

RAW Data: A Key Component for Effective Deepfake Detection

Sahar Husseinil^{*1,2} Jean-Luc Dugelay¹
¹Department of Digital Security EURECOM
²Docaposte Biometrics Lab
 Biot, France
 {husseini, dugelay}@eurecom.fr

Abstract—Current deepfake detection methods are prone to overfitting to specific deepfake artifacts and often struggle with genuine images that have undergone compression and other image processing operations. These processes can obscure indicators of forgery, leading to inaccurate decisions. This paper aims to redefine the boundary between real and fake images by narrowing the definition of authentic samples to a stage closer to the radiance of the scene as captured by the sensor, prior to any transformations by an Image Signal Processor (ISP). Our proposed method bypasses ISP processing steps, such as denoising, white balance, and de-mosaicing, which are embedded in camera hardware. This unaltered preservation makes raw data an ideal starting point for deepfake detection. Given the scarcity of large-scale datasets designed for training on raw images, we propose a methodological approach to train our model on raw image data. Our method demonstrated state-of-the-art performance on the CDF dataset and showed competitive results across other RGB domain deepfake detection datasets. The model developed in this study is available at <https://github.com/DeepFaux/Deepfake-Detection-with-RAW-Data>.

Index Terms—deepfake detection, image manipulation, ISP

I. INTRODUCTION

Figure 1 illustrates the range of deepfake generation methods; each can be understood as a controlled content creation problem, where an RGB image I_{rgb} is manipulated based on specific conditional information $C = \{\text{Image}, \text{Audio}, \text{Text}, \dots\}$. The generation process can be mathematically formulated as follows:

$$I_o = \phi_G(I_{rgb}, C) \quad (1)$$

where ϕ_G represents the specific generative network, $I_o = \{I_{rgb}^0, I_{rgb}^1, \dots, I_{rgb}^{N-1}\}$ denotes the sequence of generated contents and N is the total number of frames. This technology can be broadly categorized into four main research areas: face swapping, face reenactment, talking face generation, and facial attribute editing. Advancements in deepfake technology have enhanced applications across entertainment, art, and education. However, these developments also pose substantial risks [1]–[5]. Deepfakes pose a threat to security systems by potentially circumventing facial recognition and fooling biometric authentication software, thereby granting unauthorized access to restricted areas or sensitive data [6]. This vulnerability is particularly concerning for mobile devices used for secure unlocking, financial transactions, or access to medical records. Our method targets biometric Deepfake detection in injection

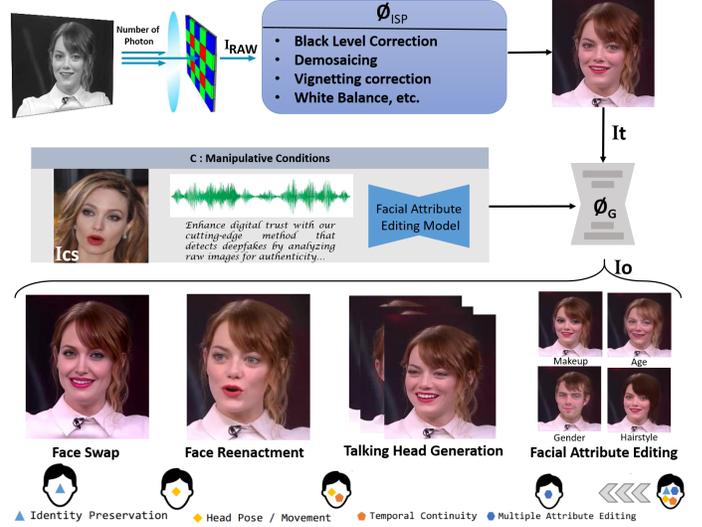


Fig. 1: Various deepfake generation methods ϕ_G manipulate the input image I_{rgb} based on conditions like audio, video frames, or text. Typically, detection models process ISP-transformed images. The ISP process starts with light focused on the CFA sensor, producing raw pixel values I_{RAW} , which undergo stages ϕ_{ISP} such as white balance and noise removal to yield the final RGB output. Our novel pipeline uses raw data (I_{RAW}) as input for improved deepfake detection.

attacks. In such cases, an attacker may use an embedded sensor and a Deepfake algorithm to generate manipulated video content in real-time. As illustrated in Figure 2, this process may involve employing a virtual camera (e.g., OBS) [7] to mimic a legitimate webcam feed, setting the Deepfake video as the source, and streaming the manipulated video to facial authentication systems by selecting the virtual camera as the input.

To mitigate the risks posed by deepfakes, different detection methods have been proposed, from early handcrafted feature-based to modern deep learning based approaches [8]–[15]. Detection tasks are framed as classification problems, applied either at the image level or the video level, depending on the practical application. These models can be represented as:

$$S_o = \phi_D(I_o), \quad (2)$$

where ϕ_D abstracts the specific detection network, and S_o represents the fake score of the generated content I_o .

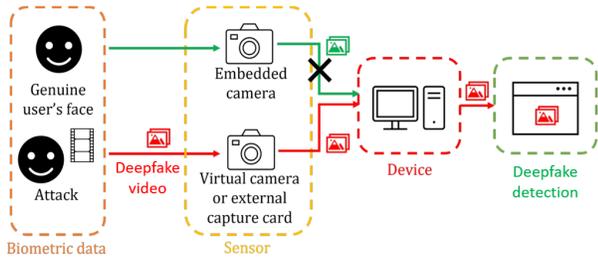


Fig. 2: Application scenario: Injection attack via virtual camera

Current deepfake detection methods often rely on supervised learning, where models are trained to recognize known deepfake artifacts. However, in real-world, the exact techniques used for facial manipulation are typically unknown, and access to the attacker’s model is generally not available. Although these models can achieve detection accuracies as high as 98%, they are susceptible to overfitting. Recent studies have recognized this challenge and aimed to enhance the robustness of detection algorithms by focusing on intrinsic indicators of forgery that go beyond relying solely on known manipulation characteristics [16]–[18]. In practical settings, content captured by cameras often undergoes various digital image and video processing operations, including post-processing such as stylization filters and beautification [13], [19], before dissemination. Recent efforts have systematically quantified the impact of such operations on detection accuracy [13], [20]–[23].

Detection models often struggle to differentiate between real and fake images because the current definition of real images encompasses both raw content captured by camera sensors and content processed through various stages of image and video processing, including linear and nonlinear adjustments. Consequently, images are considered real even after undergoing multiple processing operations such as denoising, compression, deblurring, and white balance adjustments. This paper aims to redefine the boundary between real and fake images by narrowing the definition of authentic samples to a stage closer to the radiance of the scene as captured by the sensor, prior to any transformations by an Image Signal Processor (ISP) [24], [25]. The ISP is designed to convert raw sensor data from Bayer Color Filter Arrays (CFA) into visually appealing RGB output images. This transformation process, depicted in Figure 1, begins as the camera lens focuses light onto the CFA sensor, producing a digital representation of the scene in raw pixel values. These raw images undergo several ISP stages, which primary goal is to produce images that are aesthetically pleasing to the human eye, however it introduces significant challenges in the context of deepfake detection. The challenge arises from each device using a unique ISP pipeline with distinct enhancement blocks, which obscure subtle cues critical for detecting deepfakes. This variability forces detection models to adapt to unseen ISP configurations, complicating accurate identification of real images.

Raw data, as a linear representation of scene radiance,

captures the original light distribution without alteration. This unaltered preservation makes raw data an ideal starting point for deepfake detection. In this paper, we propose a novel pipeline that utilizes raw data as input for deepfake detection. Incorporating raw data as input necessitates a modification in the existing detection formula 2, now represented as:

$$S_o = \phi_D(I_{RAW}), \quad (3)$$

where ϕ_D denotes the deepfake detection function, and I_{RAW} signifies the raw image data. Our primary objective is to focus on detection in the raw domain, assuming that attacks occur in this worst-case scenario. However, due to the scarcity of large-scale raw image datasets for training, we propose a methodological approach, detailed in Section II, to train our model for deepfake detection using raw images. This approach incorporates an auxiliary Inverse-ISP model [26] to convert processed RGB images into raw format for analysis.

The rest of this paper is organized as follows: In Section II, we present our proposed method. Section III covers the experimental setup and the results. Finally, Section IV presents the future work and conclusion of our study.

II. PROPOSED METHOD

We optimize a binary classifier using cross-entropy loss, L , to perform deepfake detection defined as:

$$L = -\frac{1}{N} \sum_{i=0}^{N-1} \{t_i \log F(x_i) + (1 - t_i) \log(1 - F(x_i))\} \quad (4)$$

where $F(x)$ denotes the probability of x being classified as “fake”, and t_i represents the binary label associated with the input image, indicating whether it is fake (1) or real (0). To enhance the generalization and robustness of our detection algorithms, we generate synthetic raw samples that embody common forgery traces. These include blending boundaries, source feature inconsistencies, and statistical anomalies in the frequency domain. All of these present significant challenges for detection. The samples undergo a conversion process from raw to RGB using ϕ_{ISP} , are manipulated within the RGB domain, and subsequently reconverted to raw using ϕ_{ISP_inv} . This novel synthetic training data, called Raw Self-Blended Images (RAW_SBI_S), is designed to improve model robustness.

Similar to the approach in [18], our synthetic data generation pipeline produces fake images by blending pseudo-source and target images, both derived from a single base image (I_{base}). This presents models with a more complex and generalized task for face forgery detection. To produce RAW_SBI_S, we develop a Source-Target Generator (STG) and a Mask Generator (MG). The STG starts by generating pairs of pseudo source and target images from single pristine images using straightforward image processing techniques, and the MG creates various blending masks based on the facial landmarks of the input images.

As illustrated in Figure 3, the generation of raw synthesized training data begins with converting a raw image (I_{RAW}) to RGB format (I_{base}) using ϕ_{ISP} . The bounding box and

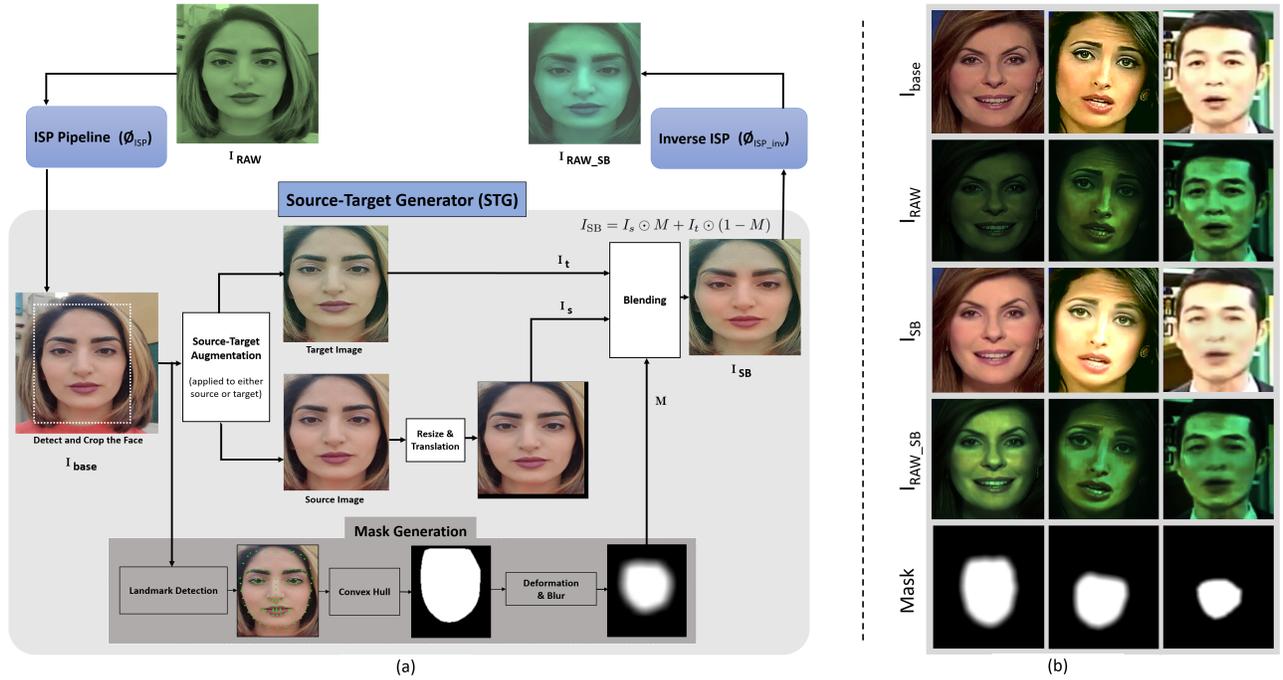


Fig. 3: (a) Overview of generating a RAW Self-Blended Image (RAW_SBI). A base image I_{base} is fed into the Source-Target Generator (STG) and the Mask Generator (MG). The STG produces pseudo source and target images from the base image using various image augmentations, while the MG creates a blending mask from facial landmarks and deforms it to enhance mask diversity. The source and target images are then blended with the mask and input into the Inverse ISP pipeline to reconstruct the raw format of the RGB input image. (b) Example images of I_{base} , I_{RAW} , I_{SB} , and I_{RAW_SB} . The I_{base} samples are sourced from FF++ dataset. I_{RAW} and I_{RAW_SB} are transformed from I_{base} and I_{SB} respectively, using inverse ISP model.

facial landmarks of the face in the input image are then detected, leading to the cropping of the face area. This cropped image serves as both the source and target image for further processing. To introduce inconsistencies between the source and target images, a series of augmentations, such as brightness and contrast adjustments, are applied to either source or target image. The source image is also resized and translated to replicate blending boundaries and landmark mismatches. Additionally, the Mask Generator (MG) produces a grayscale mask specifying the manipulated region.

The initial mask is defined as the convex hull of the facial landmarks in the input image. Given that face manipulation techniques may target various areas of the face, resulting in diverse shapes of manipulated regions (e.g., affecting only the mouth, eye, or entire face region), the mask is altered to accommodate these variations using landmark augmentation [16]. The mask is then smoothed using a Gaussian filter. Finally, the mask is applied to blend the source and target images according to Equation 5, resulting in the creation of the self-blended image (I_{SB}).

$$I_{SB} = I_s \odot M + I_t \odot (1 - M) \quad (5)$$

Where, I_s , I_t , and M represent the source image, target image, and the generated mask, respectively. After preparing the self-blended image, we use an inverse ISP model, ϕ_{ISP_inv} to transform the RGB image (I_{SB}) back into the raw image domain (I_{RAW_SB}). Finally, this synthesized raw data, along

with the original raw image (I_{RAW}), is input to the detection algorithm for training.

III. EXPERIMENT

Our proposed methodology employs raw domain data as the primary input for deepfake detection. Recognizing the scarcity of large-scale datasets specifically designed for training on raw domain images, our method reverses the data processing flow during the training phase. Instead of converting raw data into RGB images as depicted in Figure 3a, we initiate with RGB images from a publicly available dataset, labeled as I_{base} . These images are then transformed into raw format, labeled as I_{RAW} . The generated I_{RAW} and synthetically created fake images in raw format (I_{RAW_SB}) are utilized for training. This reversal strategy offers several key advantages: 1) It enables us to leverage the existing wealth of large-scale RGB data while still benefiting from training in the raw domain. 2) It allows direct comparison of our method with existing state-of-the-art RGB-based models. 3) It facilitates the application of our model in scenarios where only RGB images are available.

A. Experimental Setup

Dataset: We trained our model on the widely used benchmark dataset, FaceForensics++ (FF++) [8], using only the real videos, consisting of 720 videos for training and 140 videos for validation. To evaluate the performance of our approach, we utilized the test set from the FF++ dataset, which includes both authentic and manipulated videos. For cross-dataset evaluation,

we employed three recent deepfake datasets, namely Celeb-DF-v2 (CDF) [27], Deep-Fake Detection (DFD) [28] and DeepFake Detection Challenge public test set (DFDC) [29].

RAW Data Generation: In the absence of authentic raw image data, and to facilitate the generation of synthetic raw domain images via our proposed pipeline, we employ the state-of-the-art inverse ISP model, known as MiAlgo [26]. This method is trained to recover raw data from the RGB Huawei P20 model; however, it can effectively generalize to noisy and unseen similar sensors. Moreover, the method doesn’t require any metadata or specific camera parameters (e.g., correction matrices, digital gains), which are typically inaccessible. MiAlgo utilizes an end-to-end encoder-decoder UNet-like structure, incorporating key components such as the residual group [30] and the enhanced block [31]. The algorithm takes a full-resolution RGB image and converts the input image to a raw RGGB pattern. Consequently, we perform demosaicing on the RGGB image and then feed this raw data into a binary classifier for deepfake detection.

Evaluation Metrics: We utilize the video-level area under the receiver operating characteristic curve (AUC) for comparison with previous research. Normally, predictions at the frame level are averaged across video frames.

Training details: We utilize EfficientNet-b4 (EFNB4) [32], pretrained on ImageNet, as our classifier and train it for 100 epochs using the SAM optimizer [33]. We set the batch size to 32 and the learning rate to 0.001. During training, we sample only eight frames per video. Each batch includes both real images and their self-blended raw counterparts.

B. Experimental Results and Analysis

Our methodology is benchmarked using the public dataset described in Section III-A, against other RGB-domain deepfake detection techniques [14]–[18], [34]–[38].

1) *Quantitative Results:* Table I presents the Area Under the Curve (AUC) evaluation metrics of our method compared with existing methods. Our method, EFNB4 + $I_{RAW_{SB}}$, demonstrated state-of-the-art performance on the CDF dataset and showed competitive results across other deepfake detection datasets. More specifically, on the CDF dataset, our approach achieved an impressive AUC of 94.23%, surpassing EFNB4 + SBIs, which obtained 93.18%. For the DFD dataset, PCL + I2G attained the highest AUC of 99.07%, while our method followed closely with an AUC of 98.46%. In the DFDC dataset, EFNB4 + SBIs led with an AUC of 72.42%, and FTCN, which utilized both fake and real sets during training, achieved 71.00%. Our method secured an AUC of 69.42%, ranking as the third-best performer on this dataset. Table II presents our cross-manipulation evaluation results on FF++. We can see that even if our method is not trained on specific artifacts within the FF++ dataset, still our method performs well in recognizing different deepfake artifacts generated by different methods or nearly matches existing methods across four manipulations (99.92% on DF, 99.16% on F2F, 99.46% on FS, and 99.72% on NT) and achieves 99.56% the overall performance on the entire FF++ test set.

TABLE I: Cross dataset evaluation on CDF, DFD, and DFDC Datasets. The model is trained exclusively on the high-quality FF++ dataset using only real data. Results from previous methods are cited from their original papers. Bold values indicate the best performance, while underlined values denote the second-best performance.

Method	Input Type	Training set		Test Set AUC (%)		
		Real	Fake	CDF	DFD	DFDC
DSP-FWA [17]	Frame	✓	✓	69.30	-	-
Face X-ray + BI [16]	Frame	✓		-	93.47	-
Face X-ray + BI [16]	Frame	✓	✓	-	95.40	-
LRL [34]	Frame	✓	✓	78.26	89.24	-
FRDM [14]	Frame	✓	✓	79.4	91.9	-
PCL + I2G [15]	Frame	✓		90.03	99.07	67.52
EFNB4 + SBIs [18]	Frame	✓		<u>93.18</u>	<u>97.56</u>	72.42
Two-branch [35]	Video	✓	✓	76.65	-	-
DAM [36]	Video	✓	✓	75.3	-	-
LipForensics [37]	Video	✓	✓	82.4	-	-
FTCN [38]	Video	✓	✓	86.9	94.40	<u>71.00</u>
EFNB4 + $I_{RAW_{SB}}$ (ours)	Frame	✓		94.23	<u>98.46</u>	69.42

TABLE II: Cross-manipulation evaluation on FF++.

Method	Test Set AUC (%)				
	DF	F2F	FS	NT	FF++
Face X-ray + BI [16]	99.17	98.57	98.21	98.13	98.52
PCL + I2G [15]	100	98.97	99.86	97.63	99.11
EFNB4 + SBIs [18]	99.99	99.88	99.91	99.79	99.64
EFNB4 + $I_{RAW_{SB}}$ (ours)	99.92	99.16	99.46	99.72	99.56

2) *Qualitative Analysis of Inverse ISP:* Figure 3b presents examples of I_{base} , I_{RAW} , I_{SB} , $I_{RAW_{SB}}$ and blending mask. The I_{base} samples are sourced from FF++ dataset [8]. I_{RAW} and $I_{RAW_{SB}}$ are transformed from I_{base} and I_{SB} respectively, using inverse ISP model (see Figure 3a). Although the inverse ISP model was not specifically trained for face images, it successfully reconstructs the raw domain data with high fidelity. Furthermore, we observe that the artifacts in synthetic images (I_{SB}), generated from I_{base} , are distinctly visible and can also be recognized in $I_{RAW_{SB}}$, indicating effective transformation and consistency across image representations.

IV. CONCLUSION AND FUTURE WORK

This study presents a novel pipeline that utilizes raw data to enhance deepfake detection, addressing limitations of current models that struggle with genuine images modified by ISP pipelines. Our method has demonstrated state-of-the-art performance on the CDF dataset and competitive results on other RGB domain deepfake detection datasets. To the best of our knowledge, this study is the first to propose using RAW data for Deepfake detection. Given the limited research on RAW data attacks and the absence of diverse RAW Deepfake generators, we employed an inverse ISP model to simulate RAW data during training. This approach enabled us to make meaningful comparisons with existing RGB-based methods. Future work should focus on developing large-scale datasets from sensor outputs to improve training data effectiveness and authenticity.

REFERENCES

- [1] Felix Juefei-Xu, Run Wang, Yihao Huang, Qing Guo, Lei Ma, and Yang Liu, "Countering malicious deepfakes: Survey, battleground, and horizon," *International journal of computer vision*, vol. 130, no. 7, pp. 1678–1734, 2022. 1
- [2] Sifat Muhammad Abdullah, Aravind Cheruvu, Shravya Kanchi, Taejoong Chung, Peng Gao, Murtuza Jadhwal, and Bimal Viswanath, "An analysis of recent advances in deepfake image detection in an evolving threat landscape," in *2024 IEEE Symposium on Security and Privacy (SP)*. IEEE Computer Society, 2024, pp. 206–206. 1
- [3] Sahar Husseini and Jean-Luc Dugelay, "Metahumans help to evaluate deepfake generators," in *2023 IEEE 25th International Workshop on Multimedia Signal Processing (MMSP)*. IEEE, 2023, pp. 1–6. 1
- [4] Sahar Husseini, Jean-Luc Dugelay, Fabien Aili, and Emmanuel Nars, "A 3d-assisted framework to evaluate the quality of head motion replication by reenactment deepfake generators," in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2023, pp. 1–5. 1
- [5] Pranav Balaji, Abhijit Das, Srijan Das, and Antitza Dantcheva, "Attending generalizability in course of deep fake detection by exploring multi-task learning," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 475–484. 1
- [6] Sahar Husseini and Jean-Luc Dugelay, "Alignface: Enhancing face verification models through adaptive alignment of pose, expression, and illumination," . 1
- [7] OBS Studio, "<https://obsproject.com/kb/virtual-camera-guide>," 2024, Accessed: 2024-09-10. 1
- [8] Andreas Rossler, Davide Cozzolino, Luisa Verdoliva, Christian Riess, Justus Thies, and Matthias Nießner, "Faceforensics++: Learning to detect manipulated facial images," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 1–11. 1, 3, 4
- [9] Dennis Siegel, Christian Kraetzer, Stefan Seidlitz, and Jana Dittmann, "Media forensics considerations on deepfake detection with hand-crafted features," *Journal of Imaging*, vol. 7, no. 7, pp. 108, 2021. 1
- [10] Arash Heidari, Nima Jafari Navimipour, Hasan Dag, and Mehmet Unal, "Deepfake detection using deep learning methods: A systematic and comprehensive review," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 14, no. 2, pp. e1520, 2024. 1
- [11] Jongwook Choi, Taehoon Kim, Yonghyun Jeong, Seungryul Baek, and Jongwon Choi, "Exploiting style latent flows for generalizing deepfake detection video detection," *arXiv preprint arXiv:2403.06592*, 2024. 1
- [12] Shichao Dong, Jin Wang, Renhe Ji, Jiajun Liang, Haoqiang Fan, and Zheng Ge, "Implicit identity leakage: The stumbling block to improving deepfake detection generalization," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 3994–4004. 1
- [13] Alexandre Libourel, Sahar Husseini, Nelida Mirabet-Herranz, and Jean-Luc Dugelay, "A case study on how beautification filters can fool deepfake detectors," in *IWBF 2024, 12th IEEE International Workshop on Biometrics and Forensics*, 2024. 1, 2
- [14] Yuchen Luo, Yong Zhang, Junchi Yan, and Wei Liu, "Generalizing face forgery detection with high-frequency features," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 16317–16326. 1, 4
- [15] Tianchen Zhao, Xiang Xu, Mingze Xu, Hui Ding, Yuanjun Xiong, and Wei Xia, "Learning self-consistency for deepfake detection," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 15023–15033. 1, 4
- [16] Lingzhi Li, Jianmin Bao, Ting Zhang, Hao Yang, Dong Chen, Fang Wen, and Baining Guo, "Face x-ray for more general face forgery detection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 5001–5010. 2, 3, 4
- [17] Yuezun Li and Siwei Lyu, "Exposing deepfake videos by detecting face warping artifacts," . 2, 4
- [18] Kaede Shiohara and Toshihiko Yamasaki, "Detecting deepfakes with self-blended images," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 18720–18729. 2, 4
- [19] Nélica Mirabet-Herranz, Chiara Galdi, and Jean-Luc Dugelay, "Impact of digital face beautification in biometrics," in *2022 10th European workshop on visual information processing (EUVIP)*. IEEE, 2022, pp. 1–6. 2
- [20] Yuhang Lu and Touradj Ebrahimi, "Assessment framework for deepfake detection in real-world situations," *EURASIP Journal on Image and Video Processing*, vol. 2024, no. 1, pp. 6, 2024. 2
- [21] Federico Cocchi, Lorenzo Baraldi, Samuele Poppi, Marcella Cornia, Lorenzo Baraldi, and Rita Cucchiara, "Unveiling the impact of image transformations on deepfake detection: An experimental analysis," in *International Conference on Image Analysis and Processing*. Springer, 2023, pp. 345–356. 2
- [22] Xu Zhang, Svebor Karaman, and Shih-Fu Chang, "Detecting and simulating artifacts in gan fake images," in *2019 IEEE international workshop on information forensics and security (WIFS)*. IEEE, 2019, pp. 1–6. 2
- [23] Steven Gal and Burcu Bulgurcu, "Exploring factors influencing internet users' susceptibility to deepfake phishing," 2024. 2
- [24] Tim Brooks, Ben Mildenhall, Tianfan Xue, Jiawen Chen, Dillon Sharlet, and Jonathan T Barron, "Unprocessing images for learned raw denoising," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 11036–11045. 2
- [25] Sahar Husseini, Pouria Babahajiani, and Moncef Gabbouj, "Color constancy model optimization with small dataset via pruning of cnn filters," in *2021 9th European Workshop on Visual Information Processing (EUVIP)*. IEEE, 2021, pp. 1–6. 2
- [26] Marcos V Conde, Radu Timofte, Yibin Huang, Jingyang Peng, Chang Chen, Cheng Li, Eduardo Pérez-Pellitero, Fenglong Song, Furui Bai, Shuai Liu, et al., "Reversed image signal processing and raw reconstruction. aim 2022 challenge report," in *European Conference on Computer Vision*. Springer, 2022, pp. 3–26. 2, 4
- [27] Yuezun Li, Xin Yang, Pu Sun, Honggang Qi, and Siwei Lyu, "Celeb-df: A large-scale challenging dataset for deepfake forensics," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 3207–3216. 4
- [28] "Contributing data to deepfake detection research.," [Online; accessed 2024-04-30]. 4
- [29] Brian Dolhansky, Joanna Bitton, Ben Pflaum, Jikuo Lu, Russ Howes, Menglin Wang, and Cristian Canton Ferrer, "The deepfake detection challenge (dfd) dataset," *arXiv preprint arXiv:2006.07397*, 2020. 4
- [30] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao, "Cycleisp: Real image restoration via improved data synthesis," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 2696–2705. 4
- [31] Marcos V Conde, Steven McDonagh, Matteo Maggioni, Ales Leonardis, and Eduardo Pérez-Pellitero, "Model-based image signal processors via learnable dictionaries," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2022, vol. 36, pp. 481–489. 4
- [32] Mingxing Tan and Quoc Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning*. PMLR, 2019, pp. 6105–6114. 4
- [33] Pierre Foret, Ariel Kleiner, Hossein Mobahi, and Behnam Neyshabur, "Sharpness-aware minimization for efficiently improving generalization," in *International Conference on Learning Representations*, 2020. 4
- [34] Shen Chen, Taiping Yao, Yang Chen, Shouhong Ding, Jilin Li, and Rongrong Ji, "Local relation learning for face forgery detection," in *Proceedings of the AAAI conference on artificial intelligence*, 2021, vol. 35, pp. 1081–1088. 4
- [35] Iacopo Masi, Aditya Killekar, Royston Marian Mascarenhas, Shenoy Pratik Gurudatt, and Wael AbdAlmageed, "Two-branch recurrent network for isolating deepfakes in videos," in *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VII 16*. Springer, 2020, pp. 667–684. 4
- [36] Tianfei Zhou, Wenguan Wang, Zhiyuan Liang, and Jianbing Shen, "Face forensics in the wild," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 5778–5788. 4
- [37] Alexandros Haliassos, Konstantinos Vougioukas, Stavros Petridis, and Maja Pantic, "Lips don't lie: A generalisable and robust approach to face forgery detection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 5039–5049. 4
- [38] Yinglin Zheng, Jianmin Bao, Dong Chen, Ming Zeng, and Fang Wen, "Exploring temporal coherence for more general video face forgery detection," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 15044–15054. 4