IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 23 April 2024, accepted 3 May 2024, date of publication 9 May 2024, date of current version 20 May 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3399230

RESEARCH ARTICLE

Verifiable Facial De-Identification in Video Surveillance

SUNGJUNE PARK, HYUNSIK NA, AND DAESEON CHOI, (Member, IEEE)
Department of Software, Soongsil University, Seoul 07027, South Korea

Corresponding author: Daeseon Choi (sunchoi@ssu.ac.kr)

This work was supported in part by the Institute of Information and Communications Technology Planning and Evaluation (IITP) Grant through Korean Government, Ministry of Science and ICT (MSIT) (Robust AI and Distributed Attack Detection for Edge AI Security) under Grant 2021-0-00511; and in part by IITP Grant funded by Korea Government (MSIT) (Development of User Identity Certification and Management Technology for Self-Sovereign Identity Applications) under Grant 2021-0-00565.

This work involved human subjects or animals in its research. The authors confirm that all human/animal subject research procedures and protocols are exempt from review board approval.

ABSTRACT With the advancement of facial recognition technology, concerns over facial privacy breaches owing to data leaks and external attacks have been escalating. Existing de-identification methods face challenges with compatibility with facial recognition models and difficulties in verifying de-identified images. To address these issues, this study introduces a novel framework that combines face verification-enabled de-identification techniques with face-swapping methods, tailored for video surveillance environments. This framework employs StyleGAN, Pixel2Style2Pixel (PSP), HopSkipJumpAttack (HSJA), and FaceNet512 to achieve face verification-capable de-identification, and uses the dlib library for face swapping. Experimental results demonstrate that this method maintains high face recognition performance (98.37%) across various facial recognition models while achieving effective de-identification. Additionally, human tests have validated its sufficient de-identification capabilities, and image quality assessments have shown its excellence across various metrics. Moreover, real-time de-identification feasibility was evaluated using Nvidia Jetson AGX Xavier, achieving a processing speed of up to 9.68 fps. These results mark a significant advancement in demonstrating the practicality of high-quality de-identification techniques and facial privacy protection in the field of video surveillance.

INDEX TERMS Face de-identification, face privacy, face verification, face verifiable de-identification, privacy protection, StyleGAN.

I. INTRODUCTION

Facial recognition technology, powered by advancements in artificial intelligence and big data, plays a crucial role in various fields, including security, personalized services, and constructing social safety networks [1], [2]. This technology is extensively used for personal identification and verification, leading to the collection of massive facial data. However, this data collection and usage pose serious concerns, such as privacy infringement. Particularly, privacy breaches due to data leaks or external attacks diminish the reliability of facial recognition technology, threatening user safety and rights.

To address these issues, conventional face de-identification techniques have been developed [3]. These techniques focus on obscuring identifiable features of individuals through methods like noise addition, image blurring, JPEG compression, and pixelation. However, these methods have limitations when combined with facial recognition technology and often fail to eliminate recognizability completely. Furthermore, de-identified video information can be easily identified as such, limiting its use as training data or for other secondary purposes. Moreover, these techniques are vulnerable to removal and restoration technologies like denoising [4], [5], [6] or inpainting [7], a limitation noted in related research.

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

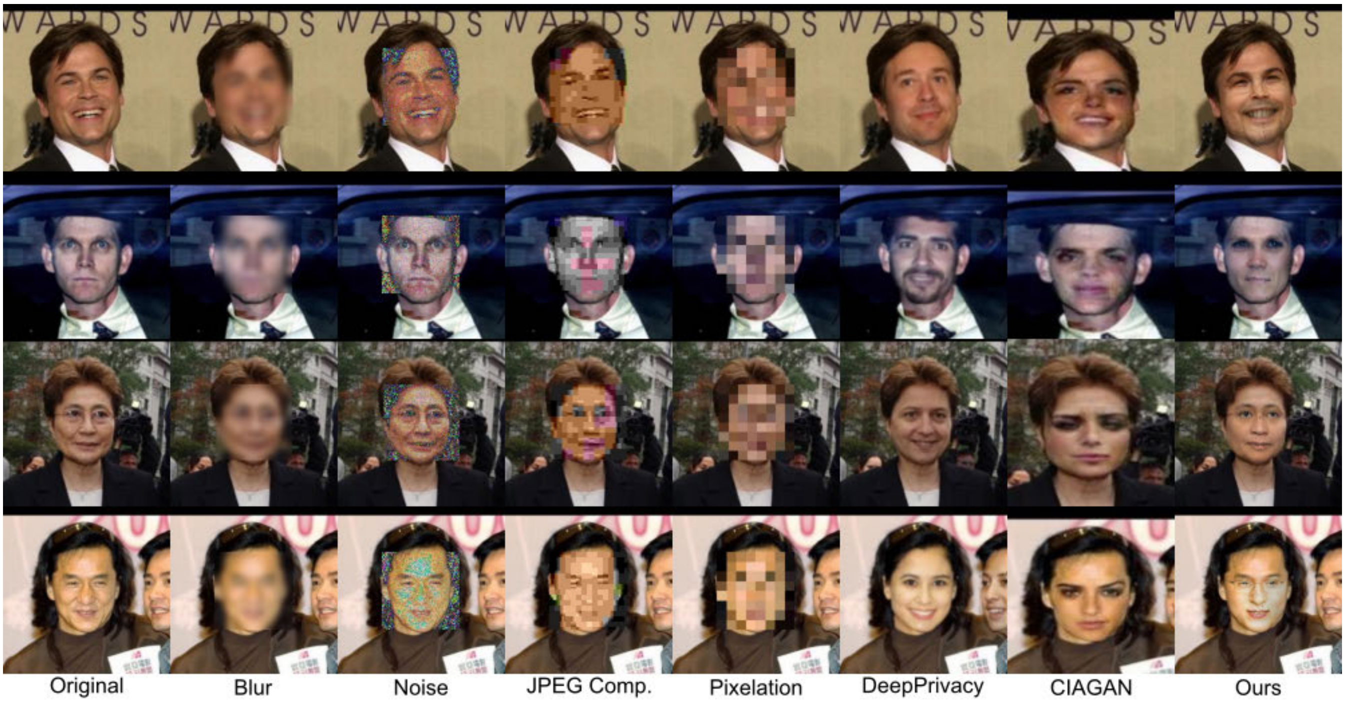


FIGURE 1. Example of de-identification using our proposed method compared to other de-identification techniques. Rows 2-5 illustrate conventional image manipulation methods for de-identification, whereas rows 6 and 7 demonstrate AI-based de-identification techniques.

These challenges suggest the need for new approaches in advancing face de-identification technology.

The evolution of deep learning technologies has opened new possibilities in this field [8], [9], [10], [11], [12], [13], [14], [15]. Deep learning-based facial de-identification methods are evolving with a focus on making faces recognizable to humans without anonymizing them or making them undetectable to facial recognition models. These approaches have the utility of protecting the privacy of individuals while maintaining the utility of facial recognition technology. Deep learning-based face de-identification methods have evolved to focus either on merely anonymizing facial features or on ensuring that, while humans can recognize the person, facial recognition models cannot. These approaches safeguard individual privacy while preserving the utility of facial recognition technologies. However, such de-identification techniques modify facial characteristics without considering the unique features and distinctiveness of individuals. Therefore, this approach impedes the possibility of re-identification and compromises the applicability of these modified facial images for various tasks, including enhanced facial recognition, emotional analysis, and other biometric-based applications, thereby undermining the broader utility and effectiveness of facial analysis technologies.

Furthermore, de-identification in real-time video environments is crucial. Videos show various changes such as facial expressions, head movements, and varying lighting conditions over time, which increase the complexity of applying facial recognition technology effectively in real-time scenarios. Accurately identifying and de-identifying an individual in environments with multiple people is significantly challenging. Effective de-identification in such environments must protect individual privacy while meeting the needs of real-time video processing. These challenges underscore the need for new approaches in the development of facial de-identification technology in video environments.

This study proposes a framework that elevates facial de-identification to a new dimension. This framework utilizes StyleGAN [16], Pixel2Style2Pixel (PSP) encoder [17], and HopSkipJumpAttack (HSJA) [18] to de-identify faces while enabling artificial intelligence(AI) models to uniquely identify individuals. It overcomes the limitations of existing facial de-identification techniques, offering an effective way to protect privacy and maintain the usefulness of facial recognition technology. By using this framework, we can enhance the reusability of facial data and open up various application possibilities for facial recognition technology, presenting a natural appearance, as shown in Fig. 1. This approach could significantly impact the future development of facial recognition technology and efforts to protect privacy.

We evaluated the framework in edge computing environments using lightweight devices, specifically focusing on real-time video processing with the Nvidia Jetson AGX Xavier [19] to demonstrate its effectiveness in real-world applications. This evaluation shows that facial de-identification technology is useful in digital evaluation and it can be applied in diverse real-world applications.

This study marks a significant advancement in the fields of data protection and privacy-preserving technologies, offering the following key contributions:

- Verifiable De-identification Techniques: We introduce verifiable de-identification techniques, a first in offering a balanced approach between privacy protection and data utility. This innovation facilitates compliance with privacy laws and enhances the value of data, pioneering new pathways in sensitive information management.
- Development of a Framework for Real-World Application: Our study results in a comprehensive framework designed for real-world implementation. It surpasses digital-level research by proving effective in real business and technical environments, highlighting its practical utility beyond theoretical or laboratory settings.
- Comprehensive Experimental Validation of Performance: Through extensive experiments, we validate the performance of our techniques in verification, de-identification, quality, and efficiency on edge devices. These tests confirm the theoretical and practical efficacy of our approaches in diverse scenarios, demonstrating their broad applicability and adaptability.

Through these contributions, our paper pioneers a new frontier in privacy-preserving technology, offering practical strategies for achieving a balance between data protection and utility. Furthermore, by providing new directions and challenges for future researchers, this work is poised to spur continuous innovation in the field. The remainder of this paper is organized as follows. First, the related works are presented. Thereafter, the proposed framework is presented. Third, the experimental settings and results are presented. Fourth, the discussions are presented, which are then followed by the conclusions and future works recommendations.

II. RELATED WORKS

A. IMAGE MANIPULATION TECHNIQUES FOR FACE DE-IDENTIFICATION

Conventional face de-identification techniques [20], [21], [22], [23] encompass various image manipulation strategies, including blurring, noise addition, JPEG compression, and pixelation. Blurring [20] is a straightforward yet effective method that obscures details of an image to hinder personal identification. This approach often employs various filters, such as Gaussian blur, to reduce distinct facial features. Although it is effective in preventing identification and retaining usability for specific tasks like object detection or demographic analysis, excessive blurring can diminish the utility of the original image.

Noise addition [21] involves introducing random pixel values into the original image, thereby impairing the accuracy of identification algorithms. A common method within this category is Gaussian Noise addition, which adds random pixel values following a Gaussian distribution to the original image. This technique subtly alters pixel values, blurring distinguishable facial patterns. The intensity of the noise can be adjusted to balance de-identification with the image's suitability for intended purposes. However, excessive noise can render the image inappropriate for certain applications. JPEG compression [22], primarily used for reducing image file size, inadvertently aids in de-identification by losing facial feature details during the compression process. Given its lossy compression nature, JPEG is particularly beneficial for large datasets where data storage and bandwidth are critical. The level of compression can be modulated to strike a balance between de-identification and image clarity. Pixelation [23] reduces image resolution by transforming details into larger pixel blocks. This technique effectively obscures facial features, offering a quick and simple application; however, the degree of pixelation necessary for effective de-identification can significantly reduce the image's utility for other purposes.

These conventional face de-identification techniques enable rapid and straightforward de-identification, effectively obscuring faces in tasks not focused on facial recognition, such as pedestrian detection, thereby protecting individual privacy [24]. However, their ease of application for anonymity and the consequent masking of facial features limits their applicability in facial recognition technology [25].

B. AI-BASED TECHNIQUES FOR FACE DE-IDENTIFICATION

The advancement of Generative Adversarial Networks (GAN) [26] has significantly impacted the field of face de-identification. The progression of GAN technology [27], [28], [29], known for its ability to generate highly realistic images, has played a crucial role in the evolution of face de-identification techniques. These technologies have become imperative tools for privacy protection, and AI-based approaches have marked significant developments in this domain. Recent research focus on striking a balance between privacy protection and the utility of data, introducing various methodologies.

Wu et al. [10] proposed Privacy-Protective-GAN (PP-GAN), a novel GAN framework that removes identifiable facial features while maintaining structural similarity. PP-GAN employs verifier and regulator modules to effectively eliminate identifiable information while preserving the utility of images. Liu et al. [8] proposed A3GAN, a method where face de-identification is approached as an attribute-based editing process. A3GAN removes identifiable information and injects controllable facial attributes, overcoming the lack of flexibility in existing methods and providing natural-looking de-identification results.

Gafni et al. [9] explored a method for de-identifying faces in real-time video streams. This approach uses an adversarial autoencoder network [30] architecture to minimize identifiable information while generating natural image sequences. Cai et al. [11] proposed the disguise algorithm, which removes identifiable information and substitutes it with a pseudo-identity while preserving utility attributes. This method uniquely maintains useful attributes while effectively eliminating identifiable information.

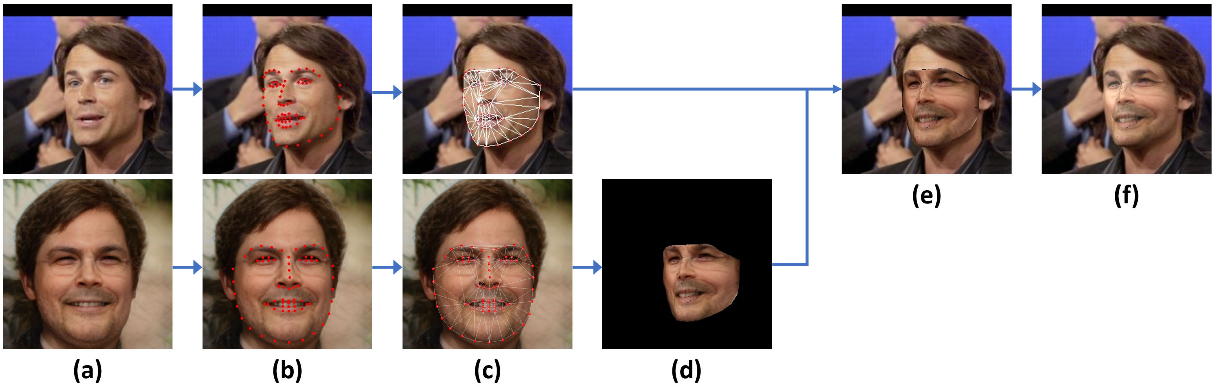


FIGURE 2. Stages of the face swapping method (FSM). The lower images are the source, whereas the upper images are the destination.

Du et al. [12] proposed the GARP-Face framework, which preserves gender, age, and race information while eliminating facial identifiers. This method leverages the active appearance model to generate de-identified faces. Chatzikyriakidis et al. [13] proposed the use of adversarial examples to minimize identifiable facial features with minimal distortion. Their P-FGVM method operates in the spatial domain of images to generate de-identified faces that maintain visual quality while being unidentifiable.

Hukkelås et al. [14] presented a method using GAN to de-identify faces in images while preserving the original data distribution. This study employs progressive growth learning techniques and a U-net architecture to create high-quality de-identified facial images. Finally, Maximov et al. [15] introduced a novel method, CIAGAN, for protecting data privacy in images and videos. CIAGAN removes identifiable characteristics while preserving essential features, producing high-quality images suitable for facial and body recognition.

These studies explore various aspects of face de-identification technologies, enhancing personal information protection and maintaining data utility while promoting technological advancements. However, these technologies diverge from the focus of our research, which explores the potential for verifiable de-identification. We aim to entirely remove the possibility of personal identification while still allowing for the re-identification of original identities under certain conditions.

III. VERIFIABLE FACIAL DE-IDENTIFICATION FRAMEWORK

A. StyleGAN-BASED FACE VERIFIABLE DE-IDENTIFICATION

1) StyleGAN

StyleGAN [16] represents an advanced GAN-based model [26], [27], [28], [29] capable of generating a diverse array of images. The architecture of this model has 18 layers, each utilizing a style vector W of size 1×512 to generate images. StyleGAN modulates features such as the shape of eyes, nose, mouth, and the overall facial structure through this vector.

The initial four layers predominantly focus on generating the basic shape of the face, whereas the subsequent four layers contribute to the precise positioning of the eyes, nose, and mouth. The final ten layers adjust detailed features like hair color and skin tone. Furthermore, the model incorporates random noise during the generation process to enhance the resolution and detail of the images.

2) Pixel2Style2Pixel

The PSP encoder [17] functions as an encoder for StyleGAN, extracting style vectors from input facial images for use in StyleGAN's generator. Unlike the conventional approach of simply replicating a 1×512 size vector, PSP independently extracts W^+ vectors of size 18×512 for each layer. This allows for more refined style adjustments, demonstrating superior performance compared to other encoders applicable to the original StyleGAN. The facial images generated using PSP closely resemble the original images, and manipulation of the W^+ vector enables the generation of facial images in desired shapes and forms as per user preference.

3) HopSkipJumpAttack

HSJA [18] is a method of decision-based attack, involving the modification of input images to query target classification models. HSJA induces misclassification by generating altered images that, while visually distinct, retain the original classification result. Unlike other attack methods that use gradient calculations via backpropagation, this approach employs random adjustments in the transformation process. HSJA is effective even in black-box situations where access to the internal parameters of the target model is not feasible. Compared to contemporary related attack techniques, HSJA has the advantage of closely approximating the target image with fewer input iterations.

B. FSM

1) YOLOv8

YOLOv8 [31], the latest iteration in the You Only Look Once (YOLO) series, represents a significant advancement in the field of real-time object detection. This model excels particularly in accurately detecting multiple and overlapping objects in complex environments while providing the essential rapid frame processing speed required for real-time

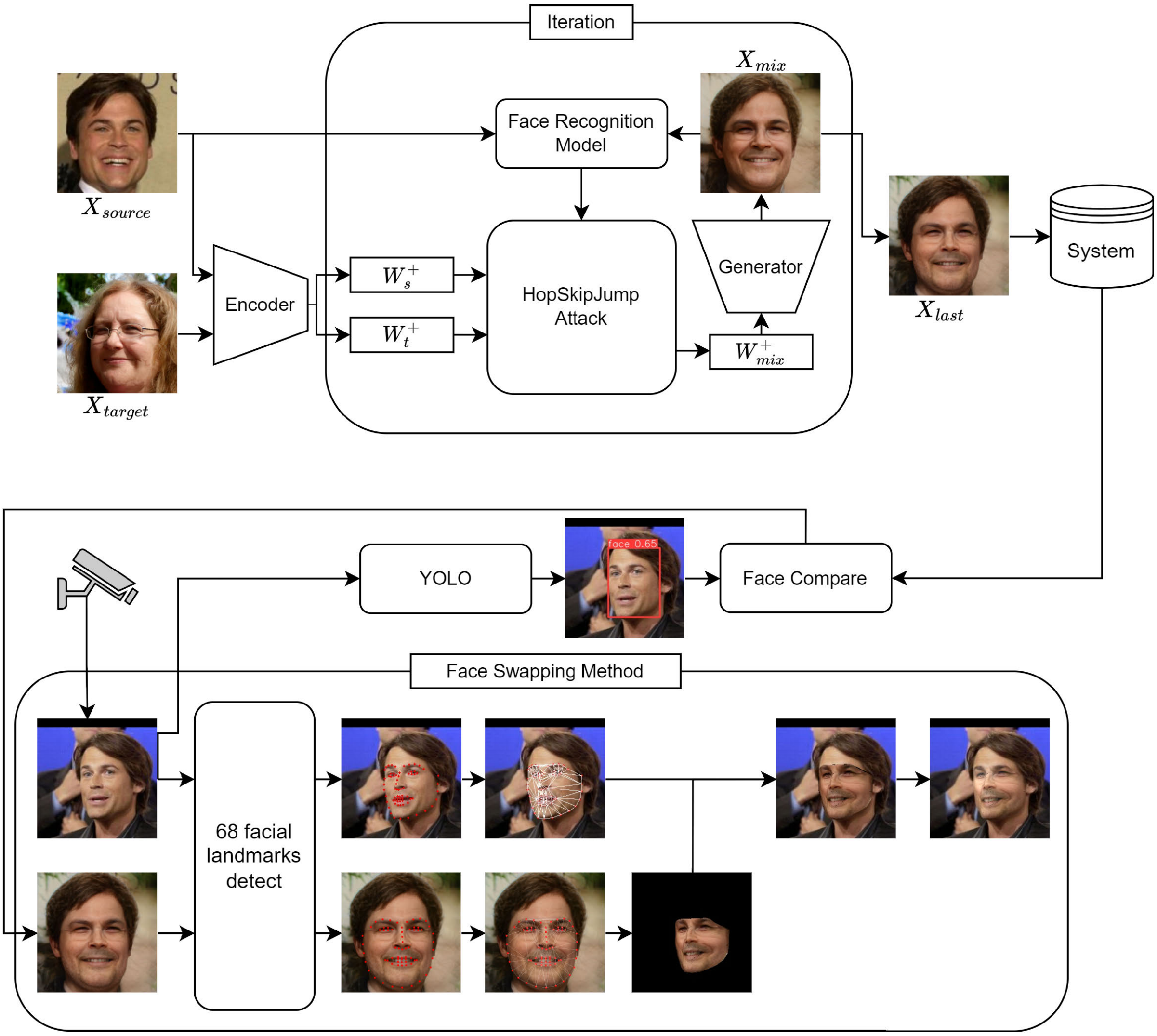


FIGURE 3. A diagram of our overall framework.

applications. Additionally, it demonstrates robustness in various lighting conditions, object sizes, and occlusions, thereby extending its applicability in real-world scenarios. The technological enhancements of YOLOv8 have a profound impact on diverse areas such as autonomous vehicles and public safety surveillance systems, offering the potential to significantly improve safety and efficiency in these domains.

2) FSM

The FSM, as shown in Fig. 2, employs the YOLOv8 for extracting bounding boxes and utilizes the dlib library [32] for detecting 68 facial landmarks. Additionally, to maintain the size and perspective of the face, the Delaunay triangulation method [33] is applied for triangulation. To adjust the color of the source image with that of the destination image, the seamlessClone function from the OpenCV library [34] is used.

The FSM process follows the following steps:

(a) Square bounding boxes for both face images are extracted using the YOLO.
(b) The dlib library is employed to detect 68 facial landmarks in both the source and destination images.
(c) Triangulation is performed on the 68 facial landmarks using the Delaunay triangulation method.
(d) Triangles are extracted and transformed from the source image to match the triangles in the destination image, ensuring they maintain the same size and perspective. Subsequently, these transformed triangles are linked together to reconstruct the face.
(e) The face in the destination image is replaced with the reconstructed face.
(f) Post-processing techniques, such as seamlessClone, are applied to adjust the color of the source image to match that of the destination image.

Through these steps, FSM effectively swaps faces between two images, maintaining the size and perspective of the faces while adjusting colors to achieve a natural-looking result.

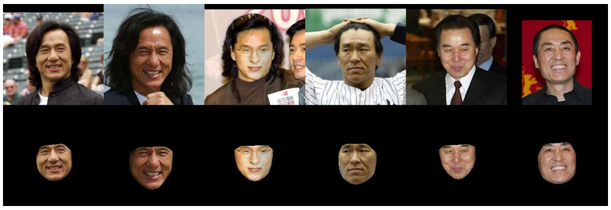


FIGURE 4. Example of human test questions. Row 1: unmasked question example. Row 2: masked question example. Column 1 shows the given examples; column 2 features the same individuals in a different frame; column 3 presents images on which our de-identification technology has been applied; columns 4-6 contain images deemed similar to the examples by FaceNet512.

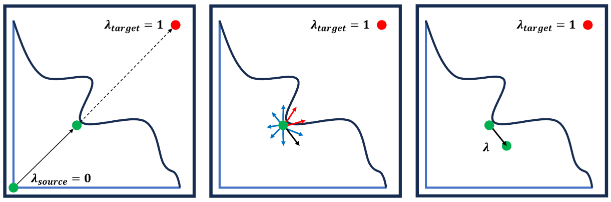


FIGURE 5. Illustration of the HSJA, detailing its three key stages. This figure aids in understanding the boundary search and image generation processes critical to our study.

C. OVERALL FRAMEWORK

This study proposes a framework for facial verifiable de-identification suitable for both video and real-time applications, as depicted in Fig. 3. The framework requires two types of images for de-identification: the source image of the individual to be de-identified, X_{source}, and a randomly generated facial image using StyleGAN, X_{target}. Through the PSP encoder, style vectors that contain the facial features of X_{source} and X_{target}, W_s^+ and W_t^+ respectively, are produced. These style vectors are aligned with the 18 layers of the StyleGAN's generator, each having dimensions of 18 × 512.

Before feeding the style vectors into the StyleGAN model, a de-identification adjustment variable, $\lambda \in [0, 1]$, is defined to ascertain the optimal level of de-identification. λ serves as the blending ratio between elements of each style vector, limited to a decimal value within the 0 to 1 range. The dimension of λ matches that of the style vectors, at 18 × 512. The equation for merging W_s^+ and W_t^+ using the established λ is as follows:

$$W_{mix}^+ = (1 - \lambda) \times W_s^+ + \lambda \times W_t^+ \tag{1}$$

As λ gets closer to 0, the StyleGAN model generates a facial image more resembling X_{source}, and as λ approaches 1, it produces an image more akin to X_{target}. Thus, by iteratively updating λ to achieve the optimal value, a facial image that is both sufficiently de-identified and capable of facial verification can be generated.

To optimize the de-identification adjustment variable λ for its intended purpose, the HSJA is utilized to challenge the facial verification model. λ starts with the initial value of $\lambda_{source} = [[0, 0, \ldots] \ldots]$, which can generate a facial image with the features of X_{source}, and aims for $\lambda_{target} = [[1, 1, \ldots] \ldots]$. The optimization process of λ, illustrated in Fig. 5, involves three repeating phases. The lower part of the Fig. 5, divided by a boundary, indicates the region where the facial verification model recognizes the input image as the user, whereas the upper part represents the class area of X_{target}.

The process begins with λ moving from λ_{source} towards λ_{target} via binary search. The basis of this search is whether the facial verification is successful for the produced X_{mix}, indicated as true or false. This step helps locate the boundary where the facial verification model recognizes X_{mix} as the user. Subsequently, λ undergoes several random perturbations, and the resulting X_{mix} is repeatedly queried against the facial verification model. Blue arrows represent scenarios where the perturbed λ and the consequent facial image X_{mix} are included in the verification area, whereas red arrows indicate scenarios where they are not verified.

Based on these query results, a directional vector towards the area capable of facial verification is formed. Finally, this directional vector is applied to adjust λ. These three steps are repeated several times, gradually bringing λ closer to the all-ones λ_{target}. The algorithm for λ optimization can be expressed as follows:

Algorithm 1 Parameter Update Algorithm

```
$N \leftarrow$ the predefined number of steps
for $i = 1$ to $N$ do
    $\lambda \leftarrow \frac{\lambda_{source}+\lambda_{target}}{2}$
    if $f(W_{mix}^+) \neq f(W_t^+)$ then
        $\lambda_{target} \leftarrow \lambda$
    else
        $\lambda_{source} \leftarrow \lambda$
    end if
end for
```

The last iteration includes a binary search to bring λ as near as possible to λ_{target}. This optimized method allows for the creation of the ultimate de-identified facial image, X_{last}, that still enables verification.

The facial image X_{last}, which is both optimally de-identified and verifiable, is subsequently stored in the System. In the facial verification process, X_{last} plays a vital role. When a camera captures an individual, the YOLO model is employed to extract the face from the image, which is then compared with the stored X_{last} images for verification. This verification process is performed by feeding the extracted image and the stored image into the face verification model, and then measuring the cosine distance of the resulting image to determine if the cosine distance is below a threshold. The equation for cosine distance ($\cos(\theta)$) is as follows:

$$\cos(\theta) = \frac{A \cdot B}{\|A\|_2 \, \|B\|_2} \tag{2}$$

If the person captured by the camera matches an image in the system, the FSM activates. The system replaces the real-time captured face with the X_{last} image, thereby safeguarding

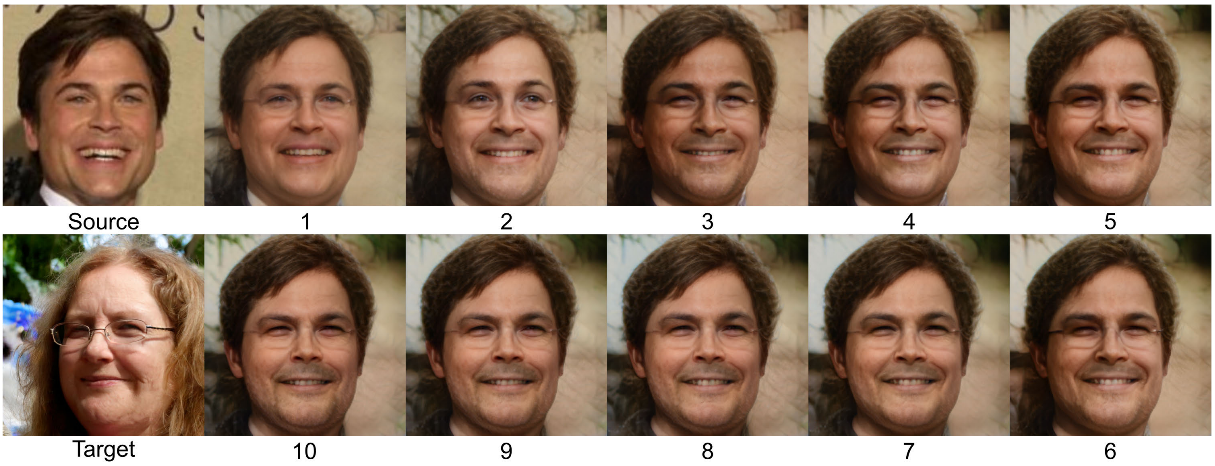


FIGURE 6. Image examples according to the iterations of our technique.

the privacy of the verificated user while simultaneously confirming their identity.

IV. EXPERIMENTAL SETTINGS AND RESULTS

A. EXPERIMENTAL SETTINGS

To conduct experiments in face verifiable de-identification, we utilized a StyleGAN pretrained on the FFHQ dataset [35], complemented by PSP as the encoder. In the target facial verification model, FaceNet512 [36] was utilized, and for the facial recognition model, the yolov8n model pre-trained on the Wider Face Dataset [37] was employed. Additionally, the number of iterations for the proposed method was set to 10 for the experiments.

To evaluate our experiments, we utilized the Labeled Faces in the Wild(LFW) dataset [38], [39]. Additionally, a webcam connected to a Jetson AGX Xavier was also employed. This setup was crucial to assess the effectiveness of the technology in real-world scenarios. To benchmark our proposed de-identification method, we applied conventional de-identification techniques such as blurring, noise addition, JPEG compression(JPEG Comp.), and pixelation to the LFW dataset. Gaussian Blur was applied for blurring, with the kernel size set to 15 by 15. Gaussian noise was used for noise addition, maintaining a mean of 0 and setting the standard deviation $\sigma = 0.07$. For JPEG compression, the JPEG quality was set to 1%; for pixelation, the pixel size was set to 8 by 8 for the experiments.

In this study, various quantitative metrics such as L2 distance, Frechet Inception Distance (FID) [40], Structural Similarity Index Measure (SSIM) [41], and Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) [42] were employed to assess the quality of de-identified images. The L2 distance is utilized for measuring the similarity between two images. This method considers images as multi-dimensional vectors, using the intensity or color values of each pixel as elements of the vector, to calculate the distance between the two images. A smaller value indicates greater similarity between the images. FID calculates the distance between the distributions of generated and original images using the Inception V3 network [43], with lower values indicating greater similarity to the original images. SSIM, a method for assessing similarity between two images, compares their structural characteristics, considering factors such as brightness, contrast, and structure. Higher SSIM values indicate greater structural similarity. BRISQUE is a method for evaluating the quality of an image without reference images. This approach analyzes the natural statistical characteristics of an image through the BRISQUE model, detecting degradation in quality caused by compression or loss. A lower BRISQUE score indicates a higher quality and more natural image. These diverse measurement methods contributed to a multi-faceted analysis of the quality of de-identified images, systematically enhancing our understanding of the impact of the de-identification process on image quality.

To quantitatively measure the degree of de-identify in de-identified images, a human test of 30 questions was conducted with 20 participants. Each question presented an original image and required the participants to identify the person in the image from five options. These options included one different frame of the same person from the original image, three images evaluated as similar through FaceNet512, and one image de-identified using the technique proposed in this study. To analyze the impact of non-facial elements on person identification, 15 of the 30 questions used images with the facial area masked, whereas the remaining 15 included images with the person's hair style, clothing, and background. Examples of such designed human tests are shown in Fig. 4.

To evaluate the verification performance of de-identified images, we utilized FaceNet512 as the target facial verification model. Additionally, to assess the transferability of the proposed technique, ArcFace [44], Dlib [32], and VGG-Face [36] were employed. The verification process involved inputting two facial images into the trained models, extracting vectors, and then calculating the cosine distance between these vectors. A value below a certain threshold indicated the same individual, whereas a higher value indicated different

TABLE 1. Result of average distance and verification rate based on verification models and de-identification methods. (The variable t represents the threshold).

Model	Method	Avg.Distance	Verified(%)
FaceNet512 ($t = 0.30$)	Blur (15 by 15)	0.549	29.59%
	JPEG Comp. (quality=1%)	0.576	48.98%
	Noise ($\sigma = 0.07$)	0.229	88.78%
	Pixel (8 by 8)	0.999	0.00%
	DeepPrivacy	0.668	0.00%
	CIAGAN	0.855	0.00%
	Ours (10 step)	**0.212**	**98.37%**
ArcFace ($t = 0.68$)	Blur (15 by 15)	0.870	4.08%
	JPEG Comp. (quality=1%)	0.838	15.31%
	Noise ($\sigma = 0.07$)	0.567	69.39%
	Pixel (8 by 8)	1.000	0.00%
	DeepPrivacy	0.615	72.44%
	CIAGAN	0.774	18.36%
	Ours (10 step)	**0.290**	**99.35%**
Dlib ($t = 0.07$)	Blur (15 by 15)	0.152	56.12%
	JPEG Comp. (quality=1%)	0.389	58.16%
	Noise ($\sigma = 0.07$)	0.137	86.73%
	Pixel (8 by 8)	0.991	0.00%
	DeepPrivacy	0.053	83.67%
	CIAGAN	0.071	47.95%
	Ours (10 step)	**0.027**	**99.02%**
VGG-Face ($t = 0.40$)	Blur (15 by 15)	0.411	67.35%
	JPEG Comp. (quality=1%)	0.489	64.29%
	Noise ($\sigma = 0.07$)	0.213	89.80%
	Pixel (8 by 8)	0.995	0.00%
	DeepPrivacy	0.303	87.75%
	CIAGAN	0.425	41.83%
	Ours (10 step)	**0.125**	**99.35%**

individuals. The default thresholds for each model were 0.30 for FaceNet512, 0.68 for ArcFace, 0.07 for Dlib, and 0.40 for VGG-Face [45], [46], [47].

To quantitatively evaluate the processing speed of the proposed de-identification method, an experimental setup based on the Nvidia Jetson AGX Xavier was established. This evaluation utilized a webcam and was conducted with two different resolution settings: 1280 × 720 and 640 × 480. Furthermore, we measured the FPS in various power modes of the Nvidia Jetson AGX Xavier (MAXN, 10W, 15W, 30W) to assess the efficiency of de-identification processing speed under each setting.

B. EXPERIMENTAL RESULTS

The comparison of visual between the proposed de-identification method and other methods is shown in Fig. 1. Visual representations of de-identification techniques for each step is presented in Fig. 6.

1) FACE VERIFICATION PERFORMANCE

To evaluate the verification performance of the proposed method, we conducted assessments to measure the extent to which original faces and faces processed through various de-identification techniques can be verified. The primary evaluation was conducted using the prominent target model, FaceNet512, as the benchmark, with additional assessments conducted using ArcFace, Dlib, and VGG-Face models to evaluate transferability. Detailed results of these evaluations are summarized in Table 1.

TABLE 2. Result of human test.

Selection Rate	Unmasked	Masked	All
Same Person Selection Rate	83.67%	75.33%	79.75%
Other Person Selection Rate	8.00%	22.67%	15.33%
De-identified Person Selection Rate	8.33%	2.00%	5.17%

The evaluation results reveal that our methodology exhibited superior verification performance compared to the conventional and GAN-based de-identification techniques. Notably, our approach achieved a remarkable verification rate of 98.37% with FaceNet512, demonstrating consistent trends with verification rates of 99.35% and 99.02% across other models, as illustrated in Table 1. Notably, similar verification rates were observed for ArcFace, Dlib, and VGG-Face model, indicating a high degree of transferability across various verification models.

Through these evaluations, our de-identification technology has convincingly demonstrated its ability to maintain high verification rates and exhibit exceptional transferability across different verification models. This suggests that our technology can be effectively applied in real-world scenarios, even in environments with diverse verification systems.

2) DE-IDENTIFICATION PERFORMANCE

We conducted a human test on 20 individuals with 30 questions to evaluate the performance of the proposed de-identification method. The results are summarized in Table 2. Out of all the questions, the de-identified individuals were selected in 5.17% of the cases, whereas the same or different individuals were chosen in 94.83% of the cases. Notably, there was a 15.33% rate of misidentification, where participants incorrectly recognized different individuals. This phenomenon is attributed to the selection of alternatives with high similarity scores obtained through FaceNet512, leading to mistaken identity.

This research analyzed the impact of a masked state on identification accuracy during person recognition. The analysis revealed that identification accuracy is lower in a masked state compared to an unmasked state, suggesting that elements other than the face, such as background, hairstyle, and clothing, play a significant role in identifying individual characteristics. Furthermore, the study focused on validating the effectiveness of de-identification techniques. In the unmasked state, de-identified individuals were selected in only 8.33% of cases, indicating that correct identification was possible in many cases (83.67%). This demonstrates the considerable effectiveness of de-identification technology.

Conversely, in the masked state, the selection of de-identified individuals considerably decreased to 2.00%. This suggests that the impact of de-identification technology is enhanced when identification factors are concentrated primarily on the face. The research particularly emphasized the de-identification of the facial region, and as the results indicate, this effectively removed identifying features of the face. Overall, de-identified individuals were selected in

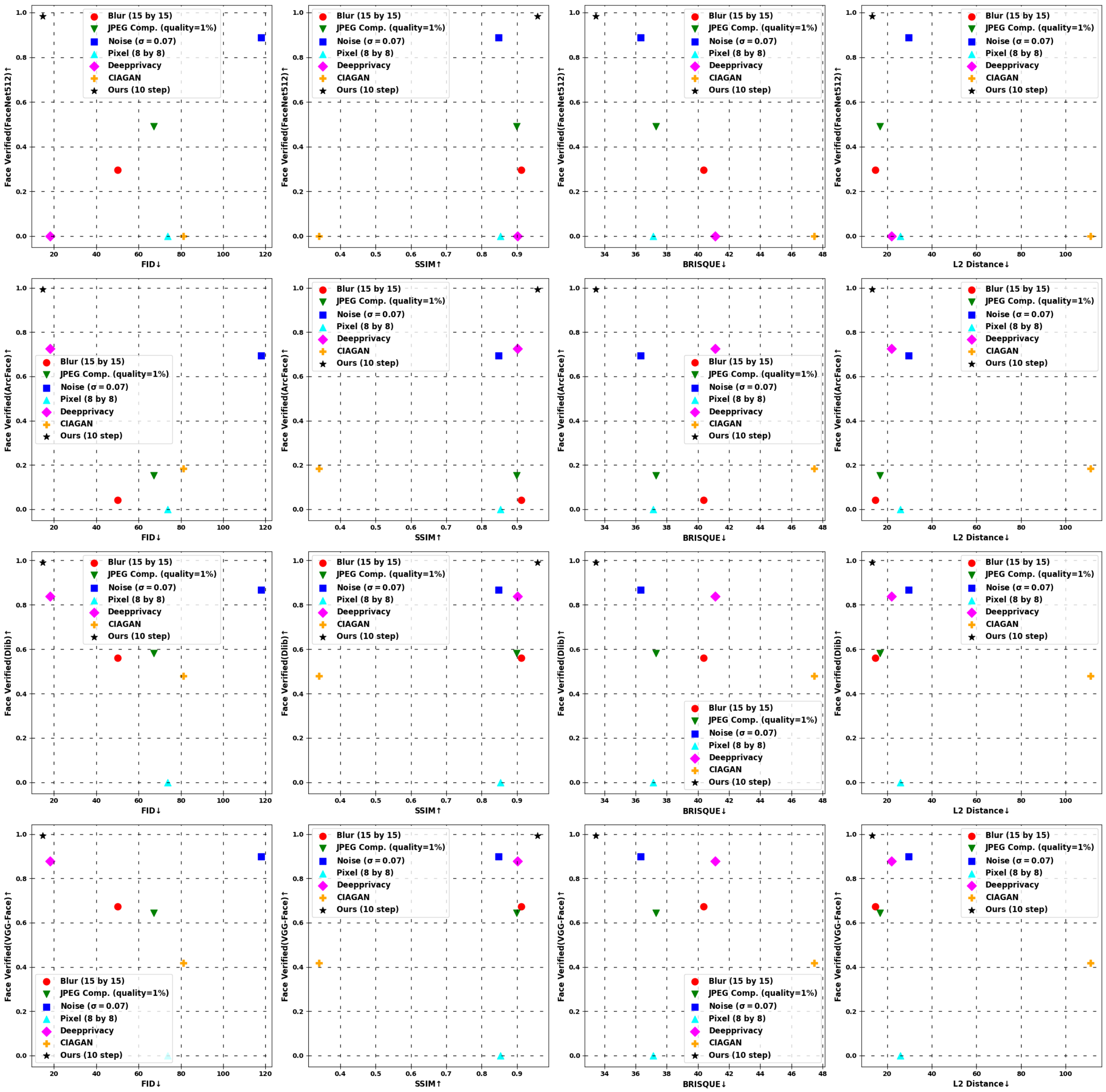


FIGURE 7. Visual comparison of image quality and face verification rates, synthesizing results from Tables 1 and 3. The top-left corner indicates the best outcomes for FID, BRISQUE, and L2 Distance, whereas the top-right indicates best outcome for SSIM.

5.17% of cases across the entire study group, irrespective of masking. These results imply that de-identification technology plays a vital role in facial privacy protection.

3) IMAGE QUALITY PERFORMANCE

To quantitatively assess the image quality of the proposed de-identification method, we compared it with various de-identification techniques. The results, including measurements of FID, SSIM, BRISQUE, and L2 distance, are summarized in Table 3. FID, SSIM, and L2 distance were calculated by comparing the original images with their de-identified counterparts, whereas BRISQUE, which does not require a reference image, was measured by inputting the de-identified images into the BRISQUE model. Our de-identification method demonstrated superior quality across all metrics, suggesting that it effectively maintains the distribution and structural similarity of the original images while achieving natural de-identification.

The comparison of image quality and verification performance between the proposed de-identification method and

TABLE 3. Results of image quality performance.

Method	FID↓	SSIM↑	BRISQUE↓	L2 Distance↓
Blur (15 by 15)	50.157	0.912	40.36	14.757
JPEG Comp. (quality=1%)	67.145	0.899	37.30	16.760
Noise ($\sigma = 0.07$)	117.847	0.848	36.34	29.616
Pixel (8 by 8)	73.837	0.853	37.14	25.927
DeepPrivacy	18.200	0.901	41.11	21.958
CIAGAN	81.237	0.340	47.46	111.281
Ours (10 step)	**14.53**	**0.958**	**33.44**	**13.295**

TABLE 4. Results of real-time performance (FPS).

Resolution	Number of people	MAXN	30W	15W	10W
640 × 480	1	9.680	5.905	5.534	4.923
	2	7.308	4.095	3.825	3.481
	3	5.936	3.255	3.339	2.559
1280 × 640	1	6.509	3.555	2.946	3.060
	2	4.038	2.185	2.292	2.158
	3	3.595	2.120	1.927	1.634

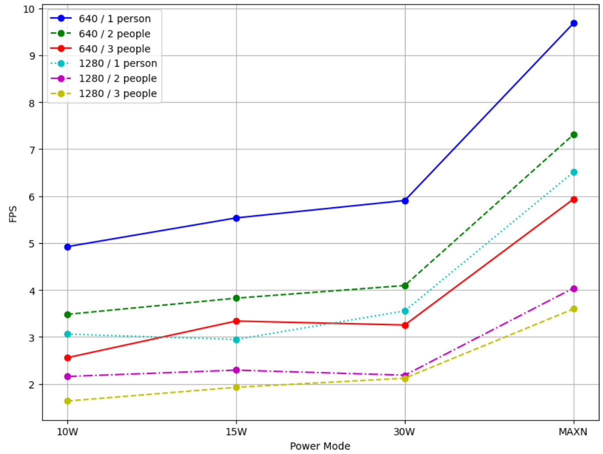


FIGURE 8. Real-time performance graph in FPS across different power modes (10 W, 15 W, 30 W, MAXN) for two resolutions (640 × 480, 1280 × 640) with varying numbers of people (1 to 3) in the scene.

other methods is shown in Fig. 7. Our method consistently outperformed others for FID, SSIM, BRISQUE, and L2 distance. Specifically, a lower FID indicates statistical similarity to the original images, positively impacting face verification performance. A higher SSIM suggests the preservation of structural quality, which correlates with higher verification performance. Our method, particularly in the Ours (10 step) variant, successfully preserves structural details while optimizing recognizability. Lower BRISQUE scores indicate better image quality, consistent with higher verification performance. The low L2 distance in Ours (10 step) signifies the generation of images closely resembling the originals, preserving features crucial for face recognition. These results indicate that the proposed method can maximize face verification performance while maintaining the visual and statistical quality of the images.

Blur (15 by 15) and JPEG Compression (quality = 1%), while maintaining structural similarity as evidenced by their SSIM and L2 distance scores, recorded high values in FID and BRISQUE, indicating an overall degradation in image quality. However, face recognition verification performance did not entirely degrade across all models, suggesting some robustness of face recognition algorithms to certain types of distortion.

Noise ($\sigma = 0.07$) recorded very high FID values and relatively lower quality in SSIM and L2 distance compared to other methods. However, it showed relatively lower BRISQUE scores. These results suggest that while structural similarity is not preserved with added noise, some non-structural quality can be preserved. The face verification performance varied across models but generally indicated high verification performance, suggesting that face verification models can be robust to a certain level of noise.

Pixel (8 by 8) showed generally low performance across all quality metrics, particularly indicating very low face verification performance. This suggests that face recognition models struggle to extract necessary information from pixelated images.

DeepPrivacy and CIAGAN, designed to disrupt original face recognition systems, unexpectedly showed high verification performance in face recognition models, except for FaceNet512. For quality, DeepPrivacy demonstrated considerably high quality, except in BRISQUE, and considerable verification performance, except for FaceNet512. CIAGAN, while showing lower performance in all quality metrics, showed considerable verification performance, albeit lower than that for DeepPrivacy.

4) REAL-TIME PERFORMANCE

In this study, the performance of the proposed method was evaluated using the Nvidia Jetson AGX Xavier. The FPS was measured according to the number of people, resolution, and power mode, and the results are summarized in Table 4. The visualization of these results according to power mode and FPS is shown in Fig. 8.

At the MAXN power mode and a resolution of 640 × 480 for a single target, the FPS was approximately 10, which corresponds to the minimum FPS threshold where continuous motion starts to be recognized. As for different power modes, increasing from 10 W to 15 W resulted in an average 6% increase in FPS at a resolution of 1280 × 640, and 18% at 640 × 480. Increasing from 15 W to 30 W showed an increase of approximately 7% and 4%, respectively; however, the change was not a relatively large. In contrast, when switching from 30 W to MAXN mode, there was a significant increase of 44% in FPS at 1280 × 640 and 75% at 640 × 480.

When switching from the lowest power mode, 10 W, to the highest mode, MAXN, there was an increase of approximately 107% at 1280 × 640 and approximately 113% at 640 × 480, nearly doubling the performance. The change in FPS with an increase in the number of people showed a decrease of approximately 29% and 32% for 1280 × 640 and 640 × 480 resolutions, respectively, when increasing from

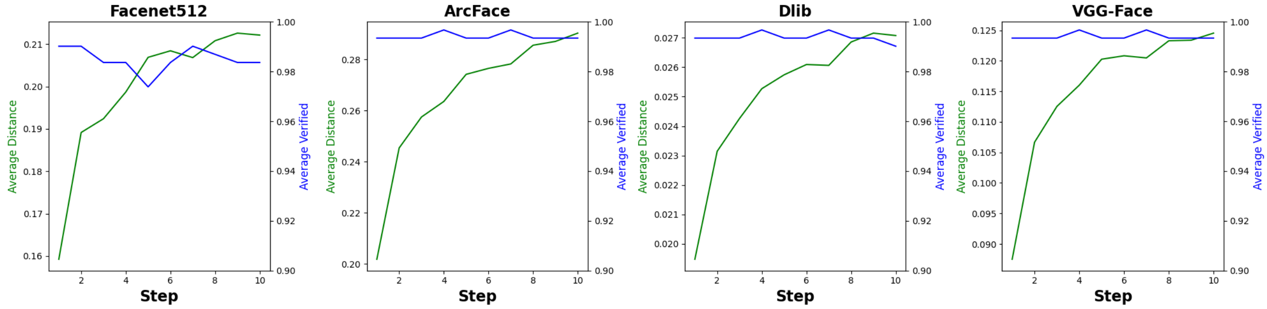


FIGURE 9. Impact of loop step increases on verification accuracy and distance metrics(Green line: major axis of the Y-axis. Blue line: minor axis of the Y-axis).

one to two people, and a decrease of 20% and 14% when increasing from two to three people.

This evaluation precisely measured the performance changes of the proposed de-identification technology on edge devices according to various variables such as resolution, power mode, and the number of subjects. Particularly, it was confirmed that the best frame rate was achieved when performing the transformation for a single subject at a resolution of 640 × 480. These results provide important guidelines for performance optimization when applying de-identification technology on edge devices. Additionally, they contribute to demonstrating the practical applicability of this technology in various operating environments.

5) ABLATION STUDY

Here, the impact of the framework's components on the overall performance was systematically analyzed. In the first experiment, the framework's performance and distance changes were measured by adjusting the number of loop steps during the processing phase. As depicted in Fig. 9, while the verification accuracy remained stable, the cosine distance gradually increased with the number of steps. Although there was a sharp increase in cosine distance at the initial stages, the change became negligible beyond a certain number of steps. This suggests that additional repetitions might decrease the framework's efficiency.

In the second experiment, the seamlessClone feature used during the face swap process was removed to analyze its effects. By excluding this feature, the impacts on verification performance and image quality were assessed. As indicated in Table 5, the cosine distance measured after the seamlessClone process was consistently lower compared to before its application across all verification models. These results demonstrate that seamlessClone processing aligns with the objectives of face verifiable de-identification proposed in this study. Further analysis of the image quality metrics in Table 5 showed that while the BRISQUE metric remained largely unchanged, improvements were noted in the FID, SSIM, and L2 distance metrics. Consequently, removing the seamlessClone feature resulted in a significant increase in cosine distance and a decrease in visual quality.

TABLE 5. Comparative analysis of verification performance with and without seamlessClone.

Measurment	Metric	Without	With
Cosine distance	Facenet512	0.281	**0.212**
	ArcFace	0.395	**0.290**
	Dlib	0.035	**0.027**
	VGG-Face	0.163	**0.125**
Image quality	FID	35.06	**14.53**
	SSIM	0.938	**0.958**
	BRISQUE	**33.32**	33.44
	L2 Distance	22.01	**13.29**

Through these experiments, this study was able to clearly identify the contributions of each component, providing crucial information for future system performance optimization. The findings offer guidance on understanding the importance of specific features and adjusting them as necessary to efficiently improve the system.

V. DISCUSSION

A. DE-IDENTIFICATION RELYING ON TARGET IMAGE

The method addressed in this study leverages the HSJA in the process of blending the styles of individuals in source images with those in generated target images. The technique mainly ensures that while individuals are perceived as different people by the human eye, they can still be verificated as the same individual by facial recognition models. This necessitates a careful combination of the styles of the source and target individuals within the verifiable boundaries of the facial recognition model.

However, a limitation revealed in our research is that when the target and source image individuals are similar, their styles become alike. Even when our technique is applied, it does not result in sufficient de-identification to the human eye. The cosine distance between the individual in the source image and the target image presented in Fig. 10., measured using FaceNet512, was relatively low at 0.598. This indicates that the facial recognition model perceives the source and target image individuals as quite similar. Analyzing both the image generated using our proposed method and an additional image to which the FSM was applied, it was confirmed that both images are still perceived as the same individual as in the source image, indicating that sufficient

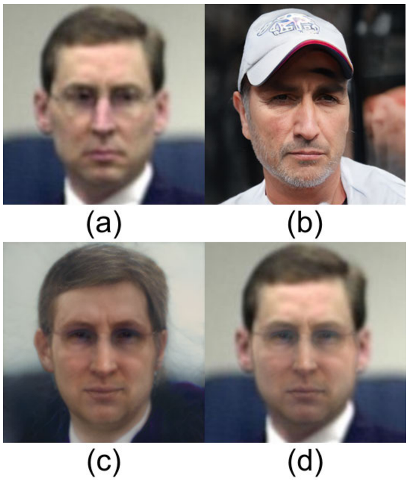


FIGURE 10. (a): source image, (b): target image. (c): image with our de-identification technique applied to the source image(10 iteration). (d): face-swapped image from (c) to (a).

de-identification was not achieved. This suggests that when the styles of the source and target are too similar, they can still be recognized as the same individual by human visual perception.

This finding provides important insights for the development and application of de-identification technologies. For verifiable de-identification, it is necessary to develop techniques that consider human visual perception more precisely, rather than just working within the boundaries of facial recognition models. Therefore, finding a balance between human visual perception and the verification boundaries of facial recognition models is an important task for future research.

B. PERFORMANCE ANALYTICS ON EDGE DEVICES

To evaluate our technology, we conducted experiments on the Nvidia Jetson AGX Xavier, an edge device engineered for AI computing. The performance benchmarks were executed in MAXN power mode at a resolution of 640 × 480 pixels. Under these conditions, the highest performance achieved was 9.680 FPS, recorded when only one person was present in the frame.

A detailed analysis of frame processing times is summarized in Table 6. The data indicates that the most time-consuming process is Face Predict, which accounts for approximately 48.37% of the total processing time. This step involves detecting faces within the frame and extracting bounding boxes, requiring significant computational power owing to the diversity and complexity of human facial features. Following closely, the Triangulation and Warp triangles process consumes 22.02% of the total time. This crucial step involves mapping and transforming the facial structure from the source face image to fit that of the destination image, which demands a high volume of computations.

TABLE 6. Analysis of processing times and percentage shares for different stages in face de-identification.

Process	Time(ms)	Percentage(%)
Frame Process	0.00095	0.98%
Face Predict	0.04692	48.37%
Face Compare	0.01322	13.63%
Face Landmark	0.00486	5.01%
Triangulation and Warp triangles	0.02136	22.02%
Replace face and seamlessCloning	0.00969	9.99%
Total	0.097	100.00%

Other processes such as Face Compare as well as Replace face and seamlessCloning occupy smaller portions of the total time, specifically 13.63% and 9.99%, respectively. Face Compare is necessary for differentiating between individuals, entailing the extraction and comparison of face embeddings. Replace face and seamlessCloning involves attaching the transformed source face image to the destination images facial structure followed by seamless cloning for a natural integration. The least time-consuming steps are Face Landmark and Frame Process, collectively accounting for less than 6% of the total time. Face Landmark detects essential facial landmarks for subsequent processing stages, whereas Frame Process includes initial frame preparation tasks such as resizing and color adjustments.

The results of this analysis are instrumental in identifying processes that are time-intensive and considering strategies for time reduction. The Face Predict process, taking up a substantial portion of the total time, is a primary candidate for performance optimization. It is anticipated that replacing the facial recognition model with a lightweight model or quantizing the existing model could reduce the time spent on this process. Moreover, for the Triangulation and Warp triangles process, the potential to decrease computation time through the adoption of GPU acceleration or parallel processing, as opposed to the current CPU-centric computation method, is promising. Such optimizations are expected to play a critical role in enhancing real-time performance on edge devices.

VI. CONCLUSION AND FUTURE WORKS

In conclusion, we propose a framework for face-verifiable de-identification in real-time video surveillance. Utilizing face-verifiable de-identification techniques and FSM, it maintains sufficient de-identification from the original source images while preserving the performance of various face verification models. By applying this method, potential violations of facial privacy in scenarios requiring both face verification and de-identification can be prevented, demonstrating feasibility for real-time applications.

In future, we will focus on addressing the key limitations identified in the research. Specifically, we aim to resolve the issue of insufficient de-identification when the original and generated images subjects closely resemble each other. This will be achieved by introducing modifications in the image generation process for more effective de-identification. Additionally, improvements in real-time performance will be pursued through facial recognition model quantization and computational optimization. Through these measures, we expect to significantly elevate the practicality and effectiveness of the research. These planned improvements are expected to significantly augment the practicality and effectiveness of our framework, paving the way for more robust and reliable facial privacy protection in an increasingly digital and surveillance-oriented world.

ACKNOWLEDGMENT

GPT-4 was used in the following ways: to help them iterate on LaTeX formatting; for text summarization; and as a copyediting tool.

REFERENCES

[1] K. Kosulin and A. Karpov, "A survey of masked face recognition methods and corpora/data," in *Proc. Int. Conf. Internet Modern Soc.* Springer, 2022, pp. 27–37.

[2] A. M. Islam, "Exploring convolutional neural networks for facial expression recognition: A comprehensive survey," *Global Mainstream J. Innov., Eng. Emerg. Technol.*, vol. 3, no. 2, pp. 14–26, 2024.

[3] M. H. A. Alkreem, R. S. Salman, and F. K. Al-Jibory, "Detect people's faces and protect them by providing high privacy based on deep learning," *Tehnicki Glasnik*, vol. 18, no. 1, pp. 92–99, Jan. 2024.

[4] K. Zhang, W. Ren, W. Luo, W.-S. Lai, B. Stenger, M.-H. Yang, and H. Li, "Deep image deblurring: A survey," *Int. J. Comput. Vis.*, vol. 130, no. 9, pp. 2103–2130, Sep. 2022.

[5] C. Tian, L. Fei, W. Zheng, Y. Xu, W. Zuo, and C.-W. Lin, "Deep learning on image denoising: An overview," *Neural Netw.*, vol. 131, pp. 251–275, Nov. 2020.

[6] L. Cavedon, L. Foschini, and G. Vigna, "Getting the face behind the squares: Reconstructing pixelized video streams," in *Proc. 5th USENIX Workshop Offensive Technol.*, 2011, pp. 1–9.

[7] J. Jam, C. Kendrick, K. Walker, V. Drouard, J. G.-S. Hsu, and M. H. Yap, "A comprehensive review of past and present image inpainting methods," *Comput. Vis. Image Understand.*, vol. 203, Feb. 2021, Art. no. 103147.

[8] Y. Liu, Q. Li, Z. Sun, and T. Tan, "A3GAN: An attribute-aware attentive generative adversarial network for face aging," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 2776–2790, 2021.

[9] O. Gafni, L. Wolf, and Y. Taigman, "Live face de-identification in video," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 9377–9386.

[10] Y. Wu, F. Yang, Y. Xu, and H. Ling, "Privacy-protective-GAN for privacy preserving face de-identification," *J. Comput. Sci. Technol.*, vol. 34, no. 1, pp. 47–60, Jan. 2019.

[11] Z. Cai, Z. Gao, B. Planche, M. Zheng, T. Chen, M. S. Asif, and Z. Wu, "Disguise without disruption: Utility-preserving face de-identification," 2023, *arXiv:2303.13269*.

[12] L. Du, M. Yi, E. Blasch, and H. Ling, "GARP-face: Balancing privacy protection and utility preservation in face de-identification," in *Proc. IEEE Int. Joint Conf. Biometrics*, Sep. 2014, pp. 1–8.

[13] E. Chatzikyriakidis, C. Papaioannidis, and I. Pitas, "Adversarial face de-identification," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2019, pp. 684–688.

[14] H. Hukkelås, R. Mester, and F. Lindseth, "Deepprivacy: A generative adversarial network for face anonymization," in *Proc. Int. Symp. Vis. Comput.* Springer, 2019, pp. 565–578.

[15] M. Maximov, I. Elezi, and L. Leal-Taixé, "CIAGAN: Conditional identity anonymization generative adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 5446–5455.

[16] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 4396–4405.

[17] E. Richardson, Y. Alaluf, O. Patashnik, Y. Nitzan, Y. Azar, S. Shapiro, and D. Cohen-Or, "Encoding in style: A styleGAN encoder for image-to-image translation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 2287–2296.

[18] J. Chen, M. I. Jordan, and M. J. Wainwright, "HopSkipJumpAttack: A query-efficient decision-based attack," in *Proc. IEEE Symp. Secur. Privacy (SP)*, May 2020, pp. 1277–1294.

[19] (2018). *Jetson Agx Xavier*. NVIDIA. Accessed: Nov. 23, 2023. [Online]. Available: https://www.nvidia.com/en-us/autonomous-machines/embedded-systems/jetson-xavier-series/

[20] C. Neustaedter, S. Greenberg, and M. Boyle, "Blur filtration fails to preserve privacy for home-based video conferencing," *ACM Trans. Comput.-Hum. Interact.*, vol. 13, no. 1, pp. 1–36, Mar. 2006.

[21] K. Mivule, "Utilizing noise addition for data privacy, an overview," 2013, *arXiv:1309.3958*.

[22] H. Liu, M. Steinebach, R. Stein, and F. Mayer, "Privacy preserving forensics for JPEG images," *Electron. Imag.*, vol. 30, no. 7, pp. 1–6, Jan. 2018.

[23] L. Fan, "Image pixelization with differential privacy," in *Data and Applications Security and Privacy XXXII*, Bergamo, Italy. Springer, 2018, pp. 148–162.

[24] R. Gross, E. Airoldi, B. Malin, and L. Sweeney, "Integrating utility into face de-identification," in *Privacy Enhancing Technologies*, Cavtat, Croatia. Springer, 2006, pp. 227–242.

[25] N. Ruchaud and J.-L. Dugelay, "Automatic face anonymization in visual data: Are we really well protected?" *Electron. Imag.*, vol. 28, no. 15, pp. 1–7, Feb. 2016.

[26] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 27, 2014, pp. 1–9.

[27] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," 2015, *arXiv:1511.06434*.

[28] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training GANs," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 29, 2016, pp. 1–9.

[29] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of GANs for improved quality, stability, and variation," 2017, *arXiv:1710.10196*.

[30] A. Makhzani, J. Shlens, N. Jaitly, I. Goodfellow, and B. Frey, "Adversarial autoencoders," 2015, *arXiv:1511.05644*.

[31] G. Jocher, A. Chaurasia, and J. Qiu. (Jan. 2023). *YOLO By Ultralytics*. [Online]. Available: https://github.com/ultralytics/ultralytics

[32] D. E. King, "Dlib-Ml: A machine learning toolkit," *J. Mach. Learn. Res.*, vol. 10, pp. 1755–1758, Jan. 2009.

[33] D. T. Lee and B. J. Schachter, "Two algorithms for constructing a Delaunay triangulation," *Int. J. Comput. Inf. Sci.*, vol. 9, no. 3, pp. 219–242, Jun. 1980.

[34] I. Culjak, D. Abram, T. Pribanic, H. Dzapo, and M. Cifrek, "A brief introduction to OpenCV," in *Proc. 35th Int. Conv. MIPRO*, May 2012, pp. 1725–1730.

[35] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 12, pp. 4217–4228, Dec. 2021.

[36] O. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *Proc. BMVC*, 2015, pp. 1–12.

[37] S. Yang, P. Luo, C. C. Loy, and X. Tang, "WIDER FACE: A face detection benchmark," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 5525–5533.

[38] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," Univ. Massachusetts, Amherst, Tech. Rep., 07-49, Oct. 2007.

[39] G. B. H. E. Learned-Miller, "Labeled faces in the wild: Updates and new reporting procedures," Univ. Massachusetts, Tech. Rep., UM-CS-2014-003, May 2014.

[40] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "GANs trained by a two time-scale update rule converge to a local Nash equilibrium," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–12.

[41] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[42] 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.

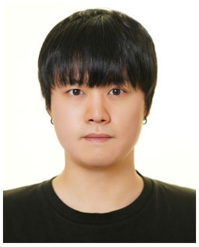
[43] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2818–2826.

[44] 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.

[45] A. Firmansyah, T. F. Kusumasari, and E. N. Alam, "Comparison of face recognition accuracy of ArcFace, facenet and Facenet512 models on deepface framework," in *Proc. Int. Conf. Comput. Sci., Inf. Technol. Eng. (ICCoSITE)*, Feb. 2023, pp. 535–539.

[46] S. I. Serengil and A. Ozpinar, "LightFace: A hybrid deep face recognition framework," in *Proc. Innov. Intell. Syst. Appl. Conf. (ASYU)*, Oct. 2020, pp. 1–5.

[47] (2020). *Fine Tuning the Threshold in Face Recognition*. Sefik Ilkin Serengil. Accessed: Apr. 15, 2024. [Online]. Available: https://sefiks.com/2020/05/22/fine-tuning-the-threshold-in-face-recognition/



SUNGJUNE PARK received the B.S. degree in software engineering from Soongsil University, South Korea, in 2023, where he is currently pursuing the M.S. degree in software convergence. His research interests include information privacy protection and AI security.

HYUNSIK NA received the B.S. degree in applied mathematics from Kongju National University, South Korea, in 2021. He is currently pursuing the Ph.D. degree in software convergence with Soongsil University, South Korea. His research interests include AI security, information security, and edge AI.

DAESEON CHOI (Member, IEEE) received the B.S. degree in computer science from Dongguk University, South Korea, in 1995, the M.S. degree in computer science from Pohang Institute of Science and Technology, South Korea, in 1997, and the Ph.D. degree in computer science from Korea Advanced Institute of Science and Technology (KAIST), South Korea, in 2009. From September 2015 to August 2020, he was a Professor with the Department of Medical Information, Kongju National University, South Korea. He is currently a Professor with the Department of Software, Soongsil University, South Korea. His research interests include identity management and information security.