Pixel Perfect or Perfectly Fake? Exploring the Robustness of Digital Forensic Tools Against Facial Deepfakes and Morphed Images

Silvia Lucia Sanna[1], Andrea Panzino[1], Simone Maurizio La Cava[1], Sara Concas[1],
Leonardo Regano[1], Davide Maiorca[1], Gian Luca Marcialis[1], Giorgio Giacinto[1,2]
[1]*Department of Electrical and Electronic Engineering*, *University of Cagliari* Cagliari, Italy
[2]*CINI, Consorzio Interuniversitario Nazionale per l'Informatica*, Italy
email: {silvial.sanna, andrea.panzino, simonem.lac, sara.concas90c,
leonardo.regano, davide.maiorca, marcialis, giorgio.giacinto}@unica.it

***Abstract*—Digital forensic investigations increasingly rely on tools that streamline digital data retrieval and analysis. Nowadays, some tools even leverage Artificial Intelligence (AI) to automatically categorize content like people's identities. However, this dependence on AI raises concerns about the robustness of these algorithms against malicious digital manipulations. This is particularly evident in the case of facial manipulation techniques like deepfakes and morphing attacks, which are capable of altering or blending identities represented in videos and images. To investigate the potential vulnerabilities of AI-enhanced digital forensic tools to these deceptive practices, we conducted a preliminary analysis of two widely used forensic tools that integrate AI for data classification: `Magnet.AI` and `Excire Photo AI`. We assessed their performance on deepfake and morphed images based on state-of-the-art image and video datasets related to celebrities, revealing that the considered forensic tools lack sufficient robustness against facial manipulations. Building on these findings, we provide recommendations for enhancing the integration of AI in digital forensic analysis regarding facial manipulations, with the final goal of enforcing the integrity and reliability of digital data.**

***Index Terms*—Digital Forensics, Artificial Intelligence, Image Forgery, Face Recognition, Deepfake, Morphing.**

I. Introduction

Digital Forensics is the science of retrieving evidence from any digital device (i.e., digital evidence) by employing specific techniques to extract, analyze and interpret data according to the target devices and operating systems. Forensics techniques are especially useful to analyze devices in the context of criminal investigations or to reconstruct cyberattacks (e.g., in industrial environments). The various phases of forensics (data acquisition, examination, analysis, etc.) must be thoroughly conducted to ensure digital data availability, reproducibility, and integrity [1].

Many tools have been developed to aid analysts in the various phases of the forensics process. In particular, some of them have integrated Artificial Intelligence (AI) techniques to ease and speed up the retrieval and examination of digital data [1]. AI is especially useful aid the analyst in selecting images and videos that feature specific individuals or belong to specific classes (e.g., drug, violence, nudity, weapons, etc.) [2], [3]. This is particularly useful to, e.g., automatically detect and classify child pornography pictures [1], preserving the analyst from emotional stress and speeding up the analysis [4].

However, images and videos can be easily manipulated with approaches aimed at deceiving AI-based algorithms. This set of techniques, known as *presentation attacks*, can subtly alter the visual content in ways that are often imperceptible to the human eye, leading the algorithm to incorrectly classify the manipulated media. Such manipulation strategies have further evolved with the development of *deepfakes*, which rely on advanced deep-learning algorithms to tamper visual and audio contents that convincingly mimic or alter the appearance, behaviour, or even identity of represented individuals [5]. Another significant threat is *morphing*, an image manipulation technique that blends facial features from multiple individuals to create synthetic images that can be simultaneously attributed to all the contributing identities [6]. These facial manipulation techniques are being specifically developed to deceive both humans and automatic identity recognition systems, even those enhanced through AI [5], [6].

Considering the stringent integrity requirements of digital evidence in forensic investigations [2], the vulnerability of AI-enhanced systems for facial data analysis raises a critical question: *Are AI-enhanced digital forensic tools robust against facial manipulation techniques?* To our knowledge, this issue has not yet been thoroughly examined. Therefore, we present a first preliminary analysis of two popular digital forensic tools, conducting an initial evaluation on the robustness of two widely used digital forensic tools against deepfake and morphing attacks. In this context, we refer to robustness as the ability to maintain accurate and reliable performance despite the presence of deceptive facial alterations in analyzed data [7], [8]. Our study leverages state-of-the-art image and video datasets featuring celebrities, utilizing deepfakes and morphed images generated through advanced techniques, to investigate the susceptibility of AI-enhanced functionalities regarding automated facial data analysis.

[1]https://www.papelesdelpsicologo.es/English/2778.pdf

The remainder of this paper is structured as follows. Section II provides an overview on deepfakes and morphing attacks. Section III provides the experimental protocol considered to analyze the effectiveness of digital forensic tools against these facial manipulations. Finally, results are reported in Section IV, and conclusions are drawn in Section V.

II. Deepfakes and Morphing Attacks

In recent years, the rapid evolution of AI techniques and deep learning has led to a shocking improvement in the development of the so-called *deepfake media*. By the term deepfake, we generally refer to images or videos within which the identity of one or more people has been altered in some way. Despite the many possible applications in the fashion or movie industries, deepfakes are often employed for illegal purposes. Popular malicious applications include the spread of misinformation and fake news, the fabrication of content that tries to influence public opinion and political outcomes, and the impersonation of individuals in a convincing manner for identity theft, phishing attacks, financial fraud, privacy violation, and blackmail [9], [10].

Deepfakes can be obtained through different types of manipulations, which span from altering a single attribute (such as an individual's gender, age, or facial movements) to synthesizing entire artificial faces [11]. One of the most dangerous manipulation techniques is *face swapping*, in which a source face replaces the identity information of a target face. This kind of forged media can easily deceive humans and several face recognition systems [12]. From traditional graphics approaches [13], this technique significantly evolved to the point of using Generative Adversarial Networks (GANs) [14] and Diffusion-based approaches [15], creating increasingly convincing and more dangerous forgeries.

Recent advances in digital data manipulation technology have also led to *morphing*, a technique that involves blending features from multiple images or videos to create seamless transitions between facial expressions, identities, or actions. For instance, morphing can generate images resembling *multiple identities*. Although initially used for artistic and entertainment purposes, morphing techniques can also be exploited for malicious activities, such as identity fraud [6]. Despite the various techniques for generating facial morphs, most of them can be summarized into two main approaches: *landmark-based* and *deep learning-based* methods [16]. The methods employing the first approach combine facial landmark detection with geometric transformations to merge source faces. The methods based on deep learning leverage models for extracting facial information and synthesizing morphs, primarily using GANs. Despite the generally higher graphics quality of the images compared to landmark-based methods and the absence of their typical artifacts, methods based on deep learning tend to be less effective in preserving identity features due to the introduced distortions. Hence, face recognition systems are generally less vulnerable to morphs generated through methods based on deep learning than those based on landmarks [17], [18].

To summarize, deepfakes and morphing pose significant challenges to the trustworthiness of digital content. In particular, these facial manipulation techniques can be employed to deceive both humans and automatic identity recognition systems, thus potentially misleading the outcomes of AI-enhanced digital forensic tools as well.

III. Experimental Framework

The experimental framework is outlined in three parts. Section III-A introduces the digital forensic tools selected for the study. Section III-B reports the datasets used for the following analysis. Finally, Section III-C describes the experiments and the considered performance metrics.

A. Tools Selection

Various forensics tools are currently available to analyze evidence. Such tools are selected according to the type of target digital device (e.g., computers, smartphones, etc.), the type of digital evidence (e.g., RAM dumps, disk partitions), and the supported Operating System (OS) (e.g., Linux, Windows, Android, iOS). For example, RAM data is usually analyzed with `Volatility` [2] while disk content is usually processed with `Autopsy` [3], `FTK Imager` [4], `Magnet Axiom` [5], `X-Ways Forensics` [6]. NAND data of mobile devices is typically processed with Cellebrite `Inseyets (UFED)` [7] and `Oxygen Forensics`[8]. Notably, the choice of the tool also influences the acquisition and analysis methodology.

To the best of our knowledge, among the widely diffused tools, `Excire Photo AI` (used by `X-Ways Forensics`) and `Magnet Axiom` represent the most popular integrations with AI technologies. `Excire Photo AI` [9] detects and recognize the faces of popular people (e.g., actors). `Magnet Axiom` performs the examination by using a module called *examine*, which features black-box AI algorithms to classify the found data (`Magnet.AI` [10]). It can automatically classify the content present in pictures and chats, thus detecting the presence of a human face, possibly AI-generated content, drugs, nudity, weapons, etc. Both the aforementioned tools are *black-box* and do not provide any explanation about the classification results. However, differently to `Excire`, `Axiom` does not recognize people by name but can detect faces in pictures and distinguish between real and artificially generated images. Hence, in a professional context the two tools are complementary to each other.

Using `Excire` to classify images according to the name of a popular person is straightforward: first, we added the directory containing the target images from our dataset to the tool by using the `Add Folders` command. Then, the

[2]https://volatilityfoundation.org/
[3]https://www.autopsy.com
[4]https://www.exterro.com/digital-forensics-software/ftk-imager
[5]https://www.magnetforensics.com/products/magnet-axiom/
[6]https://www.x-ways.net/forensics/
[7]https://cellebrite.com/en/ufed/
[8]https://www.oxygenforensics.com/en/
[9]https://excire.com/en/excire-foto/
[10]https://www.magnetai.ca/

images are then automatically analyzed by the AI algorithm. To extract the results, we used the `Find by text prompt` functionality under `Find Section` and typed the name of the queried person[11]. The tool displays all images related to the queried person, and the results are then saved to a .csv file that includes the full filepath of the detected images according to the specific query.

Concerning `Magnet.AI`, we constructed forensic evidence by dumping a USB drive containing the target images, as the tool does not allow the loading of images for classification directly. We chose a completely clean USB drive (formatted and with zeros written on it) without any installed operating system to avoid the presence of external files (e.g., default icons or images embedded in installed applications). After the acquisition process, we used `Magnet Axiom Examine` to analyze the evidence, enabling the use of `Magnet.AI` categorization on the pictures. We selected the tags related to the goals of this paper (`Possible human faces` and `Possible AI-generated content`) and saved the results in a .xlsx file with the full path of the detected images.

B. Datasets

We evaluated the robustness of the presented tools on two well-known state-of-the-art image and video datasets, typically used for deepfake forensics and face recognition analyses. In particular, the choice of the datasets described below is driven by the functionalities and limitations of the tools selected for the investigation, previously described in Section III-A.

The first dataset is Celeb-DF [19], a large-scale benchmark dataset for deepfake forensics, released in 2020. It includes 590 original videos collected from YouTube with 59 celebrities of different ages, ethnic groups, and genders, as well as 5639 corresponding deepfake videos generated through face swapping. Specifically, we employed this dataset for the analysis related to deepfakes, extracting the first frame from each video since `Excire Photo AI` performs identity recognition at the image level.

The dataset we employed for the analysis related to morphing attacks is CelebAMask-HQ [20], a large-scale dataset, released in 2020 too. It contains $30,000$ high-resolution face images of celebrities selected from the CelebA dataset [21] of 2015. Considering the purposes of this facial manipulation technique, we selected the only publicly available images presenting complete faces in neutral pose and expression without any significant occlusion. The images resulting from such a selection are 92, related to the same number of identities. From these images, we generated the morphs using a landmark-based approach, following the method outlined in [22]. This approach was chosen to reduce the presence of ghosting artifacts in the morphed picture by applying the morph only on the inner facial region of the source image. Specifically, we employed a modified version of Face Morpher[12] to comply with the following morphing procedure. Thus, given a source I_1 and a target image I_2, we obtained two morphs: M_{12} (by morphing I_1 to I_2) and M_{21} (by morphing I_2 to I_1). Therefore, we generated 5108 morphed images that, in addition to the initial 92 images, represent the dataset employed for the experiments on the robustness of digital forensic tools against morphing attacks.

It is important to remark that in the case of the considered deepfakes (i.e., face swapping) the "source" refers to the person whose face is being transferred onto another body while the "target" is the person in the original video or image whose face is replaced. On the contrary, morphing considers the "source" as the image employed as the base, while the "target" is the face that is blended with the source to create the morphed identity. Moreover, in both cases, the images were used in their original dimension without any preprocessing (e.g., cropping).

C. Performance Evaluation

As previously introduced, we conducted various experiments based on the provided AI-enhanced functionalities to assess the robustness of the selected digital forensic tools against facial manipulations. Specifically, we tested the reliability of `Magnet.AI` in extracting images containing faces and its capability to detect AI-generated images, while we assessed the reliability of `Excire Photo AI` in recognizing identities in images or involved in the forgery process.

We conducted experiments on each dataset to better reveal the vulnerability of the two digital forensic tools to different manipulation techniques. We consider several metrics commonly used to evaluate the performance in biometrics and, more in general, classification systems. Specifically, for face detection, we only assess the false negative rate (FNR) since the datasets contain facial images. For the other tasks, we also assess false positive rate (FPR) and balanced accuracy. In these contexts, a False Positive occurs when the system incorrectly detects or recognizes a face or a manipulation when it should not, while a False Negative occurs when the system fails to detect or recognize a face or a manipulation that should have been identified.

In accordance with the recent ISO/IEC 30107–3 standard [23], when considering the detection of the manipulated images, the false negatives correspond to the APCER (Attack Presentation Classification Error Rate), while the false positives correspond to the BPCER (Bona-fide Presentation Classification Error Rate). Similarly, in a face recognition scenario, the FPR and FNR correspond respectively to the False Match Rate (FMR) and False Non-Match Rate (FNMR).

We also analyzed the identity recognition capability individually for the cases concerning zero-effort attacks (i.e., unmanipulated images representing an identity different from the declared one) and presentation attacks through manipulation techniques (i.e., deepfake and morphed images). Concerning the latter case, we individually consider the cases in which the declared identity is associated with the source image, the target image, and none of the images involved in the generation.

[11] https://www.x-ways.net/Excire_Detected_Objects.txt

[12] https://github.com/alyssaq/face_morpher

TABLE I

RESULTS OF THE FACE DETECTION ANALYSIS THROUGH MAGNET AI. THE FALSE NEGATIVE RATE (FNR) REFERS TO THE IMAGES CONTAINING UNDETECTED FACES.

FNR [%]			
Celeb-DF		CelebAMask-HQ	
Real	Deepfakes	Real	Morphs
0.34	0.29	0.00	0.00

TABLE II

RESULTS OF THE DETECTION OF FACIAL MANIPULATIONS THROUGH MAGNET AI. DEEPFAKES AND MORPHED IMAGES ARE CONSIDERED THE POSITIVE CLASS IN CELEB-DF AND CELEBAMASK-HQ, RESPECTIVELY.

	FNR [%]	FPR [%]	Balanced Accuracy [%]
Celeb-DF	99.85	0.51	49.81
CelebAMask-HQ	99.90	1.09	49.51

IV. RESULTS AND DISCUSSION

In this section, we provide the results obtained through the previously described experimental protocol. The discussion follows the order of steps that would be performed by a forensic expert concerning the analysis of potentially manipulated facial data: extraction of images containing faces, detection of potential manipulations, and, finally, facial recognition.

Table I reports the results obtained from the detection of images containing faces through `Magnet.AI`. The outcomes show that the impact of morphs and deepfakes on performance is limited. This means that the tool can still be used to extract images containing faces even in the presence of altered images.

Table II shows instead that the same tool is unreliable in detecting AI-generated images. Specifically, the probability that a morph or a deepfake is recognized as an image generated through AI is even lower than that of real unmanipulated images, namely up to 0.15% (i.e. FNR = 99.85% for deepfakes) and 1.09%, respectively.

Table III revealed that identity recognition through `Excire Photo AI` in the presence of facial manipulations is unreliable as well due to a significant reduction in the overall accuracy. In particular, such a decay in performance is caused by an increment in error probability on impostors, namely deepfake and morphing images generated by considering the searched subject in the forgery process.

Through a more detailed analysis, it is possible to observe that the increment of errors is related to images employing the subject as the source, as shown in Figure 1. This is expected since, as previously introduced in Section III-B, the identity portrayed in the source image contributes the most to the outcome both in morphs and deepfakes. This represents the traditional application scenario where the attacker must generate digital media presenting the specific subject. Moreover, considering the reported results, we highlight that, due a too low error rate, the tool is unable to correctly recognize all the manipulated images related to the individual. Conversely, the error rate is too high for the tool to correctly filter such images. Thus, we can conclude that the tool is unreliable both

TABLE III

FACE RECOGNITION RESULTS. GENUINE ORIGINAL (UNMANIPULATED) IMAGES PRESENTING THE SEARCHED IDENTITY ARE CONSIDERED TO BELONG TO THE POSITIVE CLASS.

	Impostors compared with genuine data	FNR [%]	FPR [%]	Balanced Accuracy [%]
Celeb-DF	**Zero-effort impostors**	24.73	2.29	86.49
	Deepfakes		39.09	68.09
CelebAMask-HQ	**Zero-effort impostors**	9.78	8.83	90.69
	Morphed images		40.21	75.01

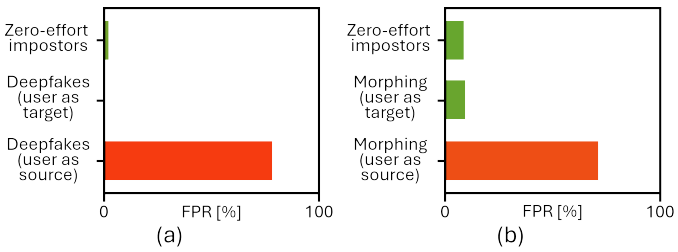

Fig. 1. FPR obtained in face recognition through `Excire Photo AI` when comparing genuine original images (positive class) with zero-effort impostors, deepfake images (a), and morphed images (b).

in cases in which forensic practitioners are required to discard manipulated images (e.g., identity recognition) and those in which such digital data must be retrieved for further analysis (e.g., cyberbullying).

V. CONCLUSIONS

In this paper, we investigated the robustness of two of the most used tools in AI-enhanced Digital Forensics against recent facial manipulation techniques, namely deepfakes and morphing attacks. Specifically, we assessed the vulnerabilities of the algorithms underlying the functionalities related to the analysis of facial images and videos, namely the extraction of data containing faces, detection of potential forgery, and identity recognition. The tests have been made on two publicly available datasets, employing deepfake and morphing generation processes representative of the literature. Despite the limited data and manipulation techniques involved in the analysis, the results are already indicative that such tools are not robust enough to the examined presentation attacks. Specifically, the detection of AI-based forgeries through `Magnet.AI` is unreliable on recent deepfakes and morphing attacks, which, however, revealed a limited impact on the capability of this tool of extracting facial images. Similarly, these facial manipulations significantly degrade the performance of face recognition through `Excire Photo AI`, making it unsuitable both for retrieving and discarding fake images (according to the use case) generated while considering the searched identity in the forgery process.

According to these findings, the AI-enhanced functionalities introduced so far in digital forensic tools must still be improved to be considered reliable when employed for the automatic selection of data in a specific forensic case. This is even more worrisome, considering that in our analysis we employed morphing and deepfake techniques that have

been available for several years. In particular, in this context, we need to consider the supervision and verification of the results by human experts, making false positives resulting from automatic data filtering a minor issue. The real criticality for forensic investigations is related to false negatives, i.e., unextracted facial images, undetected manipulations, and, finally, unrecognized original or altered images representing a specific individual. Observations on these show that AI-based functionalities are not yet ready to be actively integrated into real-world applications.

In addition to the performance enhancement through novel and complementary approaches [24], these tools must incorporate explainable mechanisms, namely explainable Artificial Intelligence algorithms (xAI), to support the understandability generally required in forensics [2]. Similarly, a value representing the confidence of such outcomes could further aid in calibrating the filtering of the potentially useful data by analysts [2]. Finally, it would be interesting to propose an extensive evaluation of the integration of distinct classification algorithms employed sequentially [16]. This analysis would allow the assessment of the global performance of the tool chain, therefore revealing the critical issues, and tailoring the selection of the tool to the specific use case.

This study sheds some light on the differences in the robustness of AI-enhanced digital forensic tools against novel facial manipulation techniques. Even though the analysis should be extended to further tools, data, and forgeries, we revealed the vulnerabilities and provided some guidelines to enhance the improvement of these tools. Therefore, we believe that our contribution adds a piece to the effective active role of digital forensic tools in sensitive forensics investigation scenarios.

ACKNOWLEDGMENT

This work was partially supported by Project SERICS (PE00000014) under the NRRP MUR program funded by the EU - NGEU and within the PRIN 2022 PNRR - BullyBuster 2 – the ongoing fight against bullying and cyberbullying with the help of artificial intelligence for the human wellbeing (CUP: P2022K39K8). This work was carried out while Silvia Lucia Sanna was enrolled in the Italian National Doctorate on Artificial Intelligence run by Sapienza University of Rome in collaboration with the University of Cagliari.

REFERENCES

[1] T. Bollé, E. Casey, and M. Jacquet, "The role of evaluations in reaching decisions using automated systems supporting forensic analysis," *Forensic Science International: Digital Investigation*, vol. 34, p. 301016, 2020.

[2] S. M. La Cava, G. Orrù, M. Drahansky, G. L. Marcialis, and F. Roli, "3D face reconstruction: the road to forensics," *ACM Computing Surveys*, vol. 56, no. 3, pp. 1–38, 2023.

[3] A. Vasilaras, N. Papadoudis, and P. Rizomiliotis, "Artificial intelligence in mobile forensics: A survey of current status, a use case analysis and ai alignment objectives," *Forensic Science International: Digital Investigation*, vol. 49, p. 301737, 2024.

[4] L. Sanchez, C. Grajeda, I. Baggili, and C. Hall, "A practitioner survey exploring the value of forensic tools, ai, filtering, & safer presentation for investigating child sexual abuse material (csam)," *Digital Investigation*, vol. 29, pp. S124–S142, 2019.

[5] T. Zhang, "Deepfake generation and detection, a survey," *Multimedia Tools and Applications*, vol. 81, no. 5, pp. 6259–6276, 2022.

[6] M. Ferrara, A. Franco, and D. Maltoni, "The magic passport," in *IEEE International Joint Conference on Biometrics*, pp. 1–7, 2014.

[7] U. Scherhag, C. Rathgeb, J. Merkle, R. Breithaupt, and C. Busch, "Face recognition systems under morphing attacks: A survey," *IEEE Access*, vol. 7, pp. 23012–23026, 2019.

[8] L. Verdoliva, "Media forensics and deepfakes: an overview," *IEEE journal of selected topics in signal processing*, vol. 14, no. 5, pp. 910–932, 2020.

[9] D. Yadav and S. Salmani, "Deepfake: A survey on facial forgery technique using generative adversarial network," in *2019 International conference on intelligent computing and control systems (ICCS)*, pp. 852–857, IEEE, 2019.

[10] P. Yu, Z. Xia, J. Fei, and Y. Lu, "A survey on deepfake video detection," *Iet Biometrics*, vol. 10, no. 6, pp. 607–624, 2021.

[11] G. Pei, J. Zhang, M. Hu, G. Zhai, C. Wang, Z. Zhang, J. Yang, C. Shen, and D. Tao, "Deepfake generation and detection: A benchmark and survey," *ArXiv*, vol. abs/2403.17881, 2024.

[12] P. Korshunov and S. Marcel, "The threat of deepfakes to computer and human visions," in *Handbook of Digital Face Manipulation and Detection: From DeepFakes to Morphing Attacks*, pp. 97–115, Springer International Publishing Cham, 2022.

[13] Y. Nirkin, I. Masi, A. T. Tuan, T. Hassner, and G. Medioni, "On face segmentation, face swapping, and face perception," in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, pp. 98–105, IEEE, 2018.

[14] Y. Nirkin, Y. Keller, and T. Hassner, "Fsgan: Subject agnostic face swapping and reenactment," in *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 7184–7193, 2019.

[15] W. Zhao, Y. Rao, W. Shi, Z. Liu, J. Zhou, and J. Lu, "Diffswap: High-fidelity and controllable face swapping via 3d-aware masked diffusion," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8568–8577, 2023.

[16] A. Panzino, S. M. La Cava, G. Orrù, and G. L. Marcialis, "Evaluating the integration of morph attack detection in automated face recognition systems," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3827–3836, 2024.

[17] N. Damer, A. M. Saladie, A. Braun, and A. Kuijper, "Morgan: Recognition vulnerability and attack detectability of face morphing attacks created by generative adversarial network," in *2018 IEEE 9th international conference on biometrics theory, applications and systems (BTAS)*, pp. 1–10, IEEE, 2018.

[18] E. Sarkar, P. Korshunov, L. Colbois, and S. Marcel, "Are gan-based morphs threatening face recognition?," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2959–2963, IEEE, 2022.

[19] Y. Li, X. Yang, P. Sun, H. Qi, and S. Lyu, "Celeb-df: A large-scale challenging dataset for deepfake forensics," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 3207–3216, 2020.

[20] C.-H. Lee, Z. Liu, L. Wu, and P. Luo, "Maskgan: Towards diverse and interactive facial image manipulation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 5549–5558, 2020.

[21] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *Proceedings of the IEEE international conference on computer vision*, pp. 3730–3738, 2015.

[22] T. Neubert, A. Makrushin, M. Hildebrandt, C. Kraetzer, and J. Dittmann, "Extended stirtrace benchmarking of biometric and forensic qualities of morphed face images," *IET Biometrics*, vol. 7, no. 4, pp. 325–332, 2018.

[23] "ISO/IEC 30107-3:2023(en), Information technology — Biometric presentation attack detection — Part 3: Testing and reporting."

[24] S. Concas, S. M. La Cava, G. Orrù, C. Cuccu, J. Gao, X. Feng, G. L. Marcialis, and F. Roli, "Analysis of score-level fusion rules for deepfake detection," *Applied Sciences*, vol. 12, no. 15, p. 7365, 2022.