IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 12 January 2024, accepted 15 February 2024, date of publication 23 February 2024, date of current version 1 March 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3369187

RESEARCH ARTICLE

eKYC-DF: A Large-Scale Deepfake Dataset for Developing and Evaluating eKYC Systems

HICHEM FELOUAT[1,2], HUY H. NGUYEN[1,2], (Member, IEEE), TRUNG-NGHIA LE[2,3,4], JUNICHI YAMAGISHI[1,5], (Senior Member, IEEE), AND ISAO ECHIZEN[1,2,6], (Senior Member, IEEE)

[1]Informatics Program, The Graduate University for Advanced Studies, SOKENDAI, Kanagawa 240-0115, Japan
[2]Information and Society Research Division, National Institute of Informatics, Tokyo 101-8430, Japan
[3]University of Science, VNU-HCM, Ho Chi Minh City 733000, Vietnam
[4]Vietnam National University, Ho Chi Minh City 700000, Vietnam
[5]Digital Content and Media Sciences Research Division, National Institute of Informatics, Tokyo 101-8430, Japan
[6]Department of Information and Communication Engineering, Graduate School of Information Science and Technology, The University of Tokyo, Tokyo 113-8656, Japan

Corresponding author: Hichem Felouat (hichemfel@nii.ac.jp)

This work was supported in part by the Japan Society for the Promotion of Science (JSPS) KAKENHI under Grant JP16H06302, Grant JP18H04120, Grant JP20K23355, Grant JP21H04907, and Grant JP21K18023; and in part by the Japan Science and Technology Agency (JST) CREST under Grant JPMJCR18A6 and Grant JPMJCR20D3.

ABSTRACT The reliability of remote identity-proofing systems (*i.e.*, electronic Know Your Customer, or eKYC, systems) is challenged by the development of deepfake generation tools, which can be used to create fake videos that are difficult to detect using existing deepfake detection models and are indistinguishable by facial recognition systems. This poses a serious threat to eKYC systems and a danger to individuals' personal information and property. Existing deepfake datasets are not particularly appropriate for developing and evaluating eKYC systems, which require specific motions, such as head movement, for liveness detection. Furthermore, they do not contain ID information or protocols for facial verification evaluation, which is vital for eKYC. We found that eKYC systems without the ability to detect deepfakes can be easily compromised. We have thus created a large-scale collection of high-quality fake videos (more than 228,000 videos) that are diverse in terms of age, gender, and ethnicity, plus a corresponding facial image subset. The videos include a variety of head movements and facial expressions. This large collection of high-quality diverse videos is well-suited for developing and evaluating various tasks related to eKYC systems. Furthermore, we provide protocols for traditional deepfake detection and facial verification, which are widely used in eKYC systems. It is worth mentioning that systematic evaluation of facial recognition systems on deepfake detection has not been reported. The entire eKYC-DF dataset, evaluation toolkit, and trained models are open access to researchers on GitHub: https://github.com/hichemfelouat/eKYC-DF.

INDEX TERMS Deepfake detection, electronic Know Your Customer, eKYC, facial verification, face swapping, face recognition.

I. INTRODUCTION

Identity proofing is the process of verifying an individual's identity and is a crucial aspect of many online transactions and processes. It is essential in cases where sensitive information or assets are accessed or transferred. In the past, identity proofing was typically done in person, with individuals presenting physical documents such as a driver's license or passport to prove their identity. However, with the increasing prevalence of remote work and online interactions, there is a growing need for remote identity-proofing systems [2], [3].

Remote identity proofing, or electronic Know Your Customer (eKYC), refers to verifying an individual's identity remotely; it is often used in online transactions, account creation, access to various services, and other scenarios where verifying an individual's identity is necessary. This type of

The associate editor coordinating the review of this manuscript and approving it for publication was Zhe Jin.

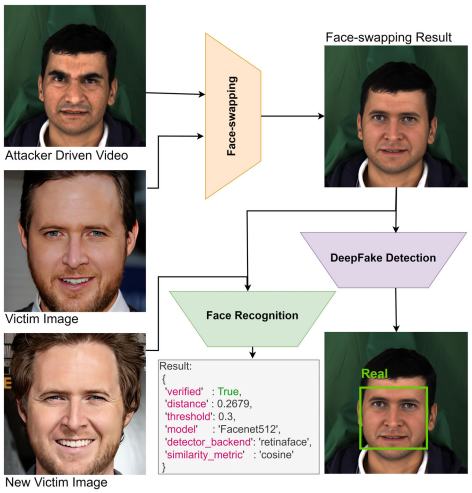


FIGURE 1. **Illustrative image of contribution provided by our eKYC-DF dataset: careful selection of source and target datasets and face-swapping methods enable the attainment of high-quality and realistic results that can fool deepfake detection models and facial recognition systems. It demonstrates that without deepfake detection or with a bad deepfake detector, eKYC systems can be easily compromised, and facial recognition systems are susceptible to high-quality deepfakes.**

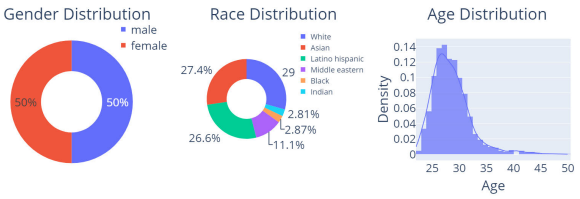


FIGURE 2. **Distribution of subjects in eKYC-DF dataset across age, gender, and ethnicity. The dataset is balanced in terms of gender and has high diversity in terms of age and ethnicity. These statistics were calculated using facebook'S DeepFace software [1].**

identity proofing is becoming increasingly common due to its convenience and flexibility.[1] It enables individuals to prove their identity from anywhere at any time, without needing to present documents or travel to a specific location physically. Various remote identity-proofing methods exist, including electronic document submission, biometric authentication, and identity verification [3], [4], [5]. Each method has advantages and disadvantages, and the most appropriate method depends on the specific needs and requirements of the organization or process in question.

Remote identity-proofing systems rely heavily on personal information and credentials, which are often stored and transmitted electronically, rendering them vulnerable to cyberattack and exploitation. The rapid advancement of enabling technologies has both positive and negative implications. While technological progress improves remote identity-proofing systems, it also provides fraudsters with tools and knowledge to manipulate these systems [3]. Identity fraud remains a widespread and concerning issue in today's digital era, encompassing the unauthorized acquisition and misuse of someone else's personal and financial data for fraudulent purposes. It can have serious consequences for an individual whose identity has been stolen or the organization that has been targeted. With the emergence of remote identity systems, the risk of identity fraud has surged, presenting substantial threats to individuals, businesses, and even governmental entities.

Deepfake attacks on video-based identity-proofing systems primarily exploit two entry points: direct presentation to the camera (physical attacks) or injection into the camera feed (logical attacks, e.g., using a virtual camera or modifying the output of a physical camera) [4]. Recent advancements enable the real-time generation of deepfake videos and interactive puppets [6]. Attackers can use their own cameras to capture video, promptly apply methods to overlay the target's face, and submit the modified content to the system. These attacks represent a huge problem for remote identity-proofing methods and their application. The section II contains further details about common steps in commercial eKYC solutions, deepfake generation, deepfake detection, and spoofing eKYC systems with deepfakes.

In this paper, we introduce the eKYC-DF dataset, a novel dataset designed to tackle the challenge of preventing identity fraud in eKYC systems. The dataset consists of real and synthetic facial videos that can be used to develop and evaluate eKYC systems in terms of deepfake detection and facial recognition systems (Figure 1). Additionally, this paper introduces five key contributions.

First, we present a comprehensive study of several critical concepts and methods related to common steps in commercial eKYC solutions, deepfake generation and detection, common datasets in deepfake detection, and spoofing eKYC with deepfakes. We offer a valuable resource for researchers and developers who aim to counter the increasing risk of deepfakes in the eKYC process by examining and analyzing these crucial areas. Our study brings together a wealth of information and insights into these complex topics, providing a solid foundation for future research and development in this field.

Second, our dataset is larger than others commonly used in deep learning applications, which makes it more effective for pattern recognition and generalization (see Table 1). Its scale enables robust model training by capturing data distribution complexity, enhancing performance on diverse tasks, and mitigating overfitting concerns.

Third, the diversity of our dataset in terms of age, gender, and ethnicity (Figure 2) ensures robust deep learning model training and bias minimization, thereby enhancing prediction accuracy across diverse individuals. Furthermore, the videos contain complex head movements and different camera poses with various facial expressions. It captures group complexity by encompassing various demographics, ensuring generalization beyond specific subsets. This diversity aids in identifying and rectifying biases, fostering a fair and equitable model for all.

Fourth, our dataset contains high-quality images and videos (Figure 13), which greatly enhance the training of deep learning models. They enable models to learn more

[1] https://www.enisa.europa.eu/publications/enisa-report-remote-id-proofing

complex subject features, resulting in more accurate and reliable output. Including diverse and high-quality images and videos ensures that the models generalize well to new and unseen data, thereby enhancing the robustness of deepfake detection models and face recognition systems.

Fifth, the paper also introduces a benchmark for assessing the dataset's effectiveness in detecting deepfakes and matching faces. A thorough and comprehensive evaluation against the most recent deepfake detection and face-matching models in various scenarios demonstrated that the eKYC-DF dataset is a valuable resource for enhancing deepfake detection models and face recognition systems. We will provide open access to the entire dataset, evaluation toolkit, and trained models upon acceptance of this paper.

II. RELATED WORK

Recent significant advancements in computer vision and image generation using generative adversarial networks and diffusion models combined with their malicious use for manipulating faces in images, spreading fake news, and hacking remote identity-proofing systems that rely on a user's face for proofing have created an urgent need for methods that can reliably detect face manipulation [3]. Many efforts have been devoted to creating face forgery detection datasets to train deep learning models [7] to address this need.

A. COMMON STEPS IN COMMERCIAL eKYC SOLUTIONS

Commercial electronic Know Your Customer (eKYC) solutions typically involve several steps to verify an individual's identity remotely. The most common steps include ID verification, face matching, and liveness detection,[2] as shown in figure 3.

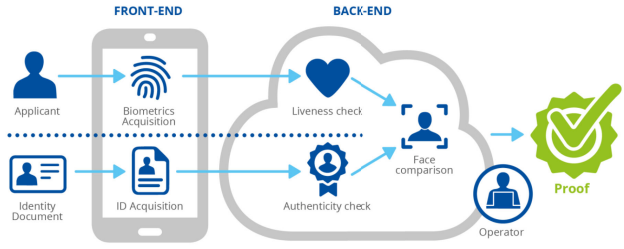


FIGURE 3. A generic proofing method diagram in eKYC systems. The applicant is physically present with the device and ID at the front end. The back end performs a high confidence match between the liveness-proven photo or video evidence and the face shown on the ID document or in the NFC chip to prove the identity.

ID Verification: The first step is to verify the customer's identity by validating the provided identity documents, such as a passport, national ID, and/or driver's license. This can be done using various methods such as optical character recognition (OCR), barcode scanning, or manual entry. The system checks the authenticity of the documents and ensures that they have not been altered.

[2]https://www.enisa.europa.eu/publications/enisa-report-remote-id-proofing

Face Matching: Once the ID has been verified, the customer is asked to take a selfie image or video, which is then compared with the photo on the ID to ensure that the person presenting the ID is the same person in the image. This is typically done using facial recognition technology. The system analyzes the customer's facial features and compares them to the photo on the ID document to ensure they are the same person.

Liveness Detection: Liveness detection is used to ensure that the person is physically present and not using a photo or video of someone else. This involves asking the person to perform a series of actions or movements, such as blinking or smiling, to demonstrate that they are a live human being and not a digital artifact. This is done using computer vision and machine learning algorithms that can detect the subtle movements and expressions of the person in the video.

These steps are designed to provide high security and accuracy in remote identity verification. By combining different types of technology, such as OCR, facial recognition, and liveness detection, eKYC solutions ensure that the customer is who they claim to be and that they are physically present during the verification process. This helps businesses comply with regulatory requirements while providing a streamlined and convenient customer onboarding experience.

B. DEEPFAKE GENERATION

Facial manipulation refers to a range of techniques used to alter the appearance of a person's face. These techniques are becoming increasingly used due to technological advances and the increased use of social media platforms. Facial manipulation can be used for various purposes, including entertainment, research, and forensics. Different types of facial manipulation methods are available (Figure 4), and we discuss several common ones in this section.

Face synthesis involves creating an entirely new synthetic face from scratch, often by using GANs or diffusion models such as the stable diffusion approach [9], in which a robust diffusion model is typically used. This type of manipulation can generate a wholly new identity or create a likeness of an existing person [10]. The process involves training a deep learning model on a large dataset of faces and then using the model to generate a new face. This technique can be used to create a new identity for use in video games or movies or to create an avatar for virtual reality environments; however, it can also be misused, such as for creating highly convincing fake profiles on social networks to spread disinformation.

Face replacement involves replacing one person's face with another's face. This can be done in two ways: Transfer involves replacing a source person's facial features with a target person's facial features while retaining the source person's facial expressions and movements [11] [12]). Identity swap involves replacing a source person's entire identity with a target person's identity [13]. This type of manipulation is commonly used in movies and television to

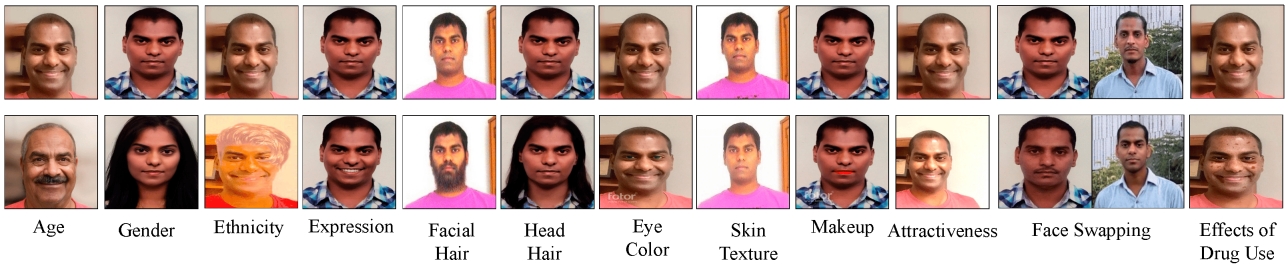


FIGURE 4. Examples of different face manipulations. The first row shows original samples, and the second row shows manipulated samples [8].

create visual effects and can also be used to create fake videos for malicious purposes.

Facial editing and attribute manipulation involve modifying specific facial features, such as the nose's shape, the hair's color, and the skin's color, or the individual's attributes, such as age, gender, and ethnicity [14]. Facial editing and attribute manipulation have become increasingly popular, particularly in the beauty industry. They enable customers to visualize how they would look with different hair colors, hairstyles, and makeup [15]. This type of manipulation can be used to alter a person's appearance in a video or image for artistic or malicious purposes [16]. For example, a deepfake video could be created to make a person appear to be wearing different clothing or appear to be of another ethnicity. The manipulation process is commonly carried out using a generative adversarial network (GAN), such as a StarGAN [17], STGAN [18], or InstructPix2Pix [19].

Facial reenactment involves creating a digital version of a person's face that can be used to mimic their facial expressions and movements [20]. This type of manipulation can make it appear as though a person is saying or doing something they did not say or do. The process involves training a deep learning model on a large dataset of facial expressions and movements and then using the model to generate new facial expressions and gestures for a person in a video. This type of manipulation can be used for malicious purposes, such as creating fake videos of politicians or celebrities [16]. There have been several outstanding developments in this field, including MegaPortraits [21], MetaPortrait [22], and PV3D [23].

C. DEEPFAKE DETECTION

Deepfake detection is an evolving challenge due to the increasing sophistication of fake videos, which are often difficult to distinguish from authentic ones. Binary classification is the most common method used to distinguish real and fake videos; an extensive dataset of training samples is required to develop accurate classification models. This section presents a comprehensive survey of methods for detecting deepfakes, focusing on video-based ones, notwithstanding that there are also methods that use images for detecting deepfakes [24], [25]. We group the video-based methods into two main categories: features-based and deep learning-based.

1) FEATURES-BASED METHODS

Features-based methods use the visual features of a video to distinguish real and fake videos. These methods rely on deepfake videos with unique features that distinguish them from real videos. According to Zhang [25], these features can be classified as biometric, model, or media.

In the context of deepfake detection, biometric features refer to the physical and behavioral characteristics that can be analyzed to determine an image's or video's authenticity. These features include eye blinking, lip-syncing, facial and head movements, head pose, color, texture, and shape. Each feature can be used to detect a deepfake video in different ways. Eye blinking is a key biometric feature that can be used to detect deepfake videos. People in deepfakes typically blink less frequently than those in untampered videos. A healthy adult human normally blinks somewhere between every 2 to 10 s, and each blink lasts 0.1 to 0.4 s. Lip-syncing is another important biometric feature that can be used to detect deepfake videos. In natural videos, the movement of the lips is synchronized with the spoken words, whereas in deepfake videos, the movement of the lips may not match the words being spoken, or there may be a delay between the two. This is a sign that the video is a deepfake. Facial expressions, head movements, and head poses are also critical biometric features that can be used to detect deepfake videos. Facial expressions, head movements, and head poses are synchronized and consistent with spoken words in natural videos. In contrast, in deepfake videos, the facial expressions, head movements, and head poses may not match the spoken words or may be inconsistent with the speaker's tone or emotion. Deepfake detection algorithms can identify whether the actions depicted in a video are consistent with those of the depicted person. The face's color, texture, and shape consistently follow the speaker's movements in natural videos, whereas in deepfake videos, these characteristics can be manipulated, leading to an artificial appearance. This can serve as a means of distinguishing between real and fake videos.

Model features in deep fake detection refer to the specific characteristics or patterns that a model left in the generated data. The most common deepfake generation methods rely on deep learning techniques, particularly GANs, to produce convincing fake images and videos. However, these fake media still contain model features that can be used to identify

them. In GAN-generated media, specific model fingerprints are present, which can be used to detect deepfake images and videos created with GANs. For instance, GAN fingerprints have been used to identify fake images and videos produced using GAN-based methods [26]. These fingerprints include unique patterns in the noise space of images generated by GANs [27] and convolutional traces generated during the GAN generative process [28]. Therefore, deepfakes can be detected by analyzing the model features present in the generated media, effectively mitigating the risks associated with deepfakes.

Certain media features, such as temporal information, inconsistencies between frames, and noise artifacts, can be used to identify deepfake media. Sabir et al. used the temporal information available in media streams to detect face manipulation in videos [29]. They found that deepfakes often exhibit unnatural changes in facial expressions that are not typical of real faces. Similarly, Li and Lyu used face-warping artifacts resulting from inconsistent illumination between frames to detect fake videos [30]. They discovered that deepfakes exhibit unnatural lighting changes within and across scenes. These media features can provide valuable clues for detecting deepfakes and mitigating their potential harm.

2) DEEP LEARNING-BASED METHODS

Deepfake videos have limited resolution, so an affine face-warping approach is required to match their configuration to that of the original one. The creation process involves scaling, rotation, and shearing of the face area, which can leave artifacts that deep learning models can detect, Figure 5. Deep learning-based detection methods typically involve training a deep neural network on a dataset of real and deepfake videos and then evaluating its performance on a separate test set. Deep learning methods are widely used for deepfake detection because they can automatically extract high-level features that are difficult to define manually. Moreover, deep learning methods can learn from large amounts of data, improving their generalization ability. Deep learning-based detection methods are also robust to noise and distortion, which are common in real-world scenarios.

FIGURE 5. Artifacts and weaknesses in fake images that limit their naturalness and facilitate fake detection [31].

Deep learning methods have shown promising results in detecting deepfake videos. The most commonly used deep learning architectures for deepfake detection include GANs. Prajapati and Pollett presented a method called MRI-GAN based on the GAN architecture, which is commonly used to generate realistic images and videos [32]. However, instead of generating new images or videos, MRI-GAN is used to learn the distribution of real images and to detect deviations from this distribution indicative of a deepfake.

Convolutional neural networks (CNNs) are commonly used for image and video analysis and can be trained to distinguish real and deepfake videos based on the presence of artifacts. Zhao et al. [33] presented an exciting approach based on a CNN. A multi-attentional network architecture consisting of three main components is used to detect deepfake videos. First, multiple spatial attention heads are used to enable the network to focus on different local parts of the input image. Second, a textural feature enhancement block is used to zoom in on subtle artifacts in shallow features. Finally, the low-level textural features and high-level semantic features are combined using attention maps. A method based on this approach outperformed state-of-the-art deepfake detection methods on several benchmark datasets.

Bonettini et al. presented an approach for detecting face manipulation in videos by using an ensemble of CNNs [34]. Recurrent neural networks (RNNs) can be used to analyze the temporal patterns in videos and to detect deepfake videos with inconsistent motion or lip-syncing. Sabir et al. proposed using recurrent RCNNs to detect face manipulation in videos by using a two-step approach [29]. The first step involves detecting, cropping, and aligning faces in a sequence of frames. The second step combines CNN and RNNs to distinguish manipulated and real face images. Their method is based on recurrent convolutional strategies and improves the accuracy of face manipulation detection in videos.

Vision transformers (ViTs) have also been proposed for deepfake detection as they can learn global features from videos and are less vulnerable to overfitting. Miao et al. introduced a method for detecting manipulated faces that enhances generalization and robustness through the use of the bag-of-local features approach [35]. Their method extends the transformer model by incorporating a bag-of-features strategy that captures an image's local characteristics by dividing the image into smaller regions and extracting features from each of them to learn local forgery features without explicit supervision. Wang et al. presented the Multi-modal Multi-scale Transformers (M2TR) method for capturing subtle manipulation artifacts at different scales using transformer models [36]. The M2TR model operates on patches of different sizes to detect local inconsistencies in images at different spatial levels. Using a cross-modality fusion block, the model learns to detect forgery artifacts in the frequency domain, which complements RGB information. Their proposed method demonstrated promising results in

detecting deepfake images, outperforming several state-of-the-art methods.

D. SPOOFING eKYC WITH DEEPFAKES

The KYC systems used by organizations to electronically verify the identity of customers rely on various methods to authenticate the person's identity, including identity documents and face matching. However, these methods can also be targeted by attackers who aim to exploit weaknesses in the system to gain access to sensitive information or commit fraud. There are various attack methods related to identity documents and face matching.

1) ATTACK METHODS RELATED TO IDENTITY DOCUMENTS

Identity documents are important tools for verifying a customer's identity. However, they can also be targeted by attackers who may attempt to use them for fraudulent or illegal purposes. Attackers can use several methods to compromise eKYC systems by exploiting identity documents.[3]

Modify one or more parts of an authentic identity document: In this attack, an attacker obtains a genuine identity document (e.g., a passport or driver's license) and modifies one or more parts of it, such as the photo, name, or expiration date. This enables the attacker to use the document to impersonate someone else or to make an expired document appear valid.

Produce a complete identity document for a real identity: In this attack, an attacker creates a complete replica of an existing identity document, including all of the information it contains. This can be achieved through various means, such as using sophisticated printing techniques or stealing the personal information needed to create a fake document.

Produce a complete identity document for a fictional identity: In this attack, an attacker creates a completely fictitious identity and produces a document to support it. This can be done by creating a false identity from scratch and producing a corresponding identity document.

Produce a complete identity document for a partially real, partially fictional identity: In this attack, an attacker may create a fake identity document that contains real information (such as a correct name or date of birth) and false information (such as an incorrect address or nationality).

Create a fantasy identity document from scratch: In this attack, an attacker creates a fictitious identity document with no basis in reality. This can be achieved through various means, such as by creating a false identity from scratch and producing a corresponding identity document.

2) ATTACK METHODS RELATED TO FACE MATCHING

Face matching has become an increasingly important tool in the field of security, authentication, and identification. However, the accuracy and reliability of face-matching systems can be compromised by various attack methods. There are several types of attacks related to face-matching systems, including photo attacks, video-of-user replay attacks, 3D mask attacks, and deepfake attacks. Figure 6[4] illustrates the distribution of different face-matching attacks observed in real-world scenarios. As can be seen, the majority of attacks fall under the category of 3D mask attack (38%) and Deepfake attack (25%). The remaining attacks are more evenly distributed, with Photo attack (13%) and Video-of-user replay attack (13%). Notably, other attacks represent the smallest category, accounting for only (11%) of observed incidents that encompass a variety of unspecified methods.

Photo attack: One of the simplest and most common types of attacks related to face-matching systems is the photo attack. In this type of attack, the attacker uses a printed or digital photo of the target to bypass the face recognition system [37]. This method is often effective as many face-matching systems cannot distinguish between a real face and a printed or digital photo of a face. To carry out a photo attack, the attacker uses a high-quality photograph of the target obtained from a social media profile, an ID card, or other sources. The attacker then prints out the photograph or displays it on a digital screen and presents it to the face-matching system to gain unauthorized access.

Video-of-user replay attack: Another type of attack related to face-matching systems is the video-of-user replay attack. In this type of attack, the attacker records a video of the target person's face and then replays the video in front of the face recognition system [37]. This attack is similar to the photo attack, but the video provides a more realistic representation of the target's face and may be more difficult for the system to detect. To carry out such an attack, the attacker uses a high-quality video of the target obtained from a surveillance camera, social media profile, or other source. The attacker then plays the video on a screen in front of the face recognition system to gain unauthorized access.

3D mask attack: A 3D mask attack is a more sophisticated attack method as it involves the creation of a three-dimensional mask that resembles the target's face. The attacker creates a physical mask that is a realistic representation of the target's face by using 3D printing technology or other method [38]. The attacker first obtains a high-quality photograph or video of the target's face. The image data is then used to create a 3D model of the target's face, which is used to print out a physical mask that matches the target's facial features. The attacker can then wear the mask and present it to the face-matching system for unauthorized access.

Deepfake attack: A deepfake attack involves using artificial intelligence (AI) and machine learning technology to

[3]https://www.enisa.europa.eu/publications/enisa-report-remote-id-proofing

[4]https://www.enisa.europa.eu/publications/enisa-report-remote-id-proofing

create highly realistic fake images and videos. The attacker may use deep learning algorithms to generate images and videos closely resembling the target's face and movements. The attacker first obtains a high-quality photograph or video of the target's face and then uses deep learning algorithms to generate a highly realistic fake image or video of the target. The attacker then performs a presentation (physical attack) or logical attack using the deepfake image or video to gain unauthorized access.

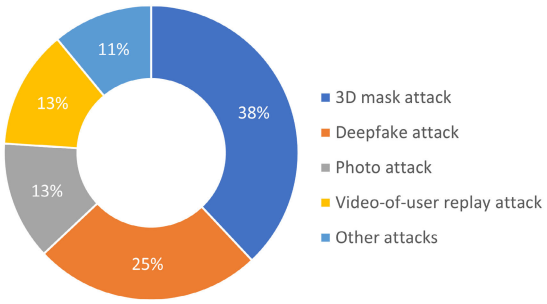


FIGURE 6. Most effective types of face-matching attacks.

E. COMMON DATASETS IN DEEPFAKE DETECTION

Since there are no special datasets for detecting face manipulation in remote identity-proofing systems, here we review the existing fake video datasets in chronological order of their introduction. We did not take into account the popularity of each dataset among researchers.

The UADFV dataset [39] contains 49 real videos, while 49 fake videos are generated using a deepfake method.

The DF-TIMIT dataset [40], which was generated based on the VidTIMIT[5] dataset, contains 10 videos for each of 43 subjects, who faced the camera and spoke short predetermined phrases. From these 43 subjects, the authors manually selected 16 subject pairs with comparable appearances and generated 10 fake videos for each of the 32 subjects using 2 versions of faceswap-GAN,[6] one for low-quality and the other for high quality. A total of 640 videos were generated.

The FaceForensics++ (FF++) dataset [7] contains fake videos generated from 1000 real videos selected from YouTube. They were used as pristine data to generate a large-scale manipulation dataset. The fake videos were synthesized by applying 4 face manipulation methods to the selected real videos, resulting in 4,000 manipulated videos.

The Deepfake Detection Dataset by Google & Jigsaw is a large-scale dataset of visual deepfakes that was created to support deepfake detection research [41]. The dataset consists of 3,068 fake videos generated from 363 videos of paid and consenting actors.

The Celeb-DF dataset [42] is a large-scale deepfake video dataset containing fake videos generated from 590 real videos selected from YouTube, which are in the form of interviews of 59 celebrities with diverse distributions in terms of age, gender, and ethnicity. The 5,639 high-quality celebrity deepfake videos in this dataset were generated using an enhanced synthesis process.

DeeperForensics-1.0 (DF-1.0) [43] is a large-scale dataset for real-world face forgery detection. This dataset was made more suitable for real-world fake face detection by focusing on quality, scale, and diversity in its creation. Source videos were carefully collected from 100 paid and consenting actors from 26 countries; 1,000 target videos were also collected from the FF++ dataset. One thousand fake videos were generated using only one face-swapping method. To compensate for using only one method, augmentation was applied to both the real and fake videos, generating 50,000 real and 10,000 fake videos.

The DeepFake Detection Challenge (DFDC) dataset [44] is one of the largest face-swap video datasets currently available to the public. It is a collaborative project between Facebook and other companies aimed at promoting competition among deepfake detection researchers and publication of their findings, as well as creating a large and useful dataset. It contains more than 120,000 manipulated videos created from more than 23,000 real videos of 960 volunteer subjects of different ages, genders, and ethnicities in different environmental settings. The manipulated videos were generated using eight different methods.

The recently created WildDeepfake dataset [45] was designed to help researchers develop deepfake detectors and evaluate their performance against real-world deepfakes. It comprises 7,314 face sequences from 707 deepfake videos acquired entirely from the Internet. The creators started by identifying the face region in each video frame; a pretrained model was then used to extract features from each face region. Next, a facial landmark was extracted and used to align the faces in the frame sequence.

The Korean DeepFake Detection Dataset (KoDF) [46] is a large-scale collection of real and fake videos focusing on individuals of Korean ethnicity. It includes 403 subjects, 62,166 real videos, and 175,776 fake videos. All original videos in the dataset were obtained from paid volunteers. The creators controlled the distribution of the 403 participants by age, gender, and recording site to maximize the diversity of the dataset. Six methods were used to generate deepfake videos. Finally, manual screening was performed to ensure the quality of the videos; the eyes and ears were used to cross-check each real and fake video for potential problems.

The recently created ForgeryNet dataset is an enormous face forgery dataset that can be used for face forgery analysis at both the image and video levels [47]. Fifteen manipulation methods were used to produce both fake images and videos, with 1,438,201 subjects for real images and 99,630 for real videos, yielding 1,457,861 fake images and 121,617 fake videos. The source data were chosen from four face datasets to increase the variety in terms of identity, angle, expression, scenario, and so on.

[5]https://conradsanderson.id.au/vidtimit/
[6]https://github.com/shaoanlu/faceswap-GAN

The novel DeepFake MNIST+1 face animation dataset [48] includes 10,000 source images selected as frames from videos in the VoxCeleb1 dataset [49] and driving videos used to animate face images with actions were selected from the Amsterdam Dynamic Facial Expression Set (ADFES) [50]. The first-order motion model for image animation [51] was used to generate videos. The total number of generated fake videos is 10,000; they portray 1 of 10 actions and depict various emotions.

The recently released FakeAVCeleb audio-video deepfake dataset [52] includes not only typical deepfake videos but also lip-synced fake audio. Five hundred source videos were selected from the VoxCeleb2 dataset [53]. They were equally distributed in terms of age, gender, and ethnicity; the selection was based on three requirements: the subject must be the only person in the video, the face must be clear and in focus, and a hat, glasses, mask, or other object must not obscure the face. The source videos were used to generate around 20,000 deepfake videos by using four deepfake generation methods.

OpenForensics dataset is a comprehensive collection designed to present significant challenges, particularly in the domain of face forgery detection and segmentation [54]. The dataset incorporates detailed face-wise annotations, enhancing its potential for deepfake prevention and general human face detection research. Furthermore, OpenForensics has established a set of benchmarks for these tasks, thoroughly evaluating cutting-edge instance detection and segmentation methods across various scenarios.

An existing deepfake face can be swapped with another face. This face-swapping process can be repeated multiple times, leading to the development of highly advanced deepfakes that effectively deceive deepfake detection methods. This problem was addressed by the development of DeePhy [55], a deepfake phylogenetic dataset comprising 5,040 deepfake videos generated from 100 source videos using three distinct deepfake generation methods. Specifically, it consists of 840 videos containing deepfakes swapped once, 2,520 videos containing deepfakes swapped twice, and 1,680 videos containing deepfakes swapped three times.

The Glitch in the Matrix dataset is a large-scale audio-visual deepfake dataset that contains 136,304 videos, of which 36,431 are real and 99,873 are fake [56]. The dataset focuses on content-driven audio-visual forgery, where the manipulations are guided by relevant words in the video transcripts. Specifically, the manipulation strategy is to replace strategic words with their antonyms, which can significantly change the statement's perceived sentiment.

DF-Platter [57] is an extensive and meticulously annotated dataset comprising low-resolution and high-resolution deepfake videos. These videos were generated from 764 source videos using three distinct deepfake generation methods and encompass single-subject as well as multiple-subject scenarios. The dataset encompasses diverse facial attributes, including gender, age, skin tone, and occlusion. It is an impressive collection of 133,260 videos.

The SWAN-DF dataset is a new public dataset of realistic audio-visual deepfakes [58]. It is the first dataset of its kind to include both face and voice deepfakes, and it is specifically designed to assess the vulnerability of automatic identity recognition systems to these types of attacks.

We quantitatively compared our eKYC-DF dataset with these available datasets; the results are compared in Table 1.

TABLE 1. Quantitative comparison of eKYC-DF dataset with existing publicly available deepfake datasets.

Dataset	Release Year	Real Videos	Fake Videos	Total Videos	Total Subjects	Methods
UADFV [39]	2018	49	49	98	49	1
DF-TIMIT [40]	2018	320	640	960	32	2
FF++ [7]	2019	1,000	4,000	5,000	-	4
GoogleDFD [41]	2019	363	3,068	3,431	363	1
CelebDF [42]	2019	590	5,639	6,229	59	1
DF-1.0 [43]	2020	50,000	10,000	60,000	100	1
DFDC [44]	2020	23,654	104,500	128,154	960	8
WildDeepfake [45]	2020	3,805	3,509	7,314	-	-
KoDF [46]	2021	62,166	175,776	237,942	403	6
ForgeryNet [47]	2021	99,630	121,617	221,247	5,400	15
MNIST+ [48]	2021	10,000	10,000	20,000	10,000	1
FakeAVCeleb [52]	2021	500	19,500	20,000	500	4
DeePhy [55]	2022	100	5,040	5,140	-	3
LAV-DF [56]	2023	36,431	99,873	136,304	153	1
DF-Platter [57]	2023	764	132,496	133,260	454	3
SWAN-DF [58]	2023	16	960	976	16	2
eKYC-DF	2023	**760**	**228,000**	**228,760**	**100**	**3**

III. eKYC-DF DATASET

The eKYC-DF dataset is developed to serve as a large public deepfake dataset for developing and evaluating eKYC systems against deepfake attacks, Figure 7. In addition to being large, it provides protocols for evaluating deepfake detection models and facial recognition systems. The dataset is diverse in terms of age, gender, and ethnicity, and the images and videos are very high quality. Furthermore, the videos include a variety of head movements and facial expressions. Existing deepfake detection datasets are not necessarily useful for developing and evaluating eKYC systems, which require a specific motion, such as head movement, for liveness detection, Figure 8. This section discusses the steps in creating the dataset using face-swapping methods, as shown in the flowchart in Figure 9. We also provide brief information on the tools and datasets we used to create the eKYC-DF dataset.

A. ASSUMPTIONS

We assume that attackers aim to bypass facial liveness challenges despite facing resource constraints. They primarily acquire victim portrait images from available online sources, such as social media profiles, personal websites, news articles, or ID documents. Obtaining portrait images is generally considered easier than acquiring video recordings. Additionally, attackers are assumed to possess the technical skills necessary to perform a logical attack on the facial liveness detection system using a virtual camera, but they

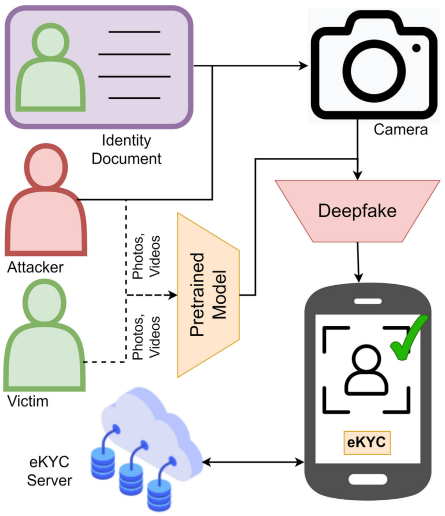


FIGURE 7. This diagram illustrates the deepfake attack on an eKYC system. The attacker uses a forged or stolen identity document of the victim and then employs a deepfake model to create a synthetic video of the victim, which is accomplished by either animating a still image of the victim's face or real-time swapping the victim's face onto an attacker's face. As a result, the attacker uses this video to impersonate the victim and get unauthorized access to the eKYC system.

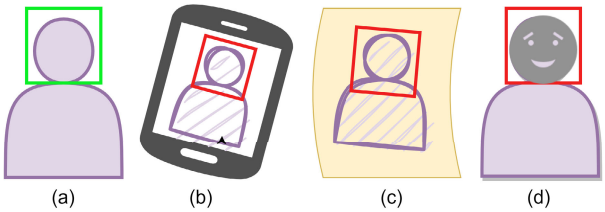


FIGURE 8. Liveness detection results against different attacks: (a) real face, (b) replay attack, (c) photo attack, and (d) 3D mask attack. Green bounding boxes indicate real faces, while red bounding boxes indicate fake faces.

lack deep expertise in machine learning and deep learning. Since neural filters are widely used these days, it may be worth mentioning that it is a reality that Victim images can also be somewhat processed and enhanced by neural filters.

B. VICTIM IMAGE

Many datasets are available for deepfake generation, *i.e.*, face swapping. Our objective was to create a dataset with a sufficient number of images with sufficient quality to train state-of-the-art deep learning models on face-swap detection tasks, especially those used in eKYC systems. We used the VGGFace2-HQ dataset,[7] which is a high-resolution version of the VGGFace2 dataset developed for academic face editing [59]. The VGGFace2-HQ dataset was generated on the basis of the GFPGAN method, which is used to restore high-quality faces from counterparts with poor quality due to low resolution, noise, blur, compression artifacts, etc. [60]. It uses InsightFace for data preprocessing, such as cropping the faces and aligning them [61]. We selected 100 uncorrupted and high-resolution images balanced between male and female, with even distributions by age and ethnicity, as verified manually using auxiliary tools. Example images are shown in Figure 10.

[7]https://github.com/NNNNAI/VGGFace2-HQ.git

C. ATTACKER DRIVEN VIDEO

During automatic authentication, the target person might be asked to move their head, open their mouth, or close their eyes. In addition, they can face the camera in different poses with various facial expressions. Therefore, to create more realistic fake videos, we must consider these movements when choosing an appropriate dataset, *i.e.*, which contains various head and facial movements from which to select the target input.

Among the available datasets, we found that the DeeperForensics-1.0 dataset was the most suitable for our objective. This large-scale dataset for real-world face forgery detection has three important characteristics that make it suitable as the source of our target data: good quality videos, a large number of videos, and highly diverse videos; also, the actors were asked to speak naturally to avoid excessive frames showing a closed mouth. As described above, the source videos in this dataset were carefully collected from 100 paid and consenting actors from 26 countries [43]. The example images in Figure 11 illustrate the diversity of the images in terms of identity, pose, expression, and illumination. To construct an input target, we selected only one light position (light uniformity), eight expressions (anger, contempt, disgust, fear, happy, neutral, sad, surprise), and one camera position (camera front). Hence, the number of target videos was 800 (100 actors $\times$ 1 light position $\times$ 8 expressions $\times$ 1 camera position); however, several videos were missing from the original dataset, so only 760 videos were obtained.

D. PREPROCESSING

We used three algorithms related to face processing in the preprocessing task to crop, align, and resize the face area to the size required for each of the three face-swapping methods we used (two require 512 x 512 pixels; one requires 224 x 224 pixels). This was done either from input images or input videos (after extracting the target video frames). All three face-swapping methods have ready-made functions for input preprocessing.

E. FACE-SWAPPING METHODS

The face-swapping methods we used to build our dataset were carefully selected based on reviewing the relevant literature and considering three key factors. First, we chose open-source methods to ensure accessibility and transparency. Second, we focused on methods that can be used for zero-shot inference, which enables face swapping without requiring specific training data. Third, we evaluated each

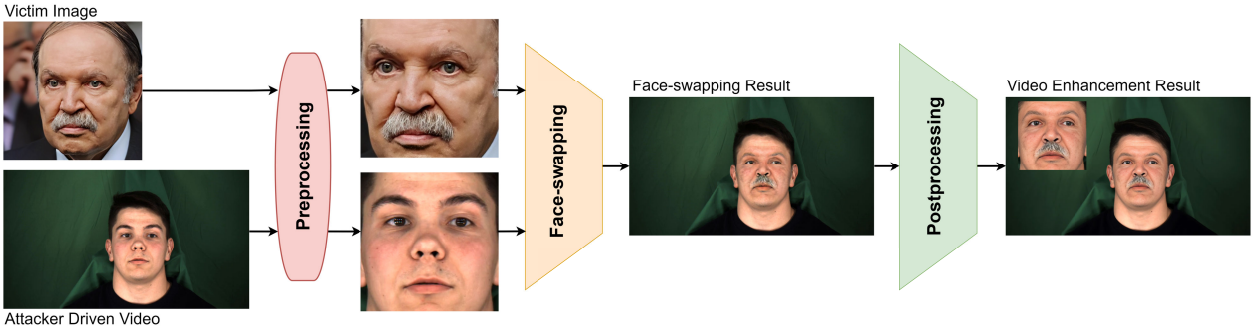


FIGURE 9. Steps in creating eKYC-DF dataset: 1) preprocessing of the victim image (images from VGGFace2-HQ dataset) and attacker-driven video (videos from DeeperForensics-1.0 dataset); 2) swapping of faces between victim and attacker images; 3) postprocessing to enhance faces in manipulated videos.

FIGURE 10. Example source images showing diversity for age, gender, and ethnicity.

FIGURE 11. Example target images showing diversity for identity, pose, facial expression, and illumination [43].

method's performance in terms of its ability to generate realistic results appropriate for the type of images and videos we needed to build our dataset. The evaluation process involved manual observation and direct comparison of several available methods. Table 2 presents a quantitative comparison of identity retrieval and pose error on the Faceforensics++ dataset with the existing state-of-the-art face swap models. After careful consideration, we selected three that met our selection criteria: SimSwap [11], FaceDancer [12], and SberSwap [13]. Despite the existence of face animation-based attacks, our search for open-source projects that met our criteria (open-source models, zero-shot inference, and the ability to generate realistic results) for constructing our dataset was unsuccessful, as shown in Figure 12.

TABLE 2. Quantitative comparison with the existing SoTA face swap models on Faceforensics++ dataset [7].

Method	Identity retrieval ↑	Pose error ↓
FaceSwap [62]	54.19	2.51
DeepFakes [63]	77.65	4.59
FaceShifter [64]	97.38	2.96
MegaFS [65]	90.83	2.64
FaceController [66]	98.27	2.65
HifiFace [67]	98.48	2.63
SberSwap [13]	98.67	3.00
SimSwap [11]	92.83	1.53
FaceDancer [12]	98.84	2.04

FIGURE 12. A visual illustration using the First-Order Motion Model (FOMM) [51] highlights the occurrence of significant distortions and artifacts in facial animation during complex head movements. Consequently, the resulting output falls short of the high-quality standards required for our dataset.

SimSwap is an effective framework aimed at generalized and high-resolution face swapping using a single trained model; it can accomplish arbitrary face swapping on images and videos [11]. Furthermore, it can maintain the attributes of the target face while transferring the identity of any arbitrary source face into any arbitrary target face, unlike previous methods, which either lack the capacity to generalize to arbitrary identities or fail to maintain features such as facial expression and glance direction. It overcomes these weaknesses in two ways. It transfers the identity information of the source face into the target face at the feature level by using a unique injection module. Then, it uses the weak feature matching loss, which implicitly helps the framework preserve the facial attributes.

FaceDancer is a model for performing high-fidelity face swapping, with the ability to consider pose and occlusion [12]. FaceDancer enables one person's face in a video to be replaced with another person's face while preserving the pose and facial expression of the original person (*e.g.*, glasses, hats) of the original face, which makes the resulting face swap even more realistic. This can be useful for various applications, such as creating realistic digital avatars or enabling deepfake videos. FaceDancer was designed to be highly accurate and to produce results that are difficult to distinguish from real video.

SberSwap is a face-swapping pipeline based on the FaceShifter architecture with several enhancements to improve the final result [13], and it fixes the problems inherent in previous architectures [68], [69]. The model's architecture and other training elements have been the subject of extensive research and testing by the developers. The improvements in quality shown by evaluation are attributed to the usage of a new special eye loss function, super-resolution block, and Gaussian-based face mask generation.

F. POSTPROCESSING

All of the aforementioned methods produced manipulated videos matched to the target input videos after inference, with the faces in the target input videos being swapped for the faces in the source input videos. Since manipulated videos can suffer from unwanted artifacts and distortions, enhancing them to remove distortions and improve resolution is necessary. Three face restoration algorithms were thus used to enhance the manipulated videos. We recommend using one of them to enhance videos before using them in research.

1) MAXIM

Multi-Axis MLP for Image Processing (MAXIM) is a generic network for the restoration and enhancement of images. It was inspired by current developments in transformer and multi-layer perceptron (MLP) models, which produce unique network architecture designs for computer vision applications [70]. Long-distance interactions are supported by MAXIM, which uses a UNet-shaped hierarchical structure that can be used as a general-purpose vision backbone for image-processing tasks. It is both efficient and adaptable.

2) GFPGAN

The GFPGAN method is a framework that uses rich and diverse priors contained in StyleGAN2, as well as the powerful generative face prior (GFP) and delicate designs to restore facial details and enhance colors with only a single forward pass while maintaining a good balance between realism and fidelity [60]. The input to GFPGAN is a face image suffering from unknown degradation; face restoration aims to estimate a high-quality image as similar as possible to the ground truth image in terms of reality and fidelity.

3) DIFFACE

Blind Face Restoration with Diffused Error Contraction is a method for restoring degraded or low-resolution facial images [71]. DIFFACE uses a deep neural network to perform the restoration; it can improve the resolution and quality of images without the need for a high-resolution reference image.

IV. EVALUATION

We evaluated the eKYC-DF dataset by focusing on our primary objectives for this dataset as framed in three research questions:

RQ1: Does the dataset offer high-quality deepfake content representing a realistic and challenging scenario?

RQ2: To what extent do facial recognition systems encounter difficulties in accurately distinguishing between real and face-swapped images in the dataset?

RQ3: Does the dataset present novel challenges and pose an effective benchmark for developing and evaluating deepfake detection models?

A. EVALUATION DATASET

We used a scaled-down version of the eKYC-DF dataset due to it being very large (the small version is over 700 GB), so the process of enhancing and cropping faces is extremely time-consuming. This scaled-down dataset consists of 6,000 fake videos (2,000 videos from each of the three face-swapping methods we used) and an equal number of real videos processed using the GFPGAN method to enhance visual quality. The videos (fake and real) were provided with three different compression levels (C0, C23, and C40) to match various real-world scenarios. To focus specifically on facial analysis, we used MediaPipe [72] to crop the faces from all videos accurately. To capture temporal information and enable dynamic analysis, we extracted 60 consecutive frames from each video, resulting in a rich and varied collection of facial expressions and movements for

comprehensive evaluation. Furthermore, to facilitate robust model training and evaluation, we divided the dataset into three distinct sets (training, validation, and test), ensuring mutual exclusivity among them. The validation set played a crucial role in determining the error threshold (equal error rate) for evaluating both deepfake detection models and face recognition systems. Table 3 provides a comprehensive overview of this scaled-down dataset.

TABLE 3. Comprehensive overview of subset created from eKYC-DF dataset.

	Real	Fake	C0	C23	C40
Number of Videos	6,000	6,000			
Selection Condition	Video duration >= 2s				
Training Set Frames	216,000	208,140	✓	✓	✓
Val Set Frames	67,200	75,600	✓	✓	✓
Test Set Frames	76,800	75,720	✓	✓	✓

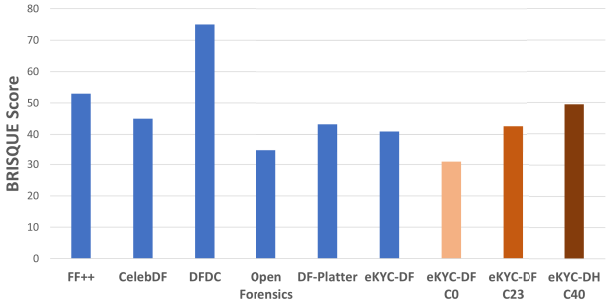


FIGURE 13. BRISQUE scores (lower scores indicate better visual quality). Blue bars show performance across the entire dataset. Other colors represent BRISQUE scores for three compression levels of the eKYC-DF dataset.

B. VISUAL QUALITY ASSESSMENT

The video frames in our dataset exhibited exceptional quality (see Figure 14), achieved through the utilization of high-quality sources and target datasets. To further enhance the realism of the generated frames, we used carefully selected face-swapping methods and applied postprocessing image enhancement techniques, as illustrated in Figure 15. This emphasis on generating images closely resembling real images makes our dataset particularly challenging for face recognition systems and deepfake detection models. To assess the visual quality of our dataset, we used the BRISQUE score [73], for which a lower value signifies better visual quality. Our dataset demonstrated competitive performance Figure 13, with a BRISQUE score of 30.83 on set C0, outperforming other existing datasets (FF++ [7], Celeb-DF [42], DFDC [44], OpenForensics [54], and DF-Platter [57]). The BRISQUE scores for sets C23 and C40 were 42.61 and 49.38, respectively. The overall average BRISQUE score for the entire dataset was 40.94, the second-highest score among the evaluation datasets. This high rank underscores the exceptional visual quality of our dataset (answer RQ1). The BRISQUE scores for the other datasets were approximated from [54], and the DF-Platter score was approximated from [57].

C. FACE RECOGNITION EVALUATION

To evaluate face recognition, we used three powerful facial recognition models: ArcFace [74], Facenet, and Facenet512 [75], along with pretrained versions of Deepface [1]. We first needed to determine the suitable threshold for each model when analyzing real video frames. We did this by extracting two frames from each real video, labeling them as "Matched", and two frames from different real videos labeled them as "Non-matched". This process was carried out for the three compression levels: C0, C23, and C40. The cosine similarities were normalized using:

normalized_similarity = (1 - cosine_similarity) / 2

By evaluating the false match rate (FMR) and false non-match rate (FNMR) of the models, we identified the optimal threshold for each. Then, we used this threshold to perform binary classification of the real set and the rest of the other sets.

We used another victim's face image of the victim himself in the face-swapping process for the fake frames, matching it with a frame from the resulting fake video (the output of face-swapping between the victim's face image and the attacker-driven video). For the other set, we paired the victim's face image with a different fake video generated from another attacker's video (different actor), labeling these "Non-matched". This enabled us to construct a comprehensive dataset for testing face recognition systems under the three compression levels. The results, summarized in Table 4 and Figure 16, demonstrate the models' impressive accuracy with real frames; on the other side, these same models were unable to distinguish whether a face in an image was real or swapped by one of the face-swapping methods, indicating the high quality of the deepfakes in our eKYC-DF dataset. Figure 1 illustrates a critical vulnerability of face recognition systems to high-quality deepfakes. In this experiment, we used a face recognition system to measure the similarity between a fake victim's face image generated through deepfakes and a real victim's face image. The system's inability to distinguish between the two images demonstrates the potential for deepfakes to be misused for deception and highlights the need for further research into improving the robustness of face recognition systems against such attacks. We obtained the same results in an additional experiment that measured the similarity between the victim's face image and a fake video. Following the same data construction process as previously described, we created a new dataset for evaluating facial recognition systems on image-video similarity. This dataset consisted of another victim's face image of the victim himself and ten frames extracted from the resulting video of face swapping between an image of the victim's face and the attacker-driven video, as shown in Figure 17.

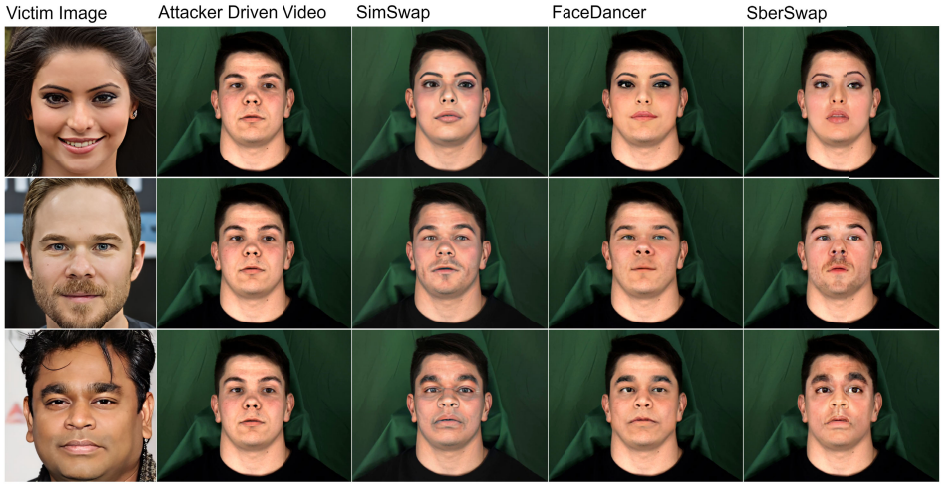


FIGURE 14. Face-swapping results generated by SimSwap, FaceDancer, and SberSwap.

FIGURE 15. Example images showing the importance of enhancement in reducing artifacts and improving visual quality. Results of three enhancement algorithms compared with original results. BRISQUE scores (shown from top to bottom for each column) were improved when enhancement was performed. Faces in images were swapped using FaceDancer.

We can, therefore, conclude that, without deepfake detection, eKYC systems can be easily compromised (answer RQ2). This presents a significant challenge, as face recognition systems need to be retrained and adjusted to keep up with advancements in deepfake generation techniques. The realistic and high-quality nature of the deepfakes in our dataset underscores the necessity for continuous development and enhancement of face recognition systems to combat the proliferation of deceptive and sophisticated deepfakes.

TABLE 4. Results obtained from evaluating our dataset on face recognition models (Acc in %; AUC in %; Thr = Threshold).

Dataset	Face-Swapping Method	Face Recognition Model: Acc % – AUC %		
		ArcFace (Thr = 16%)	Facenet (Thr = 25%)	Facenet512 (Thr = 27%)
Real	/	92.00 – 96.00	91.00 – 97.00	95.00 – 99.00
C0	SimSwap	92.00 – 97.00	96.00 – 98.00	98.00 – 99.00
	SberSwap	92.00 – 96.00	94.00 – 98.00	97.00 – 100.00
	FaceDancer	94.00 – 98.00	97.00 – 99.00	98.00 – 100.00
C23	SimSwap	92.00 – 97.00	96.00 – 98.00	98.00 – 99.00
	SberSwap	93.00 – 97.00	95.00 – 98.00	98.00 – 99.00
	FaceDancer	94.00 – 97.00	96.00 – 98.00	98.00 – 100.00
C40	SimSwap	92.00 – 97.00	95.00 – 98.00	98.00 – 100.00
	SberSwap	93.00 – 97.00	95.00 – 98.00	97.00 – 99.00
	FaceDancer	94.00 – 97.00	96.00 – 98.00	98.00 – 100.00
C0	All	93.00 – 97.00	95.00 – 98.00	98.00 – 100.00
C23	All	93.00 – 97.00	96.00 – 98.00	98.00 – 100.00
C40	All	93.00 – 97.00	95.00 – 98.00	98.00 – 99.00

TABLE 5. Results for deepfake detectors trained and tested on eKYC-DF small version dataset (Accuracy in % and AUC in %). All detectors were trained using a training set comprising samples from C0, C23, and C40.

Test Set	Models (Acc % – AUC %)		
	XceptionNet	EfficientNet (B4)	EfficientNet-v2 (B4)
C0, C23, C40	99.95 – 100.00	99.93 – 100.00	99.95 – 100.00
C0	99.93 – 100.00	99.89 – 100.00	99.93 – 100.00
C23	99.93 – 100.00	99.89 – 100.00	99.97 – 100.00
C40	99.93 – 100.00	99.89 – 100.00	99.95 – 100.00

D. DEEPFAKE DETECTION EVALUATION

To evaluate deepfake detection, we first assessed the performance of deepfake detection models specifically trained on our dataset and then investigated the generalization capabilities of deepfake detection models pretrained on external datasets when applied to ours. We also ascertained whether our dataset presents new challenges not encountered in prior datasets. By addressing these key aspects, we gained valuable insights into the suitability of our dataset for training robust deepfake detection models and its potential contributions to advancing the field.

We used three advanced deepfake detection models: XceptionNet [7], EfficientNet (B4) [76], and EfficientNet-

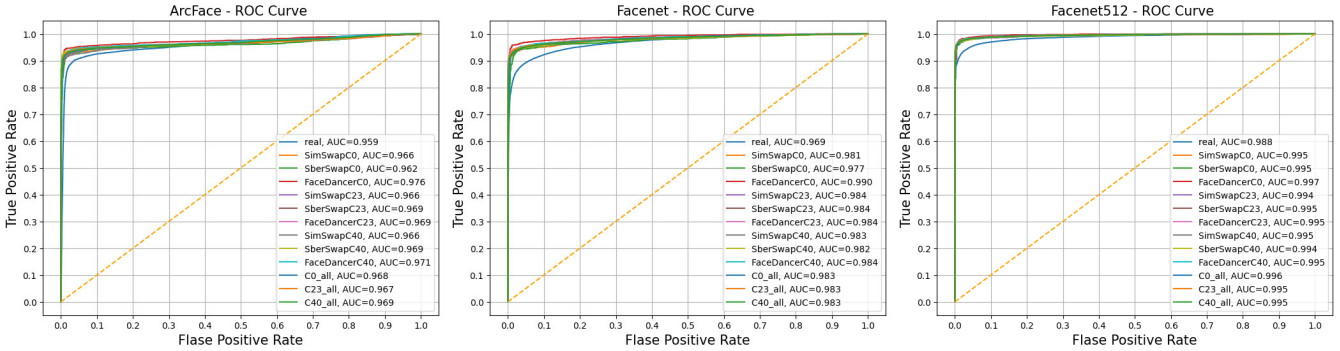


FIGURE 16. ROC curves of different face recognition models (ArcFace, Facenet, and Facenet512) on our dataset.

TABLE 6. Results of deepfake detector inference on various datasets (zero-shot inference): We report the AUC (%) on Celeb-DF-v2 (CDF) [42], DeepFake Detection (DFD) [41], and DeeperForensics (DFo) [43], as well as AUC and Acc (%) on our dataset eKYC-DF.

Model	Trained on	CDF	DFD	DFo	eKYC-DF (Acc % – AUC %, Threshold = 0.5)		
					C0	C23	C40
XceptionNet [7]	FF++ [7]	73.70	-	84.50	45.89 - 40.33	42.36 - 36.87	40.87 - 33.89
CNN-generated [78]	FF++ [7]	75.60	-	74.40	50.00 - 48.20	48.70 - 47.83	32.84 - 32.50
Self-B-Img [79]	FF++ [7]	93.18	97.87	-	51.25 – 50.72	55.17 – 55.66	49.07 – 49.07
Self-B-Img [79]	FF++ c23 [7]	92.87	98.16	-	57.25 – 57.25	63.12 – 62.56	49.34 – 49.34
AltFreezing [80]	FF++ [7]	89.50	98.50	99.30	48.88 – 48.88	50.15 – 50.15	51.05 – 51.05
HiFi-IFDL [81]	HiFi-IFDL [81]	-	-	-	45.96 – 50.00	49.83 – 50.00	49.83 – 50.00

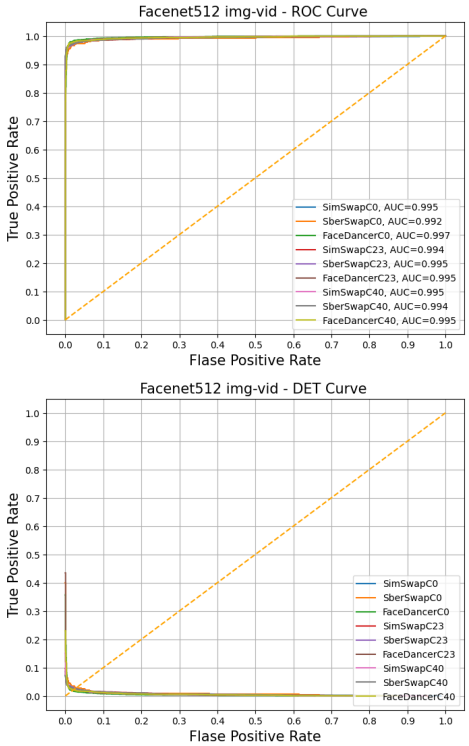


FIGURE 17. ROC and DET curves of the Facenet512 face recognition model on our dataset for image-video similarity.

v2 (B4) [77]. As shown in Table 5, all three models achieved high accuracy when trained and tested on the eKYC-DF small version dataset. This means that the eKYC-DF dataset is suitable for training deepfake detection models. However, these detectors failed to demonstrate generalization when we tested them on other datasets, such as the FF++ dataset. This lack of generalization could be attributed to model limitations. Furthermore, certain models that demonstrated strong generalization abilities on external datasets encountered difficulties when tested on the eKYC-DF dataset.

Table 6 provides a comprehensive overview of these results, illustrating instances where the models exhibited high generalizability for other datasets but failed to perform effectively on the eKYC-DF dataset (answer RQ3). Deepfake detection models primarily address artifacts that emerge during generation, which naturally differ depending on the synthesis technique. To create an ideal deepfake detection dataset, it is crucial to use a diverse set of deepfake methods and a wide array of real videos. Since no available deepfake dataset has achieved the desired level of generality, combining multiple datasets to achieve the desired generality is a practical approach.

V. CONCLUSION

Our large-scale eKYC-DF dataset is a valuable resource for researchers working to develop and protect eKYC systems as well as deepfake detection and facial recognition systems. The dataset includes diverse videos created using three deepfake methods and a range of real videos for comparison, totaling 228,760 videos. Using this dataset, researchers can develop and evaluate deep learning models that can accurately identify deepfake videos and improve

the reliability of eKYC systems. The results of experiments using this dataset demonstrated that it is a valuable resource for advancing state-of-the-art deepfake detection models and face recognition systems.

REFERENCES

[1] S. I. Serengil and A. Ozpinar, "HyperExtended LightFace: A facial attribute analysis framework," in *Proc. Int. Conf. Eng. Emerg. Technol. (ICEET)*, Oct. 2021, pp. 1–4, doi: 10.1109/ICEET53442.2021.9659697.

[2] K. Carta, C. Barral, N. El Mrabet, and S. Mouille, "Video injection attacks on remote digital identity verification solution using face recognition," in *Proc. 13th Int. Multi-Conference Complex., Informat. Cybern. (IMCIC)*, Mar. 2022, pp. 92–97.

[3] A. Nanda, S. W. A. Shah, J. J. Jeong, R. Doss, and J. Webb, "Towards higher levels of assurance in remote identity proofing," *IEEE Consum. Electron. Mag.*, vol. 13, no. 1, pp. 1–8, Jan. 2023.

[4] T.-L. Do, M.-K. Tran, H. H. Nguyen, and M.-T. Tran, "Potential attacks of DeepFake on eKYC systems and remedy for eKYC with DeepFake detection using two-stream network of facial appearance and motion features," *Social Netw. Comput. Sci.*, vol. 3, no. 6, pp. 1–17, Sep. 2022.

[5] B. Scollan, M. Shere, and R. T. Brink, "Perspectives on new forms of remote identity proofing and authentication for IRS online services," *IRS Res. Bull.*, p. 116, Jun. 2020.

[6] D. Dagar and D. K. Vishwakarma, "A literature review and perspectives in deepfakes: Generation, detection, and applications," *Int. J. Multimedia Inf. Retr.*, vol. 11, no. 3, pp. 219–289, Sep. 2022.

[7] A. Rössler, 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. (ICCV)*, Oct. 2019, pp. 1–11.

[8] Z. Akhtar, "Deepfakes generation and detection: A short survey," *J. Imag.*, vol. 9, no. 1, p. 18, Jan. 2023.

[9] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 10674–10685.

[10] J. Ren, C. Xu, H. Chen, X. Qin, C. Li, and L. Zhu, "Towards flexible, scalable, and adaptive multi-modal conditioned face synthesis," 2023, *arXiv:2312.16274*.

[11] R. Chen, X. Chen, B. Ni, and Y. Ge, "SimSwap: An efficient framework for high fidelity face swapping," in *Proc. 28th ACM Int. Conf. Multimedia*, Oct. 2020, pp. 2003–2011.

[12] F. Rosberg, E. E. Aksoy, F. Alonso-Fernandez, and C. Englund, "FaceDancer: Pose- and occlusion-aware high fidelity face swapping," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2023, pp. 3443–3452.

[13] A. Groshev, A. Maltseva, D. Chesakov, A. Kuznetsov, and D. Dimitrov, "GHOST—A new face swap approach for image and video domains," *IEEE Access*, vol. 10, pp. 83452–83462, 2022.

[14] Y. Liu, Q. Li, Q. Deng, Z. Sun, and M.-H. Yang, "GAN-based facial attribute manipulation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 12, pp. 14590–14610, Dec. 2023.

[15] A. Nickabadi, M. Saeedi Fard, N. Moradzadeh Farid, and N. Mohammadbagheri, "A comprehensive survey on semantic facial attribute editing using generative adversarial networks," 2022, *arXiv:2205.10587*.

[16] I. Perov, D. Gao, N. Chervoniy, K. Liu, S. Marangonda, C. Umé, M. Dpfks, C. S. Facenheim, L. Rp, J. Jiang, S. Zhang, P. Wu, B. Zhou, and W. Zhang, "DeepFaceLab: Integrated, flexible and extensible face-swapping framework," 2020, *arXiv:2005.05535*.

[17] 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/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8789–8797.

[18] M. Liu, Y. Ding, M. Xia, X. Liu, E. Ding, W. Zuo, and S. Wen, "STGAN: A unified selective transfer network for arbitrary image attribute editing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3668–3677.

[19] T. Brooks, A. Holynski, and A. A. Efros, "InstructPix2Pix: Learning to follow image editing instructions," 2022, *arXiv:2211.09800*.

[20] P. Kumar, M. Vatsa, and R. Singh, "Detecting Face2Face facial reenactment in videos," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2020, pp. 2578–2586.

[21] N. Drobyshev, J. Chelishev, T. Khakhulin, A. Ivakhnenko, V. Lempitsky, and E. Zakharov, "MegaPortraits: One-shot megapixel neural head avatars," 2022, *arXiv:2207.07621*.

[22] B. Zhang, C. Qi, P. Zhang, B. Zhang, H. Wu, D. Chen, Q. Chen, Y. Wang, and F. Wen, "MetaPortrait: Identity-preserving talking head generation with fast personalized adaptation," 2022, *arXiv:2212.08062*.

[23] Z. Xu, J. Zhang, J. Hao Liew, W. Zhang, S. Bai, J. Feng, and M. Zheng Shou, "PV3D: A 3D generative model for portrait video generation," 2022, *arXiv:2212.06384*.

[24] T. T. Nguyen, Q. V. H. Nguyen, D. T. Nguyen, D. T. Nguyen, T. Huynh-The, S. Nahavandi, T. T. Nguyen, Q.-V. Pham, and C. M. Nguyen, "Deep learning for deepfakes creation and detection: A survey," *Comput. Vis. Image Understand.*, vol. 223, Oct. 2022, Art. no. 103525.

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

[26] N. Yu, L. Davis, and M. Fritz, "Attributing fake images to GANs: Learning and analyzing GAN fingerprints," 2018, *arXiv:1811.08180*.

[27] J. Pu, N. Mangaokar, B. Wang, C. K. Reddy, and B. Viswanath, "NoiseScope: Detecting deepfake images in a blind setting," in *Proc. Annu. Comput. Secur. Appl. Conf.*, 2020, pp. 913–927.

[28] L. Guarnera, O. Giudice, and S. Battiato, "DeepFake detection by analyzing convolutional traces," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 2841–2850.

[29] E. Sabir, J. Cheng, A. Jaiswal, W. AbdAlmageed, I. Masi, and P. Natarajan, "Recurrent convolutional strategies for face manipulation detection in videos," *Interfaces (GUI)*, vol. 3, no. 1, pp. 80–87, 2019.

[30] Y. Li and S. Lyu, "Exposing DeepFake videos by detecting face warping artifacts," 2018, *arXiv:1811.00656*.

[31] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, and J. Ortega-Garcia, "Deepfakes and beyond: A survey of face manipulation and fake detection," *Inf. Fusion*, vol. 64, pp. 131–148, Dec. 2020.

[32] P. Prajapati and C. Pollett, "MRI-GAN: A generalized approach to detect DeepFakes using perceptual image assessment," 2022, *arXiv:2203.00108*.

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

[34] N. Bonettini, E. D. Cannas, S. Mandelli, L. Bondi, P. Bestagini, and S. Tubaro, "Video face manipulation detection through ensemble of CNNs," in *Proc. 25th Int. Conf. Pattern Recognit. (ICPR)*, Jan. 2021, pp. 5012–5019.

[35] C. Miao, Q. Chu, W. Li, T. Gong, W. Zhuang, and N. Yu, "Towards generalizable and robust face manipulation detection via bag-of-local-feature," 2021, *arXiv:2103.07915*.

[36] J. Wang, Z. Wu, W. Ouyang, X. Han, J. Chen, Y.-G. Jiang, and S.-N. Li, "M2TR: Multi-modal multi-scale transformers for DeepFake detection," in *Proc. Int. Conf. Multimedia Retr.*, 2022, pp. 615–623.

[37] J. Y. Bok, K. H. Suh, and E. C. Lee, "Verifying the effectiveness of new face spoofing DB with capture angle and distance," *Electronics*, vol. 9, no. 4, p. 661, Apr. 2020.

[38] Z. Ming, M. Visani, M. M. Luqman, and J.-C. Burie, "A survey on anti-spoofing methods for facial recognition with RGB cameras of generic consumer devices," *J. Imag.*, vol. 6, no. 12, p. 139, Dec. 2020.

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

[40] P. Korshunov and S. Marcel, "DeepFakes: A new threat to face recognition? Assessment and detection," 2018, *arXiv:1812.08685*.

[41] N. Dufour, A. Gully, P. Karlsson, A. V. Vorbyov, T. Leung, J. Childs, and C. Bregler, "DeepFakes detection dataset by Google & Jigsaw," Google, USA, 2019. [Online]. Available: https://blog.research.google/2019/09/contributing-data-to-deepfake-detection.html?m=1

[42] Y. Li, X. Yang, P. Sun, H. Qi, and S. C.-D. Lyu, "A large-scale challenging dataset for deepfake forensics," 2019, *arxiv:1909.12962*.

[43] L. Jiang, R. Li, W. Wu, C. Qian, and C. C. Loy, "DeeperForensics-1.0: A large-scale dataset for real-world face forgery detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 2886–2895.

[44] B. Dolhansky, J. Bitton, B. Pflaum, J. Lu, R. Howes, M. Wang, and C. Canton Ferrer, "The DeepFake detection challenge (DFDC) dataset," 2020, *arXiv:2006.07397*.

[45] B. Zi, M. Chang, J. Chen, X. Ma, and Y.-G. Jiang, "WildDeepfake: A challenging real-world dataset for deepfake detection," in *Proc. 28th ACM Int. Conf. Multimedia*, Oct. 2020, pp. 2382–2390.

[46] P. Kwon, J. You, G. Nam, S. Park, and G. Chae, "KoDF: A large-scale Korean DeepFake detection dataset," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, Jun. 2021, pp. 10744–10753.

[47] Y. He, B. Gan, S. Chen, Y. Zhou, G. Yin, L. Song, L. Sheng, J. Shao, and Z. Liu, "ForgeryNet: A versatile benchmark for comprehensive forgery analysis," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 4358–4367.

[48] J. Huang, X. Wang, B. Du, P. Du, and C. Xu, "DeepFake MNIST+: A DeepFake facial animation dataset," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2021, pp. 1973–1982.

[49] A. Nagrani, J. S. Chung, W. Xie, and A. Zisserman, "Voxceleb: Large-scale speaker verification in the wild," *Comput. Speech Lang.*, vol. 60, Mar. 2020, Art. no. 101027.

[50] A. Fischer, J. Schalk, S. Hawk, and B. Doosje, "The Amsterdam dynamic facial expression set (ADFES)," *Emotion*, vol. 11, no. 4, p. 907, 2011.

[51] A. Siarohin, S. Lathuilière, S. Tulyakov, E. Ricci, and N. Sebe, "First order motion model for image animation," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 7137–7147.

[52] H. Khalid, S. Tariq, M. Kim, and S. S. Woo, "FakeAVCeleb: A novel audio-video multimodal deepfake dataset," 2021, *arXiv:2108.05080*.

[53] J. Son Chung, A. Nagrani, and A. Zisserman, "VoxCeleb2: Deep speaker recognition," 2018, *arXiv:1806.05622*.

[54] T.-N. Le, H. H. Nguyen, J. Yamagishi, and I. Echizen, "OpenForensics: Large-scale challenging dataset for multi-face forgery detection and segmentation in-the-wild," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, Oct. 2021, pp. 10117–10127.

[55] K. Narayan, H. Agarwal, K. Thakral, S. Mittal, M. Vatsa, and R. Singh, "DeePhy: On deepfake phylogeny," in *Proc. IEEE Int. Joint Conf. Biometrics (IJCB)*, Oct. 2022, pp. 1–10.

[56] Z. Cai, S. Ghosh, A. Dhall, T. Gedeon, K. Stefanov, and M. Hayat, "Glitch in the matrix: A large scale benchmark for content driven audio-visual forgery detection and localization," 2023, *arXiv:2305.01979*.

[57] K. Narayan, H. Agarwal, K. Thakral, S. Mittal, M. Vatsa, and R. Singh, "DF-platter: Multi-face heterogeneous deepfake dataset," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 9739–9748.

[58] P. Korshunov, H. Chen, P. N. Garner, and S. Marcel, "Vulnerability of automatic identity recognition to audio-visual deepfakes," in *Proc. IEEE Int. Joint Conf. Biometrics (IJCB)*, Jul. 2023, pp. 1–10.

[59] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, "VGGFace2: A dataset for recognising faces across pose and age," in *Proc. 13th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, May 2018, pp. 67–74.

[60] X. Wang, Y. Li, H. Zhang, and Y. Shan, "Towards real-world blind face restoration with generative facial prior," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 9164–9174.

[61] J. Guo, J. Deng, A. Lattas, and S. Zafeiriou, "Sample and computation redistribution for efficient face detection," 2021, *arXiv:2105.04714*.

[62] (Feb. 2021). *Faceswap is the Leading Free and Open Source Multi-platform Deepfakes Software*. [Online]. Available: https://faceswap.dev/

[63] Deepfakes. *Deepfakes/Faceswap: Deepfakes Software for All*. Accessed: Nov. 2022. [Online]. Available: https://github.com/deepfakes/faceswap

[64] L. Li, J. Bao, H. Yang, D. Chen, and F. Wen, "Advancing high fidelity identity swapping for forgery detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 5073–5082.

[65] Y. Zhu, Q. Li, J. Wang, C. Xu, and Z. Sun, "One shot face swapping on megapixels," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 4832–4842.

[66] Z. Xu, X. Yu, Z. Hong, Z. Zhu, J. Han, J. Liu, E. Ding, and X. Bai, "FaceController: Controllable attribute editing for face in the wild," in *Proc. AAAI Conf. Artif. Intell.*, vol. 35, no. 4, 2021, pp. 3083–3091.

[67] Y. Wang, X. Chen, J. Zhu, W. Chu, Y. Tai, C. Wang, J. Li, Y. Wu, F. Huang, and R. Ji, "HifiFace: 3D shape and semantic prior guided high fidelity face swapping," 2021, *arXiv:2106.09965*.

[68] D. Chesakov, A. Maltseva, A. Groshev, A. Kuznetsov, and D. Dimitrov, "A new face swap method for image and video domains: A technical report," 2022, *arXiv:2202.03046*.

[69] L. Li, J. Bao, H. Yang, D. Chen, and F. Wen, "FaceShifter: Towards high fidelity and occlusion aware face swapping," 2019, *arXiv:1912.13457*.

[70] Z. Tu, H. Talebi, H. Zhang, F. Yang, P. Milanfar, A. Bovik, and Y. Li, "MAXIM: Multi-axis MLP for image processing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 5759–5770.

[71] Z. Yue and C. Change Loy, "DifFace: Blind face restoration with diffused error contraction," 2022, *arXiv:2212.06512*.

[72] C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, F. Zhang, C.-L. Chang, M. Guang Yong, J. Lee, W.-T. Chang, W. Hua, M. Georg, and M. Grundmann, "MediaPipe: A framework for building perception pipelines," 2019, *arXiv:1906.08172*.

[73] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.

[74] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "ArcFace: Additive angular margin loss for deep face recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 4685–4694.

[75] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 815–823.

[76] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 6105–6114.

[77] M. Tan and Q. V. Le, "EfficientNetv2: Smaller models and faster training," in *Proc. Int. Conf. Mach. Learn.*, 2021, pp. 10096–10106.

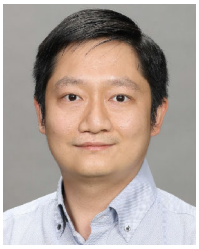
[78] S.-Y. Wang, O. Wang, R. Zhang, A. Owens, and A. A. Efros, "CNN-generated images are surprisingly easy to spot…for now," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 8692–8701.

[79] K. Shiohara and T. Yamasaki, "Detecting deepfakes with self-blended images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 18699–18708.

[80] Z. Wang, J. Bao, W. Zhou, W. Wang, and H. Li, "AltFreezing for more general video face forgery detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 4129–4138.

[81] X. Guo, X. Liu, Z. Ren, S. Grosz, I. Masi, and X. Liu, "Hierarchical fine-grained image forgery detection and localization," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 3155–3165.

HICHEM FELOUAT received the B.S. and M.S. degrees in computer science from the University of Jijel, Algeria, in 2013 and 2015, respectively, and the first Ph.D. degree in computer science from the University of Blida, Algeria, in 2021. He is currently pursuing the second Ph.D. degree with the Echizen Laboratory, National Institute of Informatics, Tokyo, Japan. His research interests include security and privacy in biometrics, neuroimaging, and machine learning.



HUY H. NGUYEN (Member, IEEE) received the Ph.D. degree in computer science from The Graduate University for Advanced Studies, SOKENDAI, Japan, in 2022. He is currently a Project Assistant Professor with the Echizen Laboratory, National Institute of Informatics, Tokyo, Japan. His research interests include security and privacy in biometrics and machine learning.

TRUNG-NGHIA LE received the B.Sc. degree (Hons.) in computer science from the University of Science, in 2012, the M.Sc. degree in computer science from the John Von Neumann Institute, in 2014, and the Ph.D. degree in computer science from the National Institute of Informatics, Japan, in 2018. He was a Project Assistant Professor with the Echizen Laboratory, National Institute of Informatics. He is currently a Senior Researcher and a Lecturer with the University of Science, VNU-HCM, Vietnam. His research interests include computer vision and applied deep learning.

JUNICHI YAMAGISHI (Senior Member, IEEE) received the Ph.D. degree from Tokyo Institute of Technology (Tokyo Tech), Tokyo, Japan, in 2006. From 2007 to 2013, he was a Research Fellow with the Centre for Speech Technology Research, University of Edinburgh, U.K. He became an Associate Professor with the National Institute of Informatics, Japan, in 2013, where he is currently a Professor. He is a Principal Investigator on the JST-CREST and ANR-supported VoicePersonae Project. His research interests include speech processing, machine learning, signal processing, biometrics, digital media cloning, and media forensics. He served as a member of the IEEE Speech and Language Technical Committee, from 2013 to 2019, as an Associate Editor for IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, from 2014 to 2017, and the Chairperson for ISCA SynSIG, from 2017 to 2021. He served as a co-organizer for the bi-annual ASVspoof Challenge and the bi-annual Voice Conversion Challenge. He is a Senior Area Editor of IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING.

ISAO ECHIZEN (Senior Member, IEEE) received the B.S., M.S., and D.E. degrees from Tokyo Institute of Technology, Japan, in 1995, 1997, and 2003, respectively. In 1997, he joined Hitachi Ltd., where he was a Research Engineer with the Systems Development Laboratory, until 2007. He was a Visiting Professor with the University of Freiburg, Germany, and the University of Halle-Wittenberg, Germany. He is currently the Director and a Professor with the Information and Society Research Division, National Institute of Informatics (NII), the Director of the Global Research Center for Synthetic Media, NII, a Professor with the Department of Information and Communication Engineering, Graduate School of Information Science and Technology, The University of Tokyo, and a Professor of the Graduate Institute for Advanced Studies, The Graduate University for Advanced Studies, SOKENDAI, Japan. He is engaged in research on multimedia security and multimedia forensics. He is also the Research Director of the CREST FakeMedia Project, Japan Science and Technology Agency (JST). He was a member of the Information Forensics and Security Technical Committee of the IEEE Signal Processing Society. He received the Best Paper Award from the IEICE, in 2023, the Best Paper Awards from the IPSJ, in 2005 and 2014, the IPSJ Nagao Special Researcher Award, in 2011, the DOCOMO Mobile Science Award, in 2014, the Information Security Cultural Award, in 2016, and the IEEE Workshop on Information Forensics and Security Best Paper Award, in 2017. He is an IEICE Fellow, the Japanese Representative on IFIP TC11 (Security and Privacy Protection in Information Processing Systems), a Member-at-Large of the Board-of-Governors of APSIPA, and an Editorial Board Member of the IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING, the *EURASIP Journal on Image and Video Processing*, and the *Journal of Information Security and Applications* (Elsevier).