New Finding and Unified Framework for Fake Image Detection

Xin Deng, *Member, IEEE*, Bihe Zhao, Zhenyu Guan, *Member, IEEE*, and Mai Xu, *Senior Member, IEEE*

***Abstract*—Recently, fake face images generated by generative adversarial network (GAN) have been widely spread in social networks, raising serious social concerns and security risks. To identify the fake images, the top priority is to find what properties make the fake images different from the real images. In this letter, we reveal an important observation about real/fake images, i.e., the GAN generated fake images contain stronger non-local self-similarity than the real images. Motivated by this observation, we propose a simple yet effective non-local attention based fake image detection network, namely NAFID, to distinguish GAN generated fake images from real images. Specifically, we develop a non-local feature extraction (NFE) module to extract the non-local features of the real/fake images, followed by a multi-stage classification module to distinguish the images with the extracted non-local features. Experimental results on various datasets demonstrate the superiority of our NAFID over state-of-the-art (SOTA) face forgery detection methods. More importantly, since the NFE module is independent from classification, we can plug it into any other forgery detection models. The results show that the NFE module can consistently improve the detection accuracy of other models, which verifies the universality of the proposed method.**

***Index Terms*—Fake face detection, generative neural network, non-local similarity.**

I. Introduction

RECENTLY, the face forgery techniques [1], [2], [3], [4], [5], [6], [7], [8] have developed rapidly due to the unprecedented success of generative adversarial network (GAN) [9]. The generated images are so realistic that can even fool the human beings. The high realism of fake faces makes them easily be used for illegal purposes, which has posed serious threats to the social security. To tackle these security issues, it is highly desirable to develop efficient face forgery detection methods. To detect the fake images, the top priority is to find the inherent differences between the fake and real images. Towards this goal, some works [11], [12], [13], [14], [15] focus on detecting the forgery clues in facial features, e.g. head poses, blinking patterns or skin colors. However, these methods can only identify the fake images with obvious changes, but fail to detect those fake images with mild distortion. Recent works [16], [17], [18], [19], [20], [21] have noticed this problem and pay more attention to the low-level subtle details, such as texture information. In addition, there are also some works [22], [23], [24] using prior knowledge from frequency domain to identify the fake face images.

Different from the existing works which rely on carefully designed networks to detect the forgery artifacts, we reveal an important observation about the difference between GAN generated fake faces and real faces. As we know, the non-local self-similarity is an important property of natural images [25], which has been used as prior knowledge to solve many inverse image restoration tasks [26], [27], [28]. Since natural images have non-local property, an intuitive question is whether the GAN generated fake images have the same property. In this letter, through deliberate experiments, we have an important observation, i.e., the fake images contain stronger non-local self-similarity property than the real images. The underlying reason is that the whole or part of the GAN generated images originate from a noise vector with a limited length, which determines that the diversity of the fake images cannot be as rich as the real images. A simple example is shown in Fig. 1, in which we add the same Gaussian noise to the fake/real images and use block-matching and 3D filtering (BM3D) algorithm [10] to denoise the images. Since BM3D achieves denoising by purely exploring the non-local similar patches in the image, the image with stronger non-local similarity tends to have better denoising result. As can be seen, the fake images have higher peak signal-to-noise ratio (PSNR) values than the real images. This observation applies to most GAN based forgery models, including face generation, face swapping, and face editting.

Based on the above observation, we propose a simple yet effective non-local attention based fake image detection network, namely NAFID. The core design of the NAFID network is a non-local feature extraction (NFE) module, where we develop a multi-head non-local block (MNLB) to extract the non-local features from the fake and real images. Then, a multi-stage classification sub-net is used to distinguish the fake and real images based on the extracted non-local features. To evaluate the effectiveness of our framework, we conduct exhaustive experiments on fake face images generated by different generative models. The experimental results show that the proposed method outperforms state-of-the-art face forgery detection methods on various datasets. In addition, since the NFE module is independent from classification, we can incorporate it into other forgery detection models, which has been demonstrated to further improve the detection accuracy of these models.

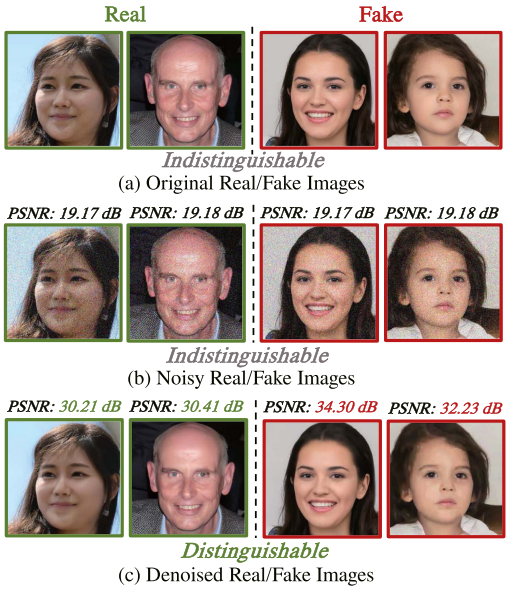
Manuscript received 1 November 2022; revised 11 January 2023; accepted 31 January 2023. Date of publication 9 February 2023; date of current version 15 February 2023. This work was supported in part by NSFC under Grant 62001016 and in part by Young Elite Scientists Sponsorship Program under Grant 2022QNRC001 by CAST. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Xiangui Kang. *(Xin Deng and Bihe Zhao contributed equally to this work.) (Corresponding author: Zhenyu Guan.)*

Xin Deng, Bihe Zhao, and Zhenyu Guan are with the School of Cyber Science and Technology, Beihang University, Beijing 100191, China (e-mail: cindydeng@buaa.edu.cn; bihezhao@buaa.edu.cn; guanzhenyu@buaa.edu.cn).

Mai Xu is with the School of Electronic and Information Engineering, Beihang University, Beijing 100191, China (e-mail: maixu@buaa.edu.cn).

The software code is available through the link https://github.com/BiheZhao/NAFID.

Digital Object Identifier 10.1109/LSP.2023.3243770



(a) Original Real/Fake Images

(b) Noisy Real/Fake Images

(c) Denoised Real/Fake Images

Fig. 1. Illustration about our observation for real/fake images. (a) shows the original real and fake images, which are not distinguishable; (b) shows the real/fake images with Gaussian noise ($\sigma = 30$) added, which are also not distinguishable; (c) shows the denoised real and fake images by BM3D [10], which can be easily distinguishable by the PSNR value.

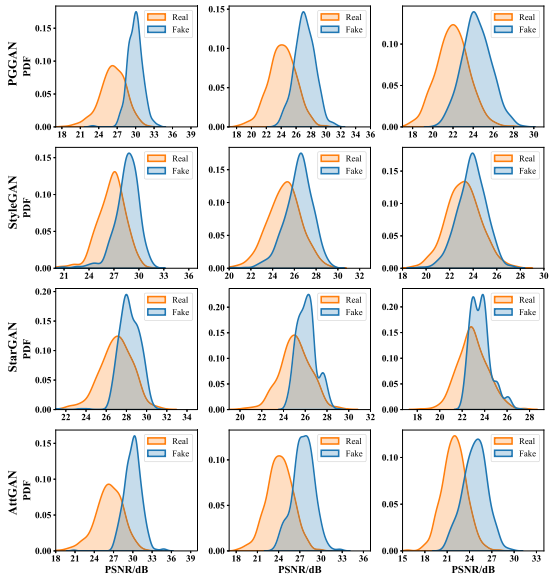


Fig. 2. The probability density function (PDF) of the PSNR for real/fake images denoised by BM3D. The three columns from left to right correspond to the noise level $\sigma = 30, 50$ and 70, respectively.

II. Proposed Method

A. Motivation

To distinguish the GAN generated fake faces from the real faces, we first analyze how these fake images are generated by GAN. There are mainly two types of face forgery methods: face generation and face manipulation. Face generation is a task to synthesize face images from a noise vector [1], [2], [3]. Specifically, the GAN-based face generation methods use a generator to synthesize fake images, and a discriminator to make the generated images indistinguishable. Face manipulation refers to the modification of a real face image, such as face swapping [4], [5] and facial attribute modification [6], [7], [8]. For face manipulation, there is also a controllable noise vector to edit the attributes of the source image. In summary, for both forgery types, the whole or part of the fake images originate from a noise vector with a limited length. In addition, in the process of image generation, there exist several upscaling layers, which enlarge the feature size through either interpolation or copying. All these make the diversity of the generated fake image not as rich as the real image. Therefore, the GAN generated faces could demonstrate more structural patterns and stronger non-local self-similarity than real faces.

Unfortunately, there is no quantitative metric that can directly measure the non-local self-similarity of an image. To tackle this issue, we design an experiment based on BM3D image denoising to evaluate the non-local self-similarity. As we know, the BM3D algorithm achieves image denoising through exploiting the non-local similar patches. The stronger non-local self-similarity of the image leads to the better image denoising result. There are six fake face datasets involved in the experiment, which are generated by the following forgery methods: PGGAN [1], StyleGAN [2], StyleGAN2 [3] belonging to the face generation category, StarGAN v2 [7], Deepfakes [4] and AttGAN [8] belonging to the face manipulation category. We use FaceForensics++ [29] as the Deepfakes dataset. For other methods, the real images are from CelebA [30] and FFHQ [2] datasets. The Gaussian noise with different levels $\sigma = 30, 50$ and 70 are added to the original clean images. After that, we use the BM3D algorithm to denoise both the noisy fake and real images.

Fig. 2 plots the probability density function (PDF) of PSNR values after denoising for both real and fake images. From Fig. 2, we can obtain two intriguing observations: 1) with the same noise level, the denoised fake faces tend to have higher PSNR values than the real faces, and this phenomenon keeps for all the six datasets in different noise levels. This observation reflects that forged faces demonstrate stronger non-local self-similarity than natural faces. 2) the PSNR distribution of fake faces is more concentrated than that of the real images. This indicates that GAN generated faces tend to have less variations and diversities than the real faces. In addition to BM3D, we also carried out the above experiments using the RNAN network [31], and similar observations can be obtained. Due to space limitation, the detailed results are not provided here.

B. Network Architecture

Motivated by the above observations, we propose a a simple yet effective non-local attention based fake image detection network, namely NAFID. We first design a non-local feature extraction (NFE) module to extract non-local features. Then, we use a multi-stage classification sub-net to distinguish real and fake faces based on the extracted non-local features. The architecture of the proposed network is shown in Fig. 3.

1) NFE Module: The NFE module is composed of a trunk branch and a non-local mask branch. The trunk branch consists of several dense blocks [32] for feature extraction. The non-local mask branch includes a multi-head non-local block (MNLB) and several dense blocks to generate non-local attention maps. As shown in Fig. 3, the MNLB replaces single-head attention in non-local block (NLB) [31] with multi-head attention to enhance its representative capability. Each attention head in

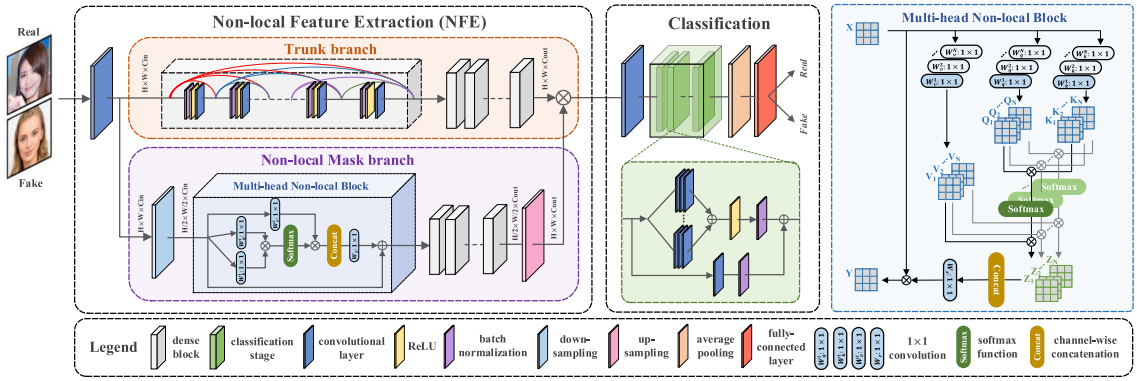


Fig. 3. The proposed NAFID network for face forgery detection. The non-local feature extraction (NFE) module extracts non-local features and the classification module distinguishes the real/fake faces based on extracted non-local features. We use only one trunk branch and one non-local mask branch in NFE module.

MNLB aims at collecting different non-local information from the input feature $\boldsymbol{X}$ with the help of non-local operation [33]. The basic form of non-local operation can be defined as follows:

$$g_q = \frac{1}{\mathcal{C}(q)} \sum_{\forall k} f(q,k) v(k), \tag{1}$$

where q is a target position in feature $\boldsymbol{X}$ and k denotes all possible positions in $\boldsymbol{X}$. $f(q,k)$ measures the similarity between q and k, and $v(k)$ is a representation of k. The $\mathcal{C}(q)$ is a normalizing factor that $\mathcal{C}(q) = \sum_{\forall k} f(q,k)$. Following Wang et al. [26], we choose Gaussian embedding function for $f(q,k)$, which makes $\frac{1}{\mathcal{C}(q)} f(q,k)$ a softmax function. In MNLB block, we have several attention heads, and each attention head aims to collect different non-local information of $\boldsymbol{X}$. Following the above descriptions, the output $\boldsymbol{Z}_i$ of the i-th attention head can be formulated as:

$$\boldsymbol{Z}_i = \mathrm{Softmax}(\boldsymbol{Q}_i \boldsymbol{K}_i^T) \boldsymbol{V}_i, \quad \text{where } i \in \{1, 2, .., N\}. \tag{2}$$

In (2), N denotes the total number of attention heads in MNLB. $\boldsymbol{Q}_i, \boldsymbol{K}_i, \boldsymbol{V}_i$ are different representations of input $\boldsymbol{X}$, i.e., $\boldsymbol{Q}_i = \boldsymbol{W}_q^i \boldsymbol{X}$, $\boldsymbol{K}_i = \boldsymbol{W}_k^i \boldsymbol{X}$, and $\boldsymbol{V}_i = \boldsymbol{W}_v^i \boldsymbol{X}$. The $\boldsymbol{W}_q^i$, $\boldsymbol{W}_k^i$ and $\boldsymbol{W}_v^i$ are weight matrices to be learned. Afterwards, the outputs of all attention heads are concatenated to go through a weight matrix $\boldsymbol{W}_z$ to yield the output $\boldsymbol{Y}$ of MNLB, as follows:

$$\boldsymbol{Y} = \boldsymbol{W}_z \mathrm{Concat}(\boldsymbol{Z}_1, \boldsymbol{Z}_2, \ldots, \boldsymbol{Z}_N) \oplus \boldsymbol{X}, \tag{3}$$

where $\mathrm{Concat}(\cdot)$ denotes channel-wise concatenation and $\oplus$ denotes element-wise addition.

As shown in Fig. 3, since we apply a 2-stride convolutional layer before the MNLB block for computational concerns, we further upsample the output of MNLB to the original size through several dense blocks and bilinear interpolation. This finishes the description of the non-local mask branch. Finally, we apply element-wise multiplication on the outputs of the trunk and non-local mask branch, to extract the non-local features from the input image. Note that the NFE module is flexible, which can be combined with any classification networks, to help improve their detection accuracy.

2) Classification Module: After extracting the non-local features through the NFE module, we simply adopt ResNeXt [34] as the classification network to distinguish the real and fake images. Note that the main contribution of this paper is the observation about the difference between real and fake images. Thus, we do not pay much attention to the design of classification network. We find that a simple ResNeXt is enough to give us advanced detection accuracy, as demonstrated in the experimental results.

III. Experimental Results and Analysis

A. Experimental Setup

1) Real/forgery Datasets: The effectiveness of our method is evaluated on six fake face datasets as mentioned in Section II-A. Each dataset contains 9 K images for training and 1 K images for testing, with equal numbers of real and fake faces. The training/testing images are split randomly in each dataset.

2) Implementation Details: We use the binary cross entropy (BCE) as the loss function to train our network. The Adam optimizer is adopted with a learning rate of 0.001. The number of attention heads is set to $N = 8$. We apply random horizontal flipping and random cropping to enlarge the training samples. The number of training epochs is 400, the batch size is 14, and the patch size is 128×128. We compare the proposed method with six state-of-the-art forgery detection methods including MesoNet [16], ResNeXt [34], Xception [29], F^3-Net [22], RFM [21] and MAT [20]. For fair comparison, the comparison methods are re-trained using the same training samples as ours.

B. Comparison Against Other Methods

Table I shows the detection accuracy of our NAFID and other forgery detection methods. As shown in this table, our NAFID achieves remarkable results on all the six datasets, especially on StyleGAN2. Many methods achieve good detection results in other datasets, but fail in StyleGAN2 which has more sophisticated GAN architecture. For example, MAT achieves 97.36% detection accuracy in StarGAN v2, but with only 73.36% in StyleGAN2. In contrast, our method achieves consistently high detection accuracy on all datasets, with 99.07% accuracy on StyleGAN2. The reason why our method performs well on different datasets is that we have observed the inherent difference between the real and fake images, which is applicable for different forgery datasets.

To further verify the effectiveness of our observation, we plug our NFE module into other comparison networks such as ResNeXt, MesoNet and Xception, to see whether their performance can be improved with NFE. For each detection network, its number of input channels is adjusted to be the same as that

TABLE I
THE DETECTION ACCURACY (%) OF OUR NAFID AND OTHER METHODS ON DIFFERENT FORGERY DATASETS

Methods	PGGAN	StyleGAN	StyleGAN2
MesoNet [16]	99.95	80.27	68.33
ResNeXt [34]	99.88	98.01	79.71
Xception [29]	99.83	99.36	59.72
F^3-Net [22]	**100.00**	99.59	85.47
RFM [21]	99.80	98.71	98.50
MAT [20]	99.71	**99.79**	73.36
NAFID (ours)	**100.00**	99.68	**99.07**
Methods	StarGAN v2	Deepfakes	AttGAN
MesoNet [16]	99.73	80.51	98.69
ResNeXt [34]	97.01	96.63	99.86
Xception [29]	99.34	95.67	99.54
F^3-Net [22]	99.89	97.48	99.95
RFM [21]	99.85	95.89	99.62
MAT [20]	97.36	97.32	99.74
NAFID (ours)	**100.00**	**98.74**	**99.98**

TABLE II
THE DETECTION ACCURACY (%) OF COMPARISON METHODS WITH OUR NFE MODULE PLUGGED IN

Methods	PGGAN	StyleGAN	StyleGAN2
MesoNet [16]	**99.95**	80.27	68.33
+Our NFE	99.22	**97.61**	**90.11**
ResNeXt [34]	99.88	98.01	82.13
+Our NFE	**100.00**	**99.68**	**99.07**
Xception [29]	99.83	99.36	59.72
+Our NFE	**99.90**	**99.69**	**98.64**
Methods	StarGAN v2	Deepfakes	AttGAN
MesoNet [16]	99.73	80.51	98.69
+Our NFE	**100.00**	**92.64**	**99.89**
ResNeXt [34]	97.01	96.63	99.86
+Our NFE	**100.00**	**98.74**	**99.98**
Xception [29]	99.34	95.67	99.54
+Our NFE	**100.00**	**97.11**	**99.81**

of output channels of our NFE module. As shown in Table II, our NFE module consistently improves the detection accuracy of the comparison methods. Take the MesoNet for example, its detection accuracy is increased from 68.33% to 90.11% on StyleGAN2, and increased from 80.51% to 92.64% on Deepfakes dataset. The similar improvement can be seen in ResNeXt and Xception models. These results verify that our NFE module is able to help improve the forgery detection accuracy of different methods.

C. Visualization Results

To visually show the effectiveness of NFE module, we apply t-SNE [35] algorithm to show the distribution of the real/fake images before and after NFE module. As shown in Fig. 4, the distribution of the original real/fake images is fully mixed with no clear clusters. In contrast, after the NFE module, the features of the real/fake images show obvious clustering effect. This phenomenon keeps for different forgery datasets, which demonstrates that the non-local features extracted by our NFE module are effective in dividing the real and fake images. Besides, since our proposed NFE module attempts to extract non-local features of fake faces, the forged faces that contain stronger non-local self-similarity can be separated from the real ones more explicitly, e.g., StarGAN.

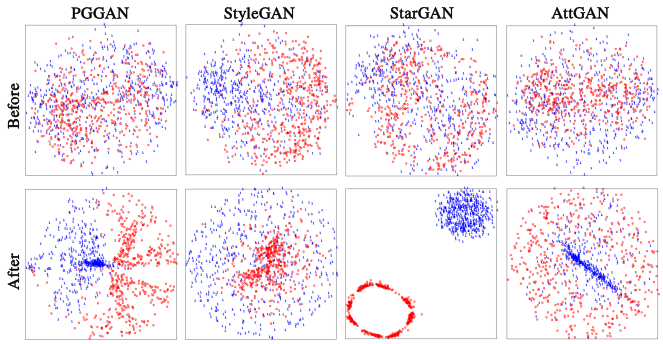


Fig. 4. The distributions of real/fake features visualized by t-SNE before and after the NFE module. The upper figures show the distributions of original real/fake images, and the lower figures show the distributions of the non-local features of real/fake images extracted by NFE module. The real images are denoted by blue dots and fake images are denoted by red ones.

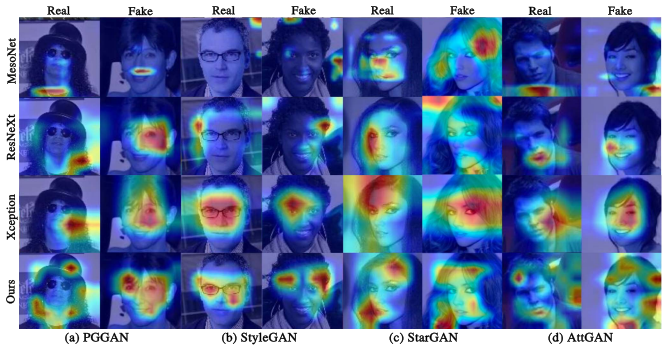


Fig. 5. The Grad-CAM heat maps of our NAFID network and other methods for both real and fake images. The last row shows the results of our method.

To further interpret how the network distinguishes the real from the fake images, we use Grad-CAM algorithm [36] to show the heat maps on the face images. Fig. 5 visualizes the heat maps generated by our NAFID and the comparison methods. As can be seen from the figure, the comparison methods rely on the local regions with forgery artifacts to detect the fake images. Compared to the local attention of the comparison methods, the heat map of our method distributes in different places across the face image. These places possess strong non-local self-similarity, such as hair and skin regions. The visualization results are consistent with the motivation of our network design, i.e., the fake images have stronger non-local properties from the real images.

IV. CONCLUSION

In this letter, we reveal an important observation for fake face detection, i.e., fake faces demonstrate stronger non-local self-similarity than real faces. This observation is demonstrated to be universal to most GAN generated fake images. Based on this observation, we propose a non-local feature extraction module to first extract the non-local features and then use a multi-stage classification module to distinguish the real and fake faces. The effectiveness of the proposed method is verified on various forgery datasets. Moreover, the NFE module is flexible to be plugged into existing forgery detection models to consistently improve their performance.

REFERENCES

[1] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of GANs for improved quality, stability, and variation," in *Proc. Int. Conf. Learn. Representations*, 2018.

[2] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 4401–4410.

[3] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, "Analyzing and improving the image quality of StyleGAN," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 8110–8119.

[4] Deepfakes, "Faceswap," Accessed on: 2017. [Online]. Available: https://github.com/deepfakes/faceswap

[5] J. Thies, M. Zollhofer, M. Stamminger, C. Theobalt, and M. Nießner, "Face2face: Real-time face capture and reenactment of RGB videos," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 2387–2395.

[6] Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, and J. Choo, "StarGAN: Unified generative adversarial networks for multi-domain image-to-image translation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 8789–8797.

[7] Y. Choi, Y. Uh, J. Yoo, and J.-W. Ha, "StarGAN v2: Diverse image synthesis for multiple domains," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 8188–8197.

[8] Z. He, W. Zuo, M. Kan, S. Shan, and X. Chen, "AttGAN: Facial attribute editing by only changing what you want," *IEEE Trans. Image Process.*, vol. 28, no. 11, pp. 5464–5478, Nov. 2019.

[9] I. Goodfellow et al., "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, vol. 27.

[10] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.

[11] Y. Li, M.-C. Chang, and S. Lyu, "In ICTU Oculi: Exposing AI created fake videos by detecting eye blinking," in *Proc. IEEE Int. Workshop Inf. Forensics Secur.*, 2018, pp. 1–7.

[12] X. Yang, Y. Li, and S. Lyu, "Exposing deep fakes using inconsistent head poses," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2019, pp. 8261–8265.

[13] U. A. Ciftci, I. Demir, and L. Yin, "FakeCatcher: Detection of synthetic portrait videos using biological signals," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, Jul. 15, 2020, doi: 10.1109/TPAMI.2020.3009287.

[14] R. Wang, Z. Yang, W. You, L. Zhou, and B. Chu, "Fake face images detection and identification of celebrities based on semantic segmentation," *IEEE Signal Process. Lett.*, vol. 29, pp. 2018–2022, 2022.

[15] B. Chu, W. You, Z. Yang, L. Zhou, and R. Wang, "Protecting world leader using facial speaking pattern against deepfakes," *IEEE Signal Process. Lett.*, vol. 29, pp. 2078–2082, 2022.

[16] D. Afchar, V. Nozick, J. Yamagishi, and I. Echizen, "Mesonet: A compact facial video forgery detection network," in *Proc. IEEE Int. Workshop Inf. Forensics Secur.*, 2018, pp. 1–7.

[17] Z. Liu, X. Qi, and P. H. Torr, "Global texture enhancement for fake face detection in the wild," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 8060–8069.

[18] L. Guarnera, O. Giudice, and S. Battiato, "Deepfake detection by analyzing convolutional traces," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, 2020, pp. 666–667.

[19] S. Chen, T. Yao, Y. Chen, S. Ding, J. Li, and R. Ji, "Local relation learning for face forgery detection," in *Proc. AAAI Conf. Artif. Intell.*, 2021, vol. 35, pp. 1081–1088.

[20] H. Zhao, W. Zhou, D. Chen, T. Wei, W. Zhang, and N. Yu, "Multi-attentional deepfake detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 2185–2194.

[21] C. Wang and W. Deng, "Representative forgery mining for fake face detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 14923–14932.

[22] Y. Qian, G. Yin, L. Sheng, Z. Chen, and J. Shao, "Thinking in frequency: Face forgery detection by mining frequency-aware clues," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 86–103.

[23] J. Li, H. Xie, J. Li, Z. Wang, and Y. Zhang, "Frequency-aware discriminative feature learning supervised by single-center loss for face forgery detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 6458–6467.

[24] H. Liu et al., "Spatial-phase shallow learning: Rethinking face forgery detection in frequency domain," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 772–781.

[25] M. Zontak and M. Irani, "Internal statistics of a single natural image," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2011, pp. 977–984.

[26] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7794–7803.

[27] C. Zuo et al., "Image denoising using quadtree-based nonlocal means with locally adaptive principal component analysis," *IEEE Signal Process. Lett.*, vol. 23, no. 4, pp. 434–438, Apr. 2016.

[28] Z. Song, B. Zhong, J. Ji, and K.-K. Ma, "A direction-decoupled non-local attention network for single image super-resolution," *IEEE Signal Process. Lett.*, vol. 29, pp. 2218–2222, 2022.

[29] A. Rossler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Nießner, "FaceForensics++ : Learning to detect manipulated facial images," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 1–11.

[30] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 3730–3738.

[31] Y. Zhang, K. Li, B. Zhong, and Y. Fu, "Residual non-local attention networks for image restoration," in *Proc. Int. Conf. Learn. Representations*, 2019.

[32] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 4700–4708.

[33] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2005, vol. 2, pp. 60–65.

[34] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 1492–1500.

[35] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, 2008.

[36] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 618–626.