. He, W. Zuo, M. Kan, S. Shan, and X. Chen. ***Attgan: Facial attribute editing by only changing what you want***. IEEE Transactions on Image Processing, 28(11):5464–5478, 2019.
- T. Karras, S. Laine, and T. Aila. ***A style-based generator architecture for generative adversarial networks***. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 4401–4410, 2019

81

Seeing Isn't Believing

82

Houston we have a problem: Deepfake is the word!

83

Prof. Sebastiano Battiato
Dipartimento di Matematica e Informatica
University of Catania, Italy

Image Processing LAB – http://iplab.dmi.unict.it

iCTLab - www.ictlab.srl

battiato@dmi.unict.it

84