Article

Fighting Deepfakes by Detecting GAN DCT Anomalies

Oliver Giudice [1,*] , Luca Guarnera [1,2] and Sebastiano Battiato [1,2]

[1] Department of Mathematics and Computer Science, University of Catania, 95125 Catania, Italy;
luca.guarnera@unict.it (L.G.); battiato@dmi.unict.it (S.B.)
[2] iCTLab s.r.l., Spinoff of University of Catania, 95125 Catania, Italy
[*] Correspondence: giudice@dmi.unict.it

Abstract: To properly contrast the Deepfake phenomenon the need to design new Deepfake detection algorithms arises; the misuse of this formidable A.I. technology brings serious consequences in the private life of every involved person. State-of-the-art proliferates with solutions using deep neural networks to detect a fake multimedia content but unfortunately these algorithms appear to be neither generalizable nor explainable. However, traces left by Generative Adversarial Network (GAN) engines during the creation of the Deepfakes can be detected by analyzing ad-hoc frequencies. For this reason, in this paper we propose a new pipeline able to detect the so-called GAN Specific Frequencies (GSF) representing a unique fingerprint of the different generative architectures. By employing Discrete Cosine Transform (DCT), anomalous frequencies were detected. The β statistics inferred by the AC coefficients distribution have been the key to recognize GAN-engine generated data. Robustness tests were also carried out in order to demonstrate the effectiveness of the technique using different attacks on images such as JPEG Compression, mirroring, rotation, scaling, addition of random sized rectangles. Experiments demonstrated that the method is innovative, exceeds the state of the art and also give many insights in terms of explainability.

Keywords: deepfake detection; Generative Adversarial Networks; multimedia forensics; image forensics

1. Introduction

Artificial Intelligence (AI) techniques to generate synthetic media and their circulation on the network led to the birth, in 2017, of the Deepfake phenomenon: altered (or created) multimedia content by ad-hoc machine learning generative models, e.g., the Generative Adversarial Network (GAN) [1]. Images and videos of famous people, available on different media like TV and Web, could appear authentic at first glance, but they may be the result of an AI process which delivers very realistic results. In this context the 96% of these media are porn (deep porn) [2]. If we think that anyone could be the subject of this alteration we can understand how a fast and reliable solution is needed to contrast the Deepfake phenomenon. Most of the techniques already proposed in literature act as a "black box" by tuning ad-hoc deep architectures to distinguish "real" from "fake" images generated by specific GAN machines. It seems not easy to find a robust detection method capable of working in the wild; even current solutions need a considerable amount of computing power. Let's assume that any generative process based on GAN, presents an automated operating principle, resulting from a learning process. In [3], it has been already demonstrated that it is possible to attack and retrieve the signature on the network's deconvolutional layers; in this paper a method to identify any anomaly of the generated "fake" signal, only partially highlighted in some preliminary studies [4,5] is presented. The Fourier domain demonstrated to be prone and robust into understanding semantic at superordinate level [6]. Spatial domain has been recently further investigated by [7–9] to gain robustness and exploit related biasness [10]. To improve the efficiency, the 8×8 DCT has been exploited, by employing similar data analysis made in [11,12] and extracting simple statistics of the underlying distribution [13]. The final classification engine based

on gradient boosting, properly manages and isolates the GAN Specific Frequencies (*GSF*), of each specific architecture, a sort of fingerprint/pattern, outperforming state-of- the-art methods. In this paper a new "white box" method of Deepfake detection called CTF-DCT (Capture the Fake by DCT Fingerprint Extraction) is proposed, based on the analysis of the Discrete Cosine Transform (DCT) coefficients. Experiments on Deepfake images of human faces proved that a proper signature of the generative process is embedded on the given spatial frequencies. In particular we stress the evidence, that such kind of images, have in common global shape and main structural elements allowing to isolate artefacts that are not only unperceivable but also capable to discriminate between the different GANs. Finally, the robust classifier is able to demonstrate its generalizing ability in the wild even on Deepfakes not generated by GAN-engines demonstrating the ability to catch artefacts related to reenactment forgeries.

The main contributions of this research are the following:

- A new high-performance Deepfake face detection method based on the analysis of the AC coefficients calculated through the Discrete Cosine Transform, which delivered not only great generalization results but also impressive classification results with respect to previous published works. The method does not require computation via GPU and "hours" of training to perform Real Vs Deepfake classifications;
- The detection method is "explainable" (white-box method). Through a simple estimation of the characterizing parameters of the Laplacian distribution, we are able to detect those anomalous frequencies generated by various Deepfake architectures;
- Finally, the detection strategy was attacked to simulated situations in the wild. Mirroring, scaling, rotation, addition of random size rectangles, position and color were applied to the images, also demonstrating the robustness of the proposed method and the ability to perform well even on video dataset never taken into account during training.

The paper is organized as follows: Section 2 presents the state-of-the-art of Deepfake generation and detection methods. The proposed approach is described in Section 3. Section 5, a discussion of GSF is reported. Experimental results, robustness test and comparison with competing methods are reported in Section 6. Section 7 concludes the paper with suggestions for future works.

2. Related Works

AI-synthetic media are generally created by techniques based on GANs, firstly introduced by Goodfellow et al. [1]. GANs train two models simultaneously: a generative model *G*, that captures the data distribution, and a discriminative model *D*, able to estimate the probability that a sample comes from the training data rather than from *G*. The training procedure for *G* is to maximize the probability of *D* making a mistake thus resulting to a min-max two-player game.

An overview on Media forensics with particular focus on Deepfakes has been recently proposed in [14,15].

Five of the most famous and effective architectures in state-of-the-art for Deepfakes facial images synthesis were taken into account (StarGAN [16], StyleGAN [17], StyleGAN2 [18], ATTGAN [19] GDWCT [20]) used in our experiments as detailed below.

2.1. Deepfake Generation Techniques of Faces

StarGAN [16], proposed by Choi et al., is a method capable of performing image-to-image translations on multiple domains (such as change hair color, change gender, etc.) using a single model. Trained on two different types of face datasets—CELEBA [21] containing 40 labels related to facial attributes such as hair color, gender and age, and RaFD dataset [22] containing 8 labels corresponding to different types of facial expressions ("happy", "sad", etc.)—this architecture, given a random label as input (such as hair color, facial expression, etc.), is able to perform an image-to-image translation operation with impressive visual result.

An interesting study was proposed by He et al. [19] with a framework called AttGAN in which an attribute classification constraint is applied in the latent representation to the generated image, in order to guarantee only the correct modifications of the desired attributes.

Another style transfer approach is the work of Cho et al. [20], proposing a groupwise deep whitening-and coloring method (GDWCT) for a better styling capacity. They used CELEBA, Artworks [23], cat2dog [24], Ink pen and watercolor classes from Behance Artistic Media (BAM) [25], and Yosemite datasets [23] as datasets improving not only computational efficiency but also quality of generated images.

Finally, one of the most recent and powerful methods regarding the entire-face synthesis is the Style Generative Adversarial Network architecture or commonly called StyleGAN [17], where, by means of mapping points in latent space to an intermediate latent space, the framework controls the style output at each point of the generation process. Thus, StyleGAN is capable not only of generating impressively photorealistic and high-quality photos of faces, but also offers control parameters in terms of the overall style of the generated image at different levels of detail. While being able to create realistic pseudo-portraits, small details might reveal the fakeness of generated images. To correct those imperfections, Karras et al. made some improvements to the generator (including re-designed normalization, multi-resolution, and regularization methods) proposing Style-GAN2 [18] obtaining extremely realistic faces. Figure 1 shows an example of facial images created by five different generative architectures.
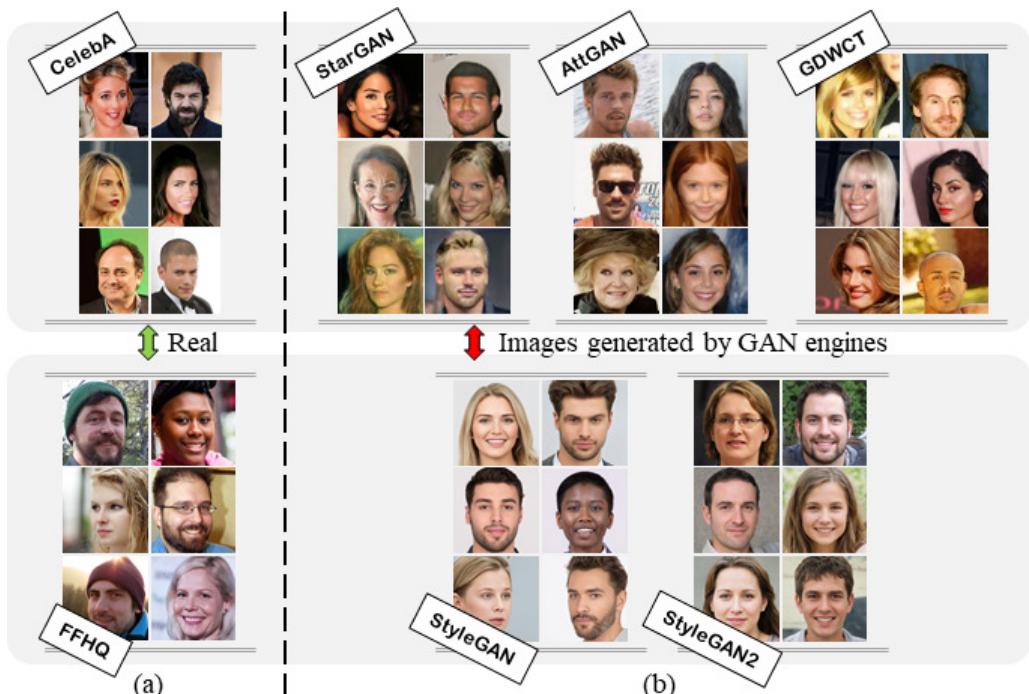
Figure 1. Example of real (**a**) and deepfake datasets (**b**) used in our experiments. The CelebA dataset was used to generate human face images with the StarGAN, AttGAN and GDWCT architectures. The FFHQ dataset was used to generate human face images with the StyleGAN and StyleGAN2 architectures.

2.2. Deepfake Detection Techniques

Almost all currently available strategies and methods for Deepfake detection are focused on anomalies detection trying to find artefact and traces of the underlying generative process. The Deepfake images could contain a pattern pointed out by the analysis of anomalous peaks appearing in the spectrum in the Fourier domain. Zhang et al. [5] analyze the artefacts induced by the up-sampler of GAN pipelines in the frequency domain. The authors proposed to emulate the synthesises of GAN artefacts. Results obtained by

the spectrum-based classifier greatly improves the generalization ability, achieving very good performances in terms of binary classification between authentic and fake images. Also Durall et al. [26] presented a method for Deepfakes detection based on the analysis in the frequency domain. The authors combined high-resolution authentic face images from different public datasets (CELEBA-HQ data set [27], Flickr-Faces-HQ data set [17]) with fakes (100K Faces project (https://generated.photos/, accessed on 14 February 2021), this person does not exist (https://thispersondoesnotexist.com/, accessed on 14 February 2021)), creating a new dataset called Faces-HQ. By means of naive classifiers they obtained effective results in terms of overall accuracy of detection.

Wang et al. [28] proposed FakeSpotter, a new method based on monitoring single neuron behaviors to detect faces generate by Deepfake technologies. The authors used in the experiments CELEBA [21] and FFHQ (https://github.com/NVlabs/ffhq-dataset, accessed on 14 February 2021) images (real datasets of faces) and compared Fakespotter with Zhang et al. [5] obtaining an average detection accuracy of more than 90% on the four types of fake faces: Entire Synthesis [18,27], Attribute Editing [16,29], Expression Manipulation [17,29], DeepFake [30,31].

The work proposed by Jain et al. [32] consists of a framework called DAD-HCNN which is able to distinguish unaltered images from those that have been retouched or generated through different GANs by applying a hierarchical approach formed by three distinct levels. The last level is able to identify the specific GAN model (STARGAN [16], SRGAN [33], DCGAN [34], as well as the Context Encoder [35]). Liu et al. [36] proposed an architecture called Gram-Net, where, through the analysis of a global image texture representations, they managed to create a robust fake image detection. The results of the experiments, done both with Deepfake (DCGAN, StarGAN, PGGAN, StyleGAN) and real images (CelebA, CelebA-HQ, FFHQ), demonstrate that this new type of detector delivers effective results.

Recently, a study conducted by Hulzebosch [37] describes that the CNN solutions presented till today for Deepfake detection are limited to lack of robustness, generalization capability and explainability, because they are extremely specific to the context in which they were trained and, being very deep, tend to extract the underlying semantics from images. For this reason, in literature new algorithms capable to find the Deepfake content without the use of deep architectures were proposed. As described by Guarnera et al. [3,38], the current GAN architectures leaves a pattern (through convolution layers) that characterizes that specific neural architecture. In order to capture this forensic trace, the authors used the Expectation-Maximization Algorithm [39] obtaining features able to distinguish real images from Deepfake ones. Without the use of deep neural networks, the authors exceeded state-of-the-art in terms of accuracy in the real Vs Deepfake classification test, using not only Deepfake images generated by common GAN architectures, but also testing images generated by modern FaceApp mobile application.

Differently from the described approaches, in this paper, the possibility to capture the underlying pattern of a possible Deepfake is investigated extracting the discriminative features through the DCT transform.

3. The CTF Approach

In [37], Dutch law enforcement experts were tasked with discriminating between images from the FFHQ dataset and StyleGAN images, which were created starting from FFHQ. The results reached only the 63% of accuracy while state-of the-art methods [28,38] are able to deliver a better outcome. Algorithms were used for extracting black-box features that likely are not related to the visible domain but are somehow encoding anomalies strictly dependent on the way Deepfakes are generated. In particular, a refined evaluation of the StyleGAN images, shows that some abnormal patterns are visible in the most structured part of the images (e.g., skin, hair, etc.). Given such a repetitive pattern, which would have to be subsisting on the middle bands of the Deepfake image frequency spectrum, a frequency-based approach might be able to detect it and describe it. To this end,

the CTF approach will take place by leveraging more than a decade of JPEG compression pipeline studies employing DCT block-based processing, which is effectively used for many computer vision and image forensics tasks not strictly related to compression itself [11,12,40–42].

The CTF approach transform and analyse images on the DCT domain in order to detect the most discriminant information related to the pattern shown in Figure 2 which is typical of the employed Generative Model (e.g., GAN).
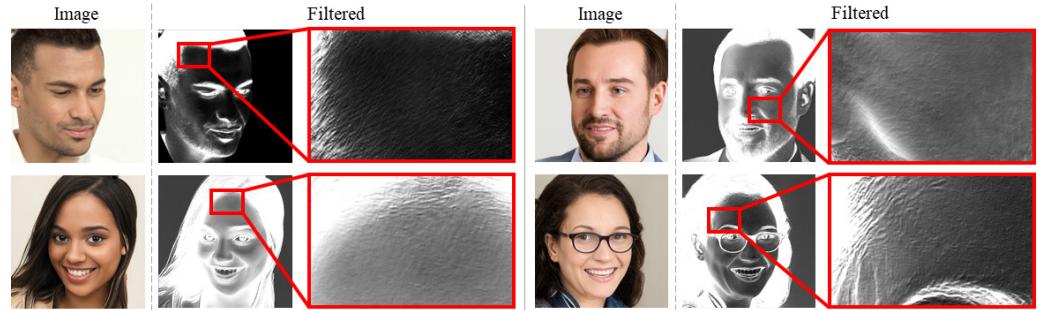

Figure 2. Example of image generated by StyleGAN properly filtered to highlight patterns resulting from the generative process.

Let *I* be a digital image. Following the JPEG pipeline, *I* is divided into non-overlapping blocks of size 8×8. The Discrete Cosine Transform (DCT) is then applied to each block, formally:

$$F[u,v] = \frac{1}{4}C(u)C(v)\left[\sum_{x=0}^{7}\sum_{y=0}^{7} I[x,y]\cos(a)\cos(b)\right] \tag{1}$$

where $a = \frac{(2x+1)u\pi}{16}$, $b = \frac{(2y+1)v\pi}{16}$, $C(u) = \begin{cases} \frac{1}{\sqrt{2}} & u = 0 \\ 1 & u > 0 \end{cases}$ and $C(v) = \begin{cases} \frac{1}{\sqrt{2}} & v = 0 \\ 1 & v > 0 \end{cases}$.

For each 8×8 block, the 64 elements $F[u,v]$ form the DCT coefficients. They are sorted into a zig-zag order starting from the top-left element to the bottom right (Figure 3). The DCT coefficient at position 0 is called DC and represents the average value of pixels in the block. All others coefficients namely AC, corresponds to specific bands of frequencies.
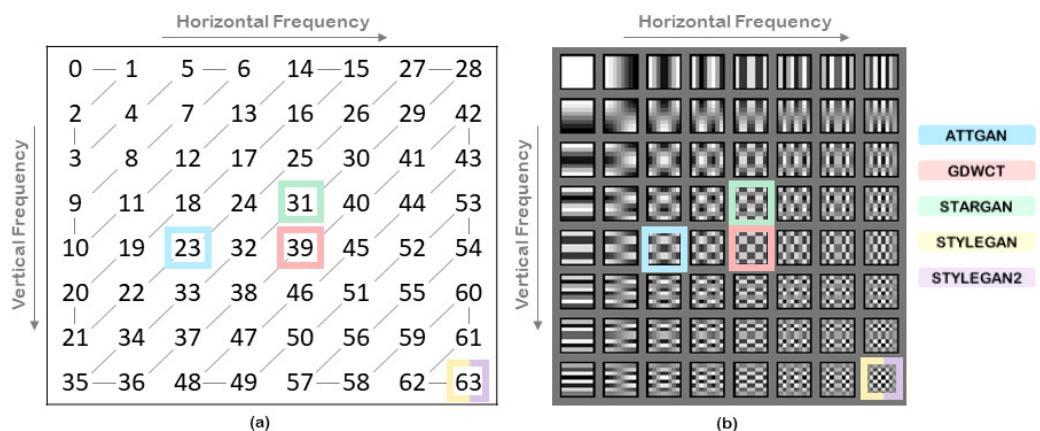

Figure 3. GSF that identify the generative architectures. (**a**) Zig-zag order after DCT transform. (**b**) DCT 8×8 frequencies.

Given all the DCT transformed 8×8 blocks of *I*, it is possible to assess some statistics for each DCT coefficient. By applying evidence reported in [13], the DC coefficient can be

modelled with a Gaussian distribution while the AC coefficients were demonstrated to follow a zero-centred Laplacian distribution described by:

$$P(x) = \frac{1}{2\beta} exp\left(\frac{-|x - \mu|}{\beta}\right) \tag{2}$$

with $\mu = 0$ and $\beta = \sigma/\sqrt{2}$ is the scale parameter where σ corresponds to the standard deviation of the AC coefficient distributions. The proposed approach is partially inspired by [11] where a GMM (Gaussian Mixture Model) over different β values has been properly adopted for scene classification at superordinate level.

An accurate estimation of such β values for each coefficient and involved GAN-engine, is crucial for the purpose achievement. Figure 4 graphically summarizes the statistical trend of the β-values of each involved datasets showing empirically the intrinsic discriminative power devoted to distinguish almost univocally images generated by GAN-engines or picked-up from real datasets. Let $\vec{\beta}_I = \{\beta_{I_1}, \beta_{I_2}, \dots, \beta_{I_N}\}$ with $N = 63$ (DC coefficient is excluded) the corresponding feature vector of the image I. We exploited related statistics on different image-datasets DT_g with g = {*StarGAN, AttGAN, GDWCT, StyleGAN, StyleGAN2, CelebA, FFHQ*}.

For the sake of comparisons in our scenario we evaluated pristine images generated by StarGAN [16], AttGAN [19], GDWCT [20], StyleGAN [17], StyleGAN2 [18], and genuine images extracted by CelebA [21] and FFHQ. E.g., $DT_{StyleGAN}$ represents all the available images generated by StyleGAN engine. For each image-set DT_g let's consider the following representation:

$$\beta_{DT_g} = \begin{pmatrix} \vec{\beta}_1 \\ \vec{\beta}_2 \\ \vdots \\ \vec{\beta}_{|DT_g|} \end{pmatrix} = \begin{pmatrix} \beta_{1,1} & \beta_{1,2} & \cdots & \beta_{1,63} \\ \beta_{2,1} & \beta_{2,2} & \cdots & \beta_{2,63} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{|DT_g|,1} & \beta_{|DT_g|,2} & \cdots & \beta_{|DT_g|,63} \end{pmatrix} \tag{3}$$

where $|.|$ is the number of images in DT_g. For sake of simplicity, in the forthcoming notation all dataset DT_g have been selected to have the same size $|DT_g| = K$. Note that β_{DT_g} have been normalised w.r.t. each column. To extract GSF we first computed the distance among the involved AC distributions modelled by β_{DT_g} for each dataset. We computed a χ^2 distance as follows:

$$\chi^2(\beta_{DT_i}, \beta_{DT_j}) = \sum_{r=1}^{K} \frac{(\beta_{DT_i}[r,c] - \beta_{DT_j}[r,c])^2}{\beta_{DT_j}[r,c]} \quad \text{with} \quad c = 1, \dots, 63 \tag{4}$$

where $i, j \in g, i \neq j$, c is the column which corresponds to the AC coefficient and r are the rows in (3) that represents all $\vec{\beta}_I$ features. The distance $\chi^2(\beta_{DT_i}, \beta_{DT_j})$ is a vector with size of 63. Finally, it is possible to define the GAN Specific Frequency (*GSF*) as follows:

$$GSF_{DT_i, DT_j} = \underset{c}{\text{argmax}} \sum_{r=1}^{K} \frac{\left(\beta_{DT_i}[r,c] - \beta_{DT_j}[r,c]\right)^2}{\beta_{DT_j}[r,c]} \tag{5}$$

where, $i, j \in g, i \neq j$. GSF allow to realize a one-to-one evaluation between image sets.

Practically, the most discriminative DCT frequency is selected among two datasets in a greedy fashion and, as proven by experiments, there is no need to add further computational steps (e.g., frequency ranking/sorting, etc.). In Figure 5c, GSF computed for a set of pair of image-sets, are highlighted just to provide a first toy example where 200 images ($K = 200$) for each set have been employed. Specifically AttGAN, StarGAN and GDWCT were compared with the originating real image-set (CelebA) and for the same reason StyleGAN and StyleGAN2 were compared with FFHQ.
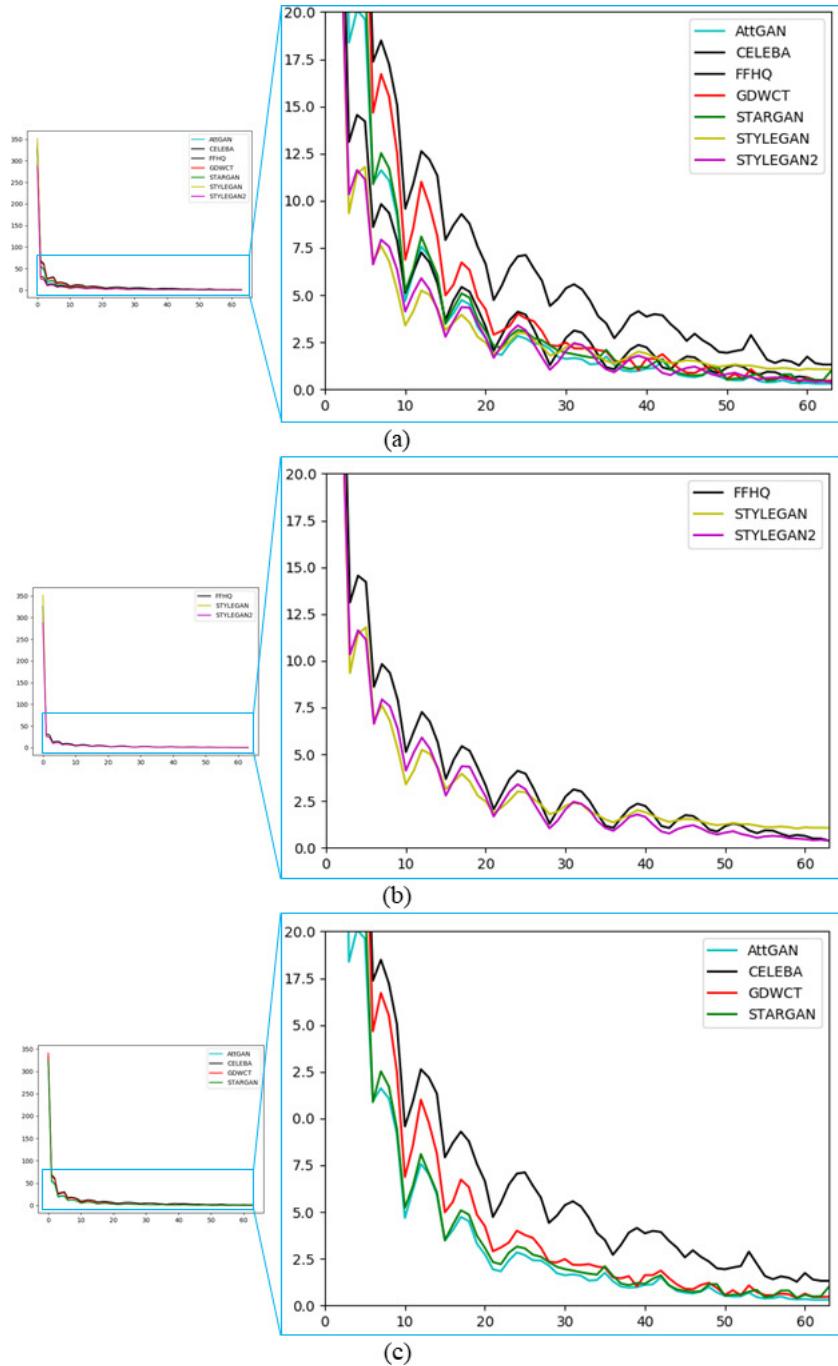
Figure 4. Plot of β statistics of each involved dataset. The average β value for each i-th coefficient is reported. (**a**) Shows the average β trend of all datasets (real and deepfake); (**b**) Shows the average β trend of StyleGAN and StyleGAN2 compared to the real image dataset used for their creation (FFHQ); (**c**) Shows the average β trend of StarGAN, AttGAN and GDWCT compared to the real image dataset used for their creation (CelebA). For each plot, the abscissa axis represents the 64 coefficients of the 8×8 block, while the ordinate axis are the respective inferred β values (in our case the average of the β values computed for all images of the respective datasets).

The β values as described in the experiments, are very discriminative when it comes to deepfake detection. Figure 4 shows the average trend of β of all images from the respective Real and Deepfake datasets. It is interesting to analyze the trend of β of the Deepfake images compared to the statistics of the Real dataset used for the generation task. Figure 4c shows StarGAN, AttGAN, and GDWCT Vs CelebA. All DCT coefficients are sorted in terms of JPEG zigzag order as shown in Figure 3a. It is worth noting that if we consider

even only one of the β values we can roughly establish if an image is a deepfake simply by properly thresholding specific frequencies according to the definition of GSF (Equation (5)). Please note that the discriminative power of the GSFs, even if in some sense they bring energies due to the involved DCT frequencies as demonstrated by the detection results, are not fully dependent by the involved resolution.
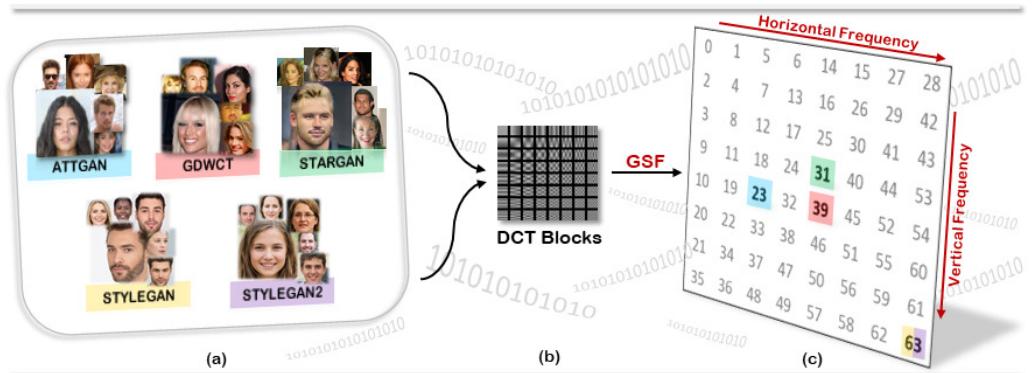
Figure 5. CTF-DCT approach: (**a**) Dataset used for our experiments; (**b**) Discrete Cosine Transform (DCT) of a given image at each 8×8 blocks; (**c**) GAN Specific Frequencies (GSF) that identify involved GAN architectures.

4. Datasets Details

Two datasets of real face images were used for the employed experimental phase: CelebA and FFHQ. Different Deepfake images were generated considering StarGAN, GDWCT, AttGAN, StyleGAN and StyleGAN2 architectures. In particular, CelebA images were manipulated using pre-trained models available on Github, taking into account StarGAN, GDWCT and AttGAN. Images of StyleGAN and StyleGAN2 created through FFHQ were downloaded ad detailed in the following:

- CelebA (CelebFaces Attributes Dataset): a large-scale face attributes dataset with more than 200 K celebrity images, containing 40 labels related to facial attributes such as hair color, gender and age. The images in this dataset cover large pose variations and background clutter. The dataset is composed by 178×218 JPEG images.
- FFHQ (Flickr-Faces-HQ): is a high-quality image dataset of human faces with variations in terms of age, ethnicity and image background. The images were crawled from Flickr and automatically aligned and cropped using dlib [43]. The dataset is composed by high-quality 1024×1024 PNG images.
- StarGAN is able to perform Image-to-image translations on multiple domains using a single model. Using CelebA as real images dataset, every image was manipulated by means of a pre-trained model (https://github.com/yunjey/stargan, accessed on 14 February 2021) obtaining a final resolution equal to 256×256.
- GDWCT is able to improve the styling capability. Using CelebA as real images dataset, every image was manipulated by means of a pre-trained model (https://github.com/WonwoongCho/GDWCT, accessed on 14 February 2021) obtaining a final resolution equal to 216×216.
- AttGAN is able to transfers facial attributes with constraints. Using CelebA as real images dataset, every image was manipulated by means of a pre-trained model (https://github.com/LynnHo/AttGAN-Tensorflow, accessed on 14 February 2021) obtaining a final resolution equal to 256×256.
- StyleGAN is able to transfers semantic content from a source domain to a target domain characterized by a different style. Images have been generated considering FFHQ as dataset in input with 1024×1024 resolution (https://drive.google.com/drive/folders/1uka3a1noXHAydRPRbknqwKVGODvnmUBX, accessed on 14 February 2021).
- StyleGAN2 improves STYLEGAN quality with the same task. Images have been generated considering FFHQ as dataset in input with 1024×1024 resolution (https:

//drive.google.com/drive/folders/1QHc-yF5C3DChRwSdZKcx1w6K8JvSxQi7, accessed on 14 February 2021).

For all the carried out experiments, 3000 Deepfake images for each GAN architecture and 3000 from CelebA and FFHQ were collected and divided into training and test set as will be reported in experimental dedicated Sections. Figure 1 shows some examples of the employed real and Deepfake images.

5. Discussion on GSF

Although differentiating between a Deepfake and a real image could be easy, given the high accuracy values demonstrated by state-of-the-art methods [44], it becomes difficult when the test is carried out on fake images obtained from a specific set of real images: for instance differentiating between FFHQ images and StyleGAN ones, which are generated from FFHQ images, is more difficult than differentiating StyleGAN vs. CelebA images. As a matter of fact, state-of-the-art methods like Fakespotter [28] employs for training, mixed sets of Deepfake and real images. Results are then unbalanced by the extremely-easy-to-spot-difference like CelebA vs. StyleGAN. This can be demonstrated by means of *GSF* analysis.

Through *GSF* it is possible to perform a one-to-one test between sets of images. This was carried out specifically for the harder case as described before: taking 200 images for each set, *GSF* was calculated for each pair of image sets, whose values obtained are shown in Figure 3b. In particular, AttGAN, StarGAN and GDWCT were compared with the starting real images (CelebA) and for the same reason StyleGAN and StyleGAN2 were compared with FFHQ.

Torralba et al. [45] demonstrated that scenes semantic-visual components are captured precisely with analogous statistics on spectral domain used also to build fast classifiers of scenes [11]. In this sense, the comparison between images that represent close-ups of faces showing the some overall visual structure raising extremely similar statistical characteristics of AC coefficients and their β values. This allows the *GSF* analysis to focus on the unnatural anomalies introduced by the convolutional generative process typical of Deepfakes. To demonstrate the discriminative power of the *GSF* a simple binary classifier (logistic regression) was trained using the β (e.g., that corresponds the set of values of a given column/coefficient in Equation (3)) of the corresponding *GSF* as unique feature.

For all the experiments carried out, the number of collected images has been equally set considering $K = 3000$. In particular the classifier was trained using only the 10% of the entire dataset, while the remaining part was used as test set. For each binary classification test, the simple classification solution obtained the results shown in Figure 6. Results demonstrated that Deepfakes are easily detectable by just looking at the β value of the *GSF* for that specific binary test. This is empirically found to be discriminative (wider range of values) than expected on natural images, given the semantic context of facial images. This finding is what state-of-the-art is exploiting with much more complex and computational intensive solutions. For instance, Fakespotter [28], at a first step compares real against fake images and finds these unnatural frequencies with an ad-hoc trained CNN. As a matter of fact, frequencies found are different for forgeries made with Photoshop which certainly do not bear traces of convolution and for this reason they are easily discriminated from the Deepfake images.

As already stated, the combination of different resolution and frequency bands image-sets is the major problem encountered in the state of the art methods, while the most problematic issue is differentiating the original images from the transformed Deepfake. Let's take into account FFHQ vs. STYLEGAN: a task in which even the human being had difficulties [37]. Applying GSF analysis among all involved proper datasets, we obtain impressive generalization results as reported in Figure 7. Further demonstration of the importance of the *GSF* will be visual. In addition to the anomalies visually identified in Figure 8, in Guarnera et al. [4] the authors already identified some strange components in the Fourier spectrum. Given an image from a specific image-set, after having computed the

GSF (Figure 5), it is possible for sake of explainability, to apply the following amplification process: to multiply in the DCT domain each DCT coefficient different from the *GSF* by a value k_1 (with $0 < k_1 \leq 1$) while the coefficient corresponding to the *GSF* by a value k_2 (with $k_2 > 1$). Figure 8 shows an example of such amplification procedure with $k_1 = 0.1$ and $k_2 = 100$. This operation will create an image where the *GSF* is amplified. Figure 8 shows that the original Fourier Spectrum and the amplified one share the same abnormal frequency appearance. Thus, *GSF* becomes an explanation of those anomalies with a clear boost of forensics analysis.

	StarGAN		AttGAN		GDWCT		StyleGAN		StyleGAN2	
	GSF	Acc(%)	GSF	Acc(%)	GSF	Acc(%)	GSF	Acc(%)	GSF	Acc(%)
CelebA	31	94%	23	89%	39	91%				
FFHQ							63	99%	63	99%

Figure 6. Average Accuracy results (%) obtained for the binary classification task employing only the *GSF*. 700 images were employed for testing, 200 images for training, 5-fold cross validated, classes are balanced.

	StarGAN		AttGAN		GDWCT		StyleGAN		StyleGAN2	
	GSF	Acc(%)	GSF	Acc(%)	GSF	Acc(%)	GSF	Acc(%)	GSF	Acc(%)
CelebA	31	94%	23	89%	39	91%	2	100%	2	100%
FFHQ	63	96%	2	95%	2	94%	63	99%	63	99%
StarGAN			63	100%	63	95%	50	100%	63	100%
AttGAN					7	72%	62	99%	2	95%
GDWCT							2	98%	2	99%
StyleGAN									54	99%

Figure 7. GSF and classification accuracy results (%) obtained for each binary classification task.

It has to be noted that the *GSF* approach described in this section is a great instrument to white-box GAN-generated image processing. A *GSF* is able to identify a set of GAN-generated images. On the other hand, it is not enough to properly being employed in the wild or against fakes not generated by neural approaches. For this reason, in the following section, we "finalize" the approach by presenting a more robust and complete feature vector but, on the other hand, we will lose explainability.

Finalizing the CTF Approach

Given the ability of the *GSF* to make one-to-one comparisons even between image-sets of GANs it is possible to use it to resolve further discrimination issues. Figure 5 shows that the two StyleGANs actually have the same $GSF_{StyleGAN,FFHQ} = GSF_{StyleGAN2,FFHQ} = 63$, while $GSF_{StyleGAN,StyleGAN2} = 54$ was obtained (Figure 7). Also upon this *GSF* it is possible to train a classifier that quickly obtains an accuracy value in the binary test between StyleGAN and StyleGAN2 close to 99%.

The *GSF* analysis can be exploited to give explainability to unusual artifacts and behaviors that appear in the Fourier domain of Deepfakes. Obviously, using only the corresponding β to *GSF* can be reductive for a scenario in the wild and this is the reason why the CTF approach will be completed by means of a robust classifier which will be outlined in the next section. Instead of using only the corresponding β to the *GSF*, it will employ a feature vector with all 63 β, consequently used as input to a Gradient Boosting classifier [46] and tested in a noisy context that includes a number of plausible attacks on the images under analysis. Gradient Boosting was selected as the best classifier for data and

the following hyper-parameters were selected by means of a 10% of the dataset employed as validation set. We selected the following hyper parameters: *number-of-estimators* = 100, *learning-rate* = 0.6, $max_{depth} = 2$.

The robust classifier thus created, fairly identify the most probable GAN from which the image has been generated, providing hints for "visual explainability". By considering the growing availability of Deepfakes to attack people reputation such aspects become fundamental to assess and validate forensics evidence. All the employed data and code will be publicly available after the review process at a public link.

| (a) | (b) | (c) |
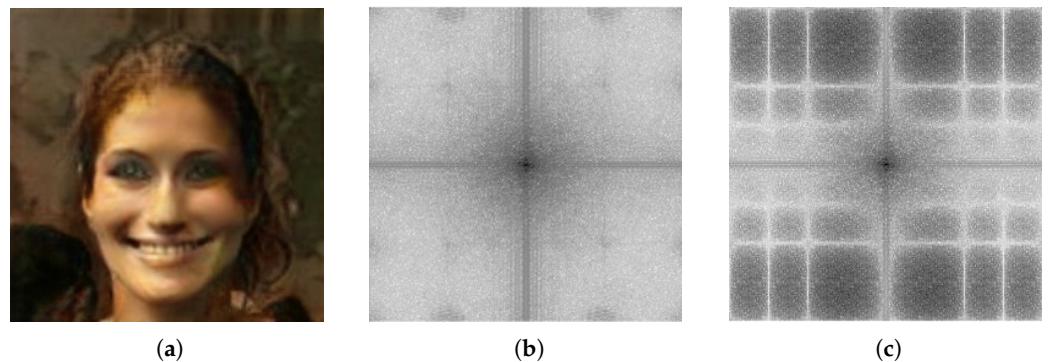
Figure 8. Abnormal frequencies inspection. (**a**) Image example from the StarGAN dataset; (**b**) Fourier Spectra of the input image (**a**); (**c**) Abnormal frequency shown by means of *GSF* amplification.

6. Experimental Results

In this section experimental results are presented. Primarily, to finalize the CTF approach, a robust classifier was trained and tested by means of several attacks on images and consequently tested in a different scenario, namely the FaceForensics++ dataset of Deepfake videos [30]. The above-mentioned deepfake dataset is used only during the testing phase to classify real Vs deepfake. 3000 real and fake images were collected to train the "robust classifier" for the validation, employing only the 10% of the entire dataset while the remaining part was used as test set. Multiple attack types augmented the dataset; Figure 9 provides examples of images after each attack. Cross-validation was carried out.

6.1. Testing with Noise

All the images collected in the corresponding *DT* have been put through different kinds of attacks as addition of a random size rectangle, position and color, Gaussian blur, rotation and mirroring, scaling and various JPEG Quality Factor compression (QF), in order to demonstrate the robustness of the CTF approach.

As shown in Table 1 this type of attacks do not destroy the *GSF* obtaining high accuracy values.

Gaussian Blur applied with different kernel sizes (3×3, 9×9, 15×15) could destroy different main frequencies in the images. This filtering preserves low frequencies by almost totally deleting the high frequencies, as the kernel size increases. It is possible to see in Table 1, that the accuracy decreases at increasing of the kernel size. This phenomenon, is particularly visible for images generated by AttGAN, GDWCT and StarGAN which have the lowest resolution.
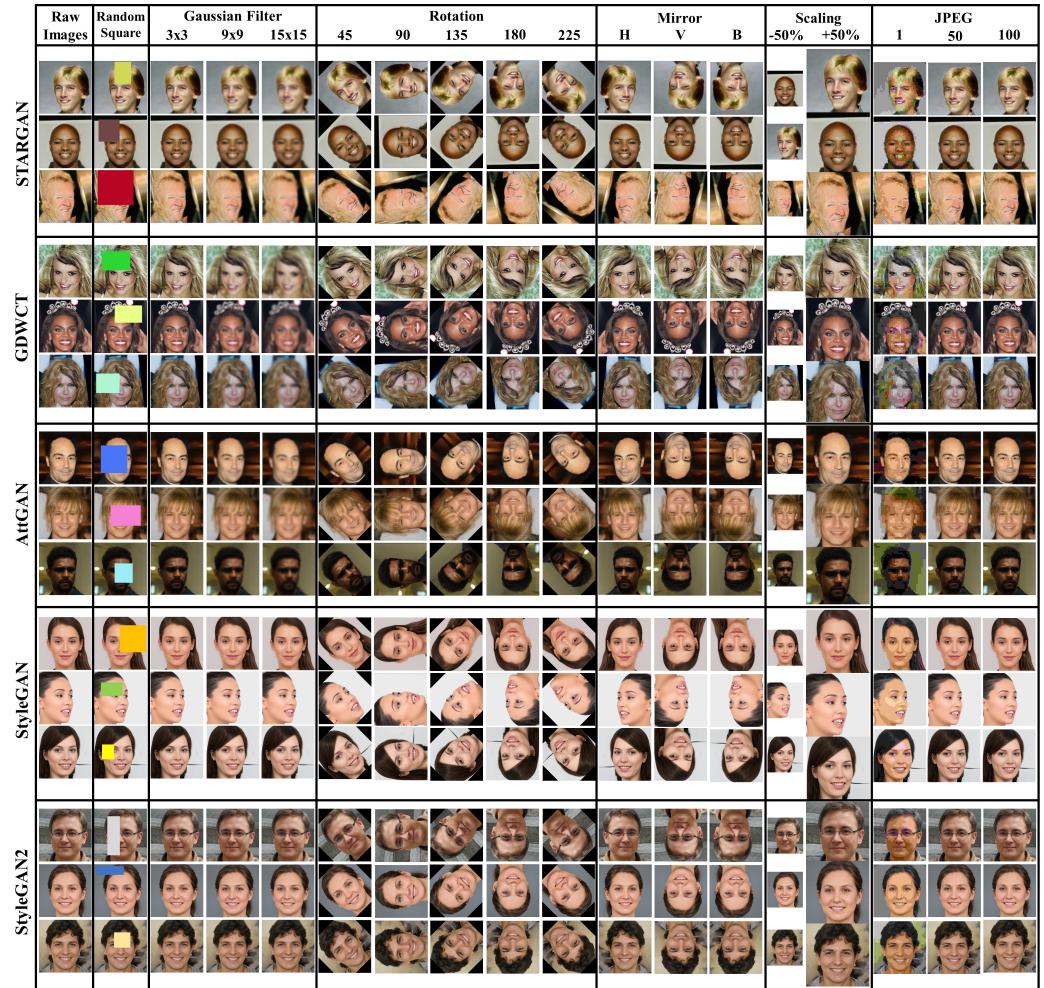
Figure 9. Examples of ATTGAN, GDWCT, STARGAN, STYLEGAN, STYLEGAN2 images in which we applied different attacks: Random Square, Gaussian Blur, Rotation, Mirror, Scaling and JPEG Compression. They were also applied in the real dataset (CelebA and FFHQ).

Several degrees of rotation $(45, 90, 135, 180, 255)$ were considered since they can modify the frequency components of the images. Rotations with angles of 90, 180, and 270 do not alter the frequencies because the $[x,y]$ pixels are simply moved to the new $[x',y']$ coordinates without performing any interpolation function, obtaining high values of detection accuracy. On the other hand, when considering different degrees of rotation, it is necessary to interpolate the neighboring pixels to get the missing ones. In this latter case, new information is added to the image that can affect the frequency information. In fact, considering rotations of 45, 135, 225 degree, the classification accuracy values decrease; except for the two StyleGANs for the same reason described for the Gaussian filter (i.e., high resolution).

The mirror attack reflects the image pixels along one axis (horizontal, vertical and both). This does not alter image frequencies, obtaining extremely high accuracy detection values.

The resizing attacks equal to -50% of resolution causes a loss of information, hence, already small images tend to totally lose high-frequency components presenting a behavior similar to low-pass filtering; in this case accuracy values are inclined to be low. Vice versa, a resizing of $+50\%$ doesn't destroy the main frequencies obtaining a high classification accuracy values.

Table 1. Percentage of Precision, Recall, F1-score and accuracy obtained in the robustness test. "Raw Images" shows the results without the attack process. For the "Real" column the CelebA and FFHQ datasets were considered. Different attacks were carried out in the datasets: Random square; Gaussian filter with different kernel size ($3 \times 3, 9 \times 9, 15 \times 15$); Rotations with $degree = \{45, 90, 135, 180, 225\}$; Mirror with Horizontal (H), Vertical (V) and Both (B) ways; Scaling ($+50\%, -50\%$); JPEG Compression with different Quality Factor ($QF = \{1, 50, 100\}$).

		Real			AttGAN			GDWCT			StarGAN			StyleGAN			StyleGAN2			Overall
		Prec	Rec	F1	Prec	Rec	F1	Prec	Rec	F1	Prec	Rec	F1	Prec	Rec	F1	Prec	Rec	F1	Accuracy
Raw Images		99	97	98	99	100	99	98	98	98	99	100	100	99	98	99	98	100	99	99
Random Square		98	94	96	90	96	93	92	89	91	100	98	99	98	99	98	99	99	99	96
Gaussian Filter	3×3	98	95	96	83	88	86	89	92	91	92	86	89	97	98	98	99	99	99	93
	9×9	98	99	98	62	59	60	70	79	74	59	53	56	99	98	99	98	99	98	81
	15×15	100	97	98	58	64	61	72	64	68	55	53	54	98	99	98	95	100	97	80
Rotation	$45°$	97	93	95	85	82	83	92	98	95	84	84	84	97	99	98	99	98	98	92
	$90°$	98	99	98	95	99	97	98	93	95	100	99	99	99	98	98	99	99	99	98
	$135°$	95	96	96	85	83	84	97	94	96	83	86	85	96	95	96	96	97	97	92
	$180°$	98	94	96	95	100	97	97	95	96	99	100	99	98	99	99	100	99	99	98
	$225°$	96	95	95	88	85	87	96	96	96	86	89	88	96	97	97	97	97	97	93
Mirror	H	99	96	98	99	100	99	98	99	98	99	100	99	99	99	99	100	100	100	99
	V	99	96	98	99	100	99	97	99	98	99	100	100	99	99	99	100	100	100	99
	B	99	94	97	98	100	99	97	99	98	99	100	100	99	99	99	100	100	100	99
Scaling	$+50\%$	99	98	99	94	95	95	95	93	94	98	99	99	99	99	99	99	100	99	97
	-50%	74	95	84	77	66	71	74	72	73	81	77	79	82	85	84	90	81	85	80
JPEG	1	78	69	73	63	65	64	59	67	63	59	57	58	78	83	80	84	80	82	70
	50	93	95	94	98	99	98	87	80	83	84	89	86	88	88	88	90	89	89	90
	100	99	99	99	100	99	99	98	98	98	99	100	99	99	99	99	99	99	99	99

Finally, different JPEG compression quality factors were applied ($QF = 1, 50, 100$). As expected in Table 1, a compression with $QF = 100$ does not affect the results. The overall accuracy begins to be affected as the QF decreases, among other things, destroying the DCT coefficients. However, at $QF = 50$ the mid-level frequencies are still preserved and the results maintain a high level of accuracy. This is extremely important given that this level of QF is employed by the most common social platforms such as Whatsapp or Facebook, thus demonstrating that the CTF approach is extremely efficient in real-world scenarios.

6.2. Comparison and Generalization Tests

The CTF approach is extremely simple, fast, and demonstrates a high level of accuracy even in real-world scenarios. In order to better understand the effectiveness of the technique, a comparison with state-of-the-art methods was performed and reported in this section. The trained robust classifier was compared to the most recent work in the literature and in particular Zhang et al. [5] (AutoGAN), Wang et al. [28] (FakeSpotter) and Guarnera et al. [38] (Expectation-Maximization) were considered for the use of a few GAN architectures in common with the analysis performed in this paper: StyleGAN, StyleGAN2, StarGAN. Table 2 shows that the CTF approach achieves the best results with an unbeatable accuracy of 99.9%.

Table 2. Comparison with state-of-the-art methods [5,28,38]. Classification of Real images (CelebA and FFHQ) vs. Deepfake images. Accuracy values (%) of each classification task are reported.

	StarGAN	**StyleGAN**	**StyleGAN2**
AutoGAN [5]	65.6	79.5	72.5
FakeSpotter [28]	88	99.1	91.9
EM [38]	90.55	99.48	99.64
CTF (our)	99.9	100	100

Another comparison was made on the detection of StyleGAN and StarGAN with respect to [38,44]. The obtained results are shown in the Table 3 in which the average classification values of each classification task are reported.

A specific discussion is needed for testing the FaceForensics++ dataset [30] which is a challenging dataset of fake videos of people speaking and acting in several contexts. The fake videos were created by means of four different techniques (Face2Face [47] among them) on videos taken from YouTube. By means of OpenCV's face detectors, cropped images of faces were taken from fake videos of FF++ (with samples from all four categories, at different compression levels) and a dataset of 3000 images with different resolutions ($min_{resolution} = 162 \times 162$ px, $max_{resolution} = 895 \times 895$ px). The CTF approach was employed to construct the β feature vector computed on the DCT coefficients and the robust classifier (trained in the Section 6.1), was used for binary classification in order to perform this "in the wild" test. We emphasize that the latter datasets were only used in the testing phase with the robust classifier. Since the classifier detected FaceForensics++ images as well as StyleGAN images, we also tried to calculate the *GSF* by comparing FaceForensics++ images with FFHQ obtaining a value of 61 which is extremely close to the *GSF* of Style-GANs. This leads to the explanation that the *GSF*s are also dependent not only on the generative process but also to the reenactment phase done on images. The reenactment is done analytically in Face2Face and trained in StyleGANs as a part of the model (similarly to Face2face but as a cost function).

The results obtained on FaceForensics++ are reported in Table 3 showing how the CTF approach is an extremely simple and fast method capable of beating the state-of-the-art even on datasets on which it has not been trained and being able to catch not only convolutional artefacts but also those created by reenactment phase which is an important part for the most advanced Deepfake techniques.

Table 3. Comparison with state-of-the-art methods [38,44]. Classification of Real images (CelebA and FFHQ) vs. Deepfake images. The CTF approach was tested and compared also considering the dataset of Deepfake video's FaceForensics++ (FF++). Average Precision values (%) of each classification task are reported.

	StyleGAN	StarGAN	FF++
Wang [44]	96.3	100	98.2
EM [38]	99	93	98.8
CTF (our)	99.9	99.9	99.9

7. Conclusions

In this paper, the CTF approach was presented as a detection method for Deepfake images. The approach is extremely fast, explainable, and does not need intense computational power for training. By exploiting and analyzing the overall statistics of the DCT coefficients it is possible to discriminate among all known GAN's by means of the GAN Specific Frequency band (*GSF*). The *GSF* has many interesting properties demonstrated through empirical and visual analysis; among others it is possible to give some explainability to the underlying generation process, especially for forensics purposes. In order to achieve higher accuracy values, all frequency bands must be taken into account and the CTF approach is finalized by means of a G-boost classifier which demonstrated to be robust to attacks and able to generalize even in a dataset of Deepfake videos (FaceForensics++) not used during training. Further investigation could be carried out on *GSF* frequencies in order to detect not only GAN artefacts but also information coming from the reenactment phase. Finally, the CTF approach could give useful suggestions for the *GSF* analysis (explainability, etc.) in new scenarios with more challenging modalities (attribute manipulation, expression swap, etc.) and media (audio,video).

Author Contributions: Conceptualization, O.G. and L.G.; Data curation, O.G. and L.G.; Formal analysis, O.G. and L.G.; Investigation, O.G. and L.G.; Methodology, O.G. and L.G.; Resources, O.G. and L.G.; Software, L.G.; Supervision, S.B.; Validation, L.G. and S.B.; Writing—original draft, O.G. and L.G.; Writing—review and editing, O.G., L.G and S.B. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: https://iplab.dmi.unict.it/mfs/Deepfakes/PaperGANDCT-2021/, accessed on 14 February 2021.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 8–13 December 2014; pp. 2672–2680.
2. Vaccari, C.; Chadwick, A. Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Soc. Media+ Soc.* **2020**, *6*, 2056305120903408. [CrossRef]
3. Guarnera, L.; Giudice, O.; Battiato, S. DeepFake Detection by Analyzing Convolutional Traces. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 666–667.
4. Guarnera, L.; Giudice, O.; Nastasi, C.; Battiato, S. Preliminary Forensics Analysis of DeepFake Images. In Proceedings of the 2020 AEIT International Annual Conference (AEIT), Catania, Italy, 23–25 September 2020; pp. 1–6. [CrossRef]
5. Zhang, X.; Karaman, S.; Chang, S.F. Detecting and simulating artifacts in gan fake images. In Proceedings of the 2019 IEEE International Workshop on Information Forensics and Security (WIFS), Delft, The Netherlands, 9–12 December 2019; pp. 1–6.
6. Oliva, A.; Torralba, A. Modeling the shape of the scene: A holistic representation of the spatial envelope. *Int. J. Comput. Vis.* **2001**, *42*, 145–175. [CrossRef]

7. Xu, K.; Qin, M.; Sun, F.; Wang, Y.; Chen, Y.K.; Ren, F. Learning in the frequency domain. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 1740–1749.
8. Xu, Z.Q.J.; Zhang, Y.; Xiao, Y. Training behavior of deep neural network in frequency domain. In *International Conference on Neural Information Processing*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 264–274.
9. Yin, D.; Gontijo Lopes, R.; Shlens, J.; Cubuk, E.D.; Gilmer, J. A Fourier Perspective on Model Robustness in Computer Vision. In Proceedings of the Advances in Neural Information Processing Systems, Vancouver, BC, Canada, 8–14 December 2019; Volume 32, pp. 13276–13286.
10. Rahaman, N.; Baratin, A.; Arpit, D.; Draxler, F.; Lin, M.; Hamprecht, F.; Bengio, Y.; Courville, A. On the spectral bias of neural networks. In Proceedings of the International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019; pp. 5301–5310.
11. Farinella, G.M.; Ravì, D.; Tomaselli, V.; Guarnera, M.; Battiato, S. Representing scenes for real-time context classification on mobile devices. *Pattern Recognit.* **2015**, *48*, 1086–1100. [CrossRef]
12. Ravì, D.; Bober, M.; Farinella, G.M.; Guarnera, M.; Battiato, S. Semantic segmentation of images exploiting DCT based features and random forest. *Pattern Recognit.* **2016**, *52*, 260–273. [CrossRef]
13. Lam, E.Y.; Goodman, J.W. A mathematical analysis of the DCT coefficient distributions for images. *IEEE Trans. Image Process.* **2000**, *9*, 1661–1666. [CrossRef] [PubMed]
14. Tolosana, R.; Vera-Rodriguez, R.; Fierrez, J.; Morales, A.; Ortega-Garcia, J. Deepfakes and beyond: A survey of face manipulation and fake detection. *arXiv* **2020**, arXiv:2001.00179.
15. Verdoliva, L. Media Forensics and DeepFakes: An overview. *IEEE J. Sel. Top. Signal Process.* **2020**, *14*, 910–932. [CrossRef]
16. Choi, Y.; Choi, M.; Kim, M.; Ha, J.W.; Kim, S.; Choo, J. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 8789–8797.
17. Karras, T.; Laine, S.; Aila, T. A style-based generator architecture for generative adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 4401–4410.
18. Karras, T.; Laine, S.; Aittala, M.; Hellsten, J.; Lehtinen, J.; Aila, T. Analyzing and improving the image quality of stylegan. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 8110–8119.
19. He, Z.; Zuo, W.; Kan, M.; Shan, S.; Chen, X. Attgan: Facial attribute editing by only changing what you want. *IEEE Trans. Image Process.* **2019**, *28*, 5464–5478. [CrossRef] [PubMed]
20. Cho, W.; Choi, S.; Park, D.K.; Shin, I.; Choo, J. Image-to-image translation via group-wise deep whitening-and-coloring transformation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 10639–10647.
21. Liu, Z.; Luo, P.; Wang, X.; Tang, X. Deep learning face attributes in the wild. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 3730–3738.
22. Langner, O.; Dotsch, R.; Bijlstra, G.; Wigboldus, D.H.; Hawk, S.T.; Van Knippenberg, A. Presentation and validation of the Radboud Faces Database. *Cogn. Emot.* **2010**, *24*, 1377–1388. [CrossRef]
23. Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2223–2232.
24. Lee, H.Y.; Tseng, H.Y.; Huang, J.B.; Singh, M.; Yang, M.H. Diverse image-to-image translation via disentangled representations. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 35–51.
25. Wilber, M.J.; Fang, C.; Jin, H.; Hertzmann, A.; Collomosse, J.; Belongie, S. Bam! the behance artistic media dataset for recognition beyond photography. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 1202–1211.
26. Durall, R.; Keuper, M.; Pfreundt, F.J.; Keuper, J. Unmasking deepfakes with simple features. *arXiv* **2019**, arXiv:1911.00686.
27. Karras, T.; Aila, T.; Laine, S.; Lehtinen, J. Progressive growing of gans for improved quality, stability, and variation. *arXiv* **2017**, arXiv:1710.10196.
28. Wang, R.; Ma, L.; Juefei-Xu, F.; Xie, X.; Wang, J.; Liu, Y. Fakespotter: A simple baseline for spotting AI-synthesized fake faces. *arXiv* **2019**, arXiv:1909.06122.
29. Liu, M.; Ding, Y.; Xia, M.; Liu, X.; Ding, E.; Zuo, W.; Wen, S. Stgan: A unified selective transfer network for arbitrary image attribute editing. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 3673–3682.
30. Rossler, A.; Cozzolino, D.; Verdoliva, L.; Riess, C.; Thies, J.; Nießner, M. Faceforensics++: Learning to detect manipulated facial images. In Proceedings of the IEEE International Conference on Computer Vision, Seoul, Korea, 27–28 October 2019; pp. 1–11.
31. Li, Y.; Yang, X.; Sun, P.; Qi, H.; Lyu, S. Celeb-DF: A Large-scale Challenging Dataset for DeepFake Forensics. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 3207–3216.
32. Jain, A.; Majumdar, P.; Singh, R.; Vatsa, M. Detecting GANs and Retouching based Digital Alterations via DAD-HCNN. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 672–673.

33. Ledig, C.; Theis, L.; Huszár, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z.; et al. Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4681–4690.

34. Radford, A.; Metz, L.; Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv* **2015**, arXiv:1511.06434.

35. Pathak, D.; Krahenbuhl, P.; Donahue, J.; Darrell, T.; Efros, A.A. Context encoders: Feature learning by inpainting. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2536–2544.

36. Liu, Z.; Qi, X.; Torr, P.H. Global texture enhancement for fake face detection in the wild. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 8060–8069.

37. Hulzebosch, N.; Ibrahimi, S.; Worring, M. Detecting CNN-Generated Facial Images in Real-World Scenarios. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 642–643.

38. Guarnera, L.; Giudice, O.; Battiato, S. Fighting Deepfake by Exposing the Convolutional Traces on Images. *IEEE Access* **2020**, *8*, 165085–165098. [CrossRef]

39. Moon, T.K. The expectation-maximization algorithm. *IEEE Signal Process. Mag.* **1996**, *13*, 47–60. [CrossRef]

40. Jing, X.Y.; Zhang, D. A face and palmprint recognition approach based on discriminant DCT feature extraction. *IEEE Trans. Syst. Man Cybern. Part B (Cybern.)* **2004**, *34*, 2405–2415. [CrossRef] [PubMed]

41. Thai, T.H.; Retraint, F.; Cogranne, R. Camera model identification based on DCT coefficient statistics. *Digit. Signal Process.* **2015**, *40*, 88–100. [CrossRef]

42. Lam, E.Y. Analysis of the DCT coefficient distributions for document coding. *IEEE Signal Process. Lett.* **2004**, *11*, 97–100. [CrossRef]

43. King, D.E. Dlib-ml: A Machine Learning Toolkit. *J. Mach. Learn. Res.* **2009**, *10*, 1755–1758.

44. Wang, S.Y.; Wang, O.; Zhang, R.; Owens, A.; Efros, A.A. CNN-generated images are surprisingly easy to spot... for now. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; Volume 7.

45. Zhou, B.; Zhao, H.; Puig, X.; Fidler, S.; Barriuso, A.; Torralba, A. Scene parsing through ade20k dataset. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 633–641.

46. Bishop, C.M. *Pattern Recognition and Machine Learning*; Springer: Berlin/Heidelberg, Germany, 2006.

47. Thies, J.; Zollhofer, M.; Stamminger, M.; Theobalt, C.; Nießner, M. Face2face: Real-time face capture and reenactment of rgb videos. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2387–2395.