Achieving Privacy-Preserving Multi-View Consistency with Advanced 3D-Aware Face De-identification

Jingyi Cao
Institute of Image Communication and Network Engineering, Shanghai Jiao Tong University
Shanghai, China
cjycaojingyi@sjtu.edu.cn

Bo Liu
School of Computer Science
University of Technology Sydney
Sydney, Australia
bo.liu@uts.edu.au

Yunqian Wen
Institute of Image Communication and Network Engineering, Shanghai Jiao Tong University
Shanghai, China
wenyunqian@sjtu.edu.cn

Rong Xie
Institute of Image Communication and Network Engineering, Shanghai Jiao Tong University
Shanghai, China
xierong@sjtu.edu.cn

Li Song
Institute of Image Communication and Network Engineering, Shanghai Jiao Tong University
Shanghai, China
song_li@sjtu.edu.cn

ABSTRACT

The widespread application of face recognition technology has exacerbated privacy threats. Face de-identification is an effective means of protecting visual privacy by concealing identity information. While deep learning-based methods have greatly improved de-identification results, most existing algorithms rely on 2D generative models that struggle to produce identity-consistent results for multiple views. In this paper, we focus on identity disentanglement within the latest 3D-aware face generation model, and propose an advanced face de-identification framework that can be applied to various scenarios. Our proposed framework disentangles identity from other facial features, modifies only the former and generates the de-identified face using a 3D generator. This approach results in high-quality, identity-consistent de-identification that preserves other facial features. We demonstrate our approach on StyleNeRF, one of the most widely-used style-based neural radiation field models. Through extensive experiments, we demonstrate the effectiveness of our approach in achieving face de-identification both for a single image and group images with the same identity. Our work is a significant step forward in the field of face de-identification, opening up new possibilities for practical applications.

CCS CONCEPTS

• **Computing methodologies** → **Computer vision representations**; • **Security and privacy** → **Privacy protections**; **Usability in security and privacy**;

KEYWORDS

Face De-identification, Image Privacy, 3D-aware Generative Models

1 INTRODUCTION

The proliferation of computer vision technologies like surveillance cameras and online video conferencing has made large-scale visual data collection possible. Additionally, the private images people share on social media are at risk of malicious attacks, posing a privacy threat. Privacy issues are gradually receiving attention, and there is an urgent need for more advanced protection methods.

Face de-identification is considered an effective way to protect sensitive information. Ribaric et al. [27] defined de-identification in multimedia content as *"the process of concealing or removing personal identifiers, or replacing them with surrogate personal identifiers"*. Traditional methods, such as blurring and pixelation, tend to scramble the image content directly on pixels, which may greatly impair the visual quality and provide limited protection against face recognition models. Recently, Generative Adversarial Networks (GANs) have achieved impressive success in face image generation, and they have also been applied to privacy protection [17, 19, 21], significantly improving the quality of de-identification results. Furthermore, research on semantic exploration in latent space has proved that face images can be semantically divided into identity-related representations and identity-independent attributes. Identity disentanglement enables us to de-identify images through more precise editing for identity [1, 14, 20, 34] while still maintaining as much visual similarity to the original image as possible.

However, existing face de-identification methods are always based on general GANs, which only operate in a single view and fail to obtain multi-view-consistent image synthesis. In contrast, 3D-aware generative models are excel in handling the relationship between facial content and viewing directions. De-identification based on 3D generative models can better preserve the multi-view

identity consistency, thus adapting to richer application scenarios such as computer animation and beyond.

Due to the excellent performance of Neural Radiance Fields (NeRFs) [23] in scene reconstruction, there has been a mainstream trend to enhance 3D structures by incorporating them into face generation [11, 38, 39]. Drawing on style-based generators [16], style-based NeRFs [2, 9, 31] have also been proposed to address the computationally expensive problem of NeRFs in rendering high-resolution images. 3D-aware GANs perform well in explicitly modeling objects' geometry, but there is still limited research on NeRF inversion or controlled face editing by 3D-aware generators.

With a focus on exploring the latent space of 3D-aware generators, we aim to address the problem of identity disentanglement and investigate de-identification for various applications. Our proposed framework disentangles face features by two encoders, where E_{id} extracts identity and E_{aux} extracts auxiliary information including non-identity attributes and camera directions. The extracted attributes are combined with identity and fed into identity conversion mapper M_c to create a new representation in latent space. Finally, we generate the de-identified results by 3D-aware generator G_{3d}, which is hoped to have a different identity but high visual similarity. We take StyleNeRF [9] as an example to validate the effectiveness of key components. Specifically, our approach is highly adaptable and can be implemented using other generators.

The major contributions of our work can be summarized as:

(1) We propose a novel solution to the challenging problem of face de-identification in 3D radiance fields, which can only modify the disentangled identity. Our approach is versatile and can be applied to various scenarios, including single-view, multi-view, and group de-identification.
(2) Our novel framework is based on a two-stage process. First, we employ two encoders that disentangle identity and auxiliary information. Next, a mapping network transforms the identity, while a 3D generator reconstructs the face with high fidelity, preserving non-identity attributes.
(3) We conduct extensive experiments to evaluate our framework's performance, and the results show that our approach achieves superior image quality and identity consistency for multi-view rendering, making it a compelling choice for various real-world applications.

2 RELATED WORK

2.1 Exploration of 3D-aware face generation

Early methods [24] attempted to learn pose from 2D-GANs by disentangling pose representations, but they always rely on annotation or 3D Morphable Model (3DMM) auxiliary information [33] for supervision. The successful introduction of NeRFs leads to a new paradigm for 3D-aware face generation. Some frameworks [3, 6, 25, 30] utilize NeRF as a 3D representation for GAN generation, and NeRF-GAN can learn ensemble information from unlabeled images and provide explicit control based on volume rendering.

Some algorithms [2, 9, 26] draw on the success of StyleGAN [16] to propose style-based NeRFs, which provide an efficient manner for high-resolution geometry-aware image generation and also facilitate the explicit control [7, 11, 18, 37] of 3D-aware generated content. For example, CoRF [37] embeds motion features in hierarchical latent space, enabling editing for identity, viewing direction, and motion. HeadNeRF [11] is a novel NeRF-based parametric head model that can directly control the rendering pose, identity, expression, and appearance. SURF-GAN [18] successfully discovers semantic attributes and controls them in an unsupervised manner.

Nevertheless, most of these methods require training for generators under supervised conditions. We target disentanglement and image synthesis with pre-trained style-based NeRF, which can exploit their advanced generative capabilities, expressive latent space and without a training burden.

2.2 Face de-identification

Conventional algorithms scrambled images directly at the pixel level, failing to achieve a satisfactory trade-off between privacy and utility. Deep learning techniques have led to significant leaps in the quality of face de-identification results. This was accomplished by revisiting the face completion or face synthesis task initially. Sun et al. [32] proposed to generate head inpainting based on landmarks. DeepPrivacy [13] can automatically replace the original face region with a generated one without altering posture and background. This type of algorithm can ensure the deletion of all privacy-sensitive information, but it cannot effectively retain similarity. Another type of de-identification method utilizes auxiliary modules or loss functions. PP-GAN [35] introduced two additional modules: verificator and regulator, which are used to constrain the identity away from the original sample and to preserve the structural similarity, respectively. Zhao et al. [36] employed an adjustable privacy-related loss in the training process, allowing the generated results to have a certain identity distance from the original. The study of latent spaces and disentanglement has made face de-identification more targeted and fine-grained. IdentityDP [34] applied differential privacy mechanisms for adjustable privacy control on individual identity embedding. FICGAN [14] extracted potential encodings of ID and non-ID and designed a layer-wise generator structure to ensure de-identification and attribute retention control. However, most of the existing algorithms are based on traditional GANs and designed for still frontal faces, which cannot adapt to multiple views or large angles. We hope to introduce the 3D advantage of NeRF-GAN image generation to de-identification.

3 PROBLEM FORMULATION

Face de-identification is a rapidly growing research area with a wide range of applications in security, privacy and data protection. Our method can be applied to three categories for face de-identification, including single-view images, multi-view images and image sets of the same identity, with the following definitions.

Single-View De-identification. To protect a single image X, general de-identification algorithm $\mathcal{F}$ can be formulated as,

$$ID(\mathcal{F}(X)) \neq ID(X), \tag{1}$$

where $\mathcal{F}(X)$ indicates the de-identified result and $ID(X)$ represents the identity of the input image X. Considering the image utility, we prefer that $\mathcal{F}(X)$ looks similar to X as well as keeping other identity-irrelevant attributes and viewpoints consistent.

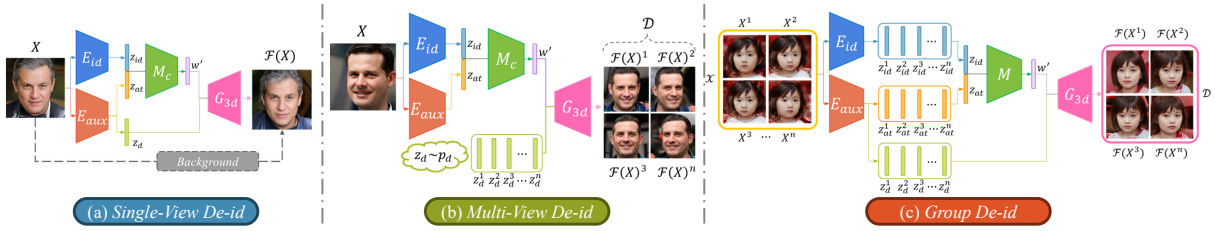

Figure 1: Different de-identification applications.

Multi-View De-identification. Our approach supports the generation of multi-view de-identification results for a single image. For the input image X, we can obtain multi-view de-identification results $\mathcal{D} = \{\mathcal{F}(X)^1, \mathcal{F}(X)^2, \ldots, \mathcal{F}(X)^n\}$ as,

$$ID(\mathcal{D}) \neq ID(X), \quad ID(\mathcal{F}(X)^i) = ID(\mathcal{F}(X)^j) \ \ \forall i, j \in [1, n], \quad (2)$$

where $\mathcal{F}(X)^i$ indicates the de-identified result $\mathcal{F}(X)$ from multiple views $z_d = \{z_d^1, z_d^2, \cdots, z_d^n\}$.

Group De-identification. In our approach, it could also achieve effective anonymization for a series of images $\mathcal{X} = \{X^1, X^2, \ldots, X^n\}$ of the same person under different viewpoints, and the de-identified set $\mathcal{D} = \{\mathcal{F}(X^1), \mathcal{F}(X^2), \ldots, \mathcal{F}(X^n)\}$ are still of the same identity under the same conditions, which can be formulated as,

$$ID(\mathcal{D}) \neq ID(\mathcal{X}), \quad ID(\mathcal{F}(X^i)) = ID(\mathcal{F}(X^j)) \ \ \forall i, j \in [1, n]. \quad (3)$$

4 OUR APPROACH

4.1 Overview

As shown in Figure 1, our proposed method can be used for different applications. All the above frameworks share the same fundamental module composition and the same training process. The auxiliary encoder E_{aux} and identity conversion mapper M_c are trainable while the pre-trained identity encoder E_{id} and 3D-aware generator G_{3d} remain frozen. The training process consists of two training stages, firstly for basic disentanglement in latent space and secondly for more fine-grained tuning of de-identification. More details will be further explained in Subsec. 4.2 and Subsec. 4.3.

4.2 Training Stage 1: Disentanglement

As illustrated in Figure 2, we input two images in each iteration, noted as X_i and X_j, where z_{id} is provided by X_i while z_{at} and z_d by X_j. We hope the information contained in z_{id} and z_{at} that satisfies (1) representative of the entire face, while (2) as independent as possible. These two points are achieved by the following training modes: (1) face reconstruction ($X_i = X_j$), using the extracted z_{id} and z_{at} to restore, and (2) face swapping ($X_i \neq X_j$), using two images to provide z_{id} and z_{at}, the results obtained can maintain the correspondence with the two images respectively. To unify the notation, we use $\hat{X}_{i\to j}$ to denote the generated results in both modes in the subsequent equations.

We employ identity consistency loss between X_i and the generated result $\hat{X}_{i\to j}$ as,

$$\mathcal{L}_{id} = \left\| E_{id}(X_i) - E_{id}\left(\hat{X}_{i\to j}\right) \right\|_2, \quad (4)$$

where E_{id} is a pre-trained face recognition model [29] that provides supervision. Similarly, the training of E_{aux} involves attributes consistency loss $\mathcal{L}_{attr}$ and camera direction loss $\mathcal{L}_d$ as,

$$\mathcal{L}_{attr} = \left\| z_{at}(X_j) - z_{at}(\hat{X}_{i\to j}) \right\|_2, \ \mathcal{L}_d = \left\| z_d(X_j) - z_d(\hat{X}_{i\to j}) \right\|_1. \quad (5)$$

To emphasize the face region while preserving geometric features and expressions, we introduce a face parsing net [1] as E_{mask}, which segments the entire face into different categories and calculates the loss on the overlapping regions of pairs as,

$$\mathcal{L}_{mask} = \left\| \left(E_{mask}(X_j) - E_{mask}(\hat{X}_{i\to j}) \right) \odot \left(X_j - \hat{X}_{i\to j} \right) \right\|_2, \quad (6)$$

where $\odot$ denotes an element-wise product.

Apart from the above basic loss used in both training strategies, there are some other constraints utilized solely. In reconstruction, $\mathcal{L}_{inv}$ in Equ.(7) is applied to enhance the training stability. We random sample latent codes $\boldsymbol{w}$ and camera poses $\boldsymbol{d}$ as ground truth.

$$\mathcal{L}_{inv(w)} = \left\| \boldsymbol{w} - \boldsymbol{w}(\hat{X}_{i\to j}) \right\|_1, \ \mathcal{L}_{inv(d)} = \left\| \boldsymbol{d} - z_d(\hat{X}_{i\to j}) \right\|_1. \quad (7)$$

Additionally, a reconstruction loss $\mathcal{L}_{rec}$ followed pSp [28] is applied to capture both perceptual-level $\mathcal{L}_{LPIPS}$ and pixel-level $\mathcal{L}_2$ loss.

In face swapping, $\hat{X}_{i\to j}$ is encouraged to have the same identity with X_i and other identity-independent features are in accordance with X_j. Inspired by contrastive learning [4, 15], we applied $\mathcal{L}_{con}$ to encourage the generated identity to be close to z_{id}^i while being away from z_{id}^j, which can be expressed as,

$$\left\| z_{id}^i - \hat{z}_{id}^{i\to j} \right\|_2^2 + m < \left\| \hat{z}_{id}^{i\to j} - z_{id}^j \right\|_2^2, \quad (8)$$

where $m > 0$ is the margin used to define the distance difference between the positive pair and negative. To ensure the mapped new latent codes $\hat{\boldsymbol{w}}_{i\to j}$ does not deviate far from the latent space, we use a latent discriminator denoted as D (shown in Figure 3), which is trained in an adversarial scheme with the non-saturating GAN loss [8] and R1 regularization [22]. We set corresponding hyperparameters to adjust the weight of losses and control different training modes.

4.3 Training Stage 2: De-identification

The key to de-identification is to alter the identity in face images. Therefore, in the second training stage, we incorporate a privacy parameter τ to transform the original identity loss $\mathcal{L}_{id}$ into a de-identification loss, denoted as $\mathcal{L}_{de-id}$. We then fine-tune the network with the modified loss to regulate the identity distance between the generated output and the source image. The $\mathcal{L}_{de-id}$ can be formulated as,

$$\mathcal{L}_{de-id} = \left| \tau - \left\| E_{id}(X) - E_{id}(X') \right\|_2 \right|, \quad (9)$$

[1] Code available at https://github.com/zllrunning/face-parsing.PyTorch

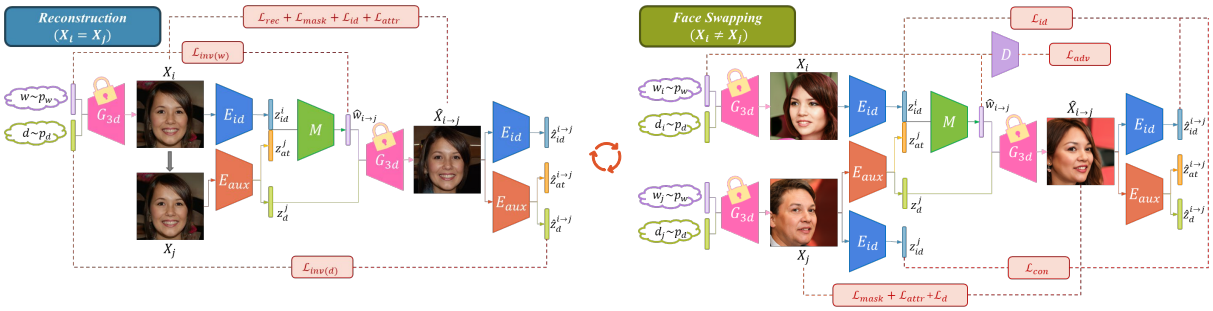

Figure 2: The framework of the disentanglement training, where two strategies are used alternately to achieve disentanglement.

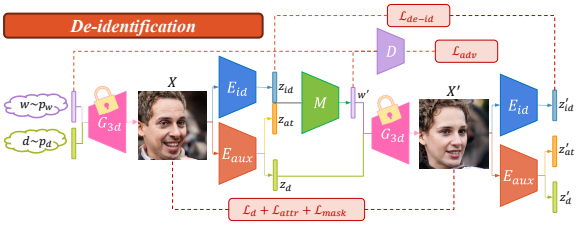

Figure 3: An overview of de-identification training phases.

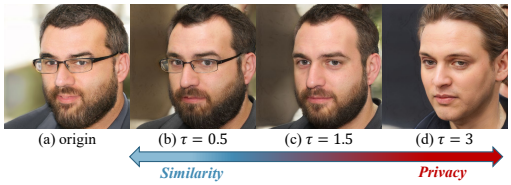

Figure 4: The impact of the privacy parameter τ in Eq. (9).

and the flow is shown in Figure 3.

The main objective is to achieve de-identification and expand the range of identities that can be generated, even including new identities that may not exist in the training dataset. On the other hand, it is equivalent to a certain relaxation of the identity constraint and can lead to better retention of desirable facial properties. We also use attributes consistency loss $\mathcal{L}_{attr}$, direction loss $\mathcal{L}_d$ in Eq. (5), and mask loss $\mathcal{L}_{mask}$ in Eq. (6) to ensure the de-identified images maintain a high similarity and the same viewpoint with the original image. As the de-identification procedure involves generating new samples, we employ adversarial loss $\mathcal{L}_{adv}$ as well, and the total loss $\mathcal{L}_{dt}$ can be summarized in Eq. (10). After this stage, our model will no longer require other reference images to generate de-identified results.

$$\mathcal{L}_{dt} = \mathcal{L}_{de-id} + \lambda_{attr}\mathcal{L}_{attr} + \lambda_d\mathcal{L}_d + \lambda_{mask}\mathcal{L}_{mask} + \lambda_{adv}\mathcal{L}_{adv}. \quad (10)$$

5 EXPERIMENTS

5.1 Implementation Details

Network Architecture. The identity encoder E_{id} is a pre-trained face recognition model [29] to obtain $z_{id} \in \mathcal{R}^{512}$. The auxiliary encoder E_{aux} consists of *ResBlock* as the base module to extract $z_{attr} \in \mathcal{R}^{7\times512}$ and $z_d \in \mathcal{S}^3$. The identity conversion mapper M_c is a five-layer MLP. All our experiments are conducted by StyleNeRF [9] pretrained on FFHQ [16] as the 3D-aware generative model.

Datasets. We train E_{aux} and M_c using multi-view synthesized images from StyleNeRF, which contain 50,000 images with the resolution of 512×512 generated by random latent codes $\boldsymbol{w}$ and camera poses $\boldsymbol{d}$. In particular, the latent codes are ideally view-independent, so that the dataset also contains the images generated with the same latent codes and different camera poses.

Experimental Settings. We optimize the adversarial loss and non-adversarial losses separately for a more stable training procession. We train our network using Adam with $\beta_1 = 0.9$, $\beta_2 = 0.999$, and the learning rate is set as 6×10^{-5} when optimizing non-adversarial losses. For adversarial learning, we set the learning rate 5×10^{-6} for encoders and 2×10^{-6} for latent discriminator D. The tradeoff parameters are set to $\lambda_{id} = \lambda_{mask} = \lambda_{rec} = 1, \lambda_{attr} = 100, \lambda_d = \lambda_{inv(d)} = 10, \lambda_{inv(w)} = 0.1, \lambda_{con} = 0.5$. In *Training Stage1*, the two processes of reconstruction and face swapping are alternated in a ratio of 2:1. The network is trained end-to-end on a single GeForce RTX 3090 GPU with a batch size of 8.

5.2 Qualitative Analysis

Baselines. We mainly compare our approach with deep learning-based face de-identification methods, including AMT-GAN [12], DeepPrivacy [13], CIAGAN [21], Gu et al. [10] and IdentityDP ($\varepsilon = 0.5$) [34]. These algorithms have different design approaches and represent advanced techniques in the field of de-identification.

Privacy Parameter. As demonstrated by the experimental results presented in Figure 4, we discovered that the privacy parameter τ in Eq. (9) was too small to effectively transform the identity, while a value that was too large could negatively impact attribute consistency. Therefore, we set the privacy parameter as $\tau = 1.5$ to balance the tradeoff between privacy and similarity.

5.2.1 Single-View De-identification. The de-identification results compared with baselines are shown in Figure 5. AMT-GAN [12] (line-(b)) is designed based on adversarial perturbation combined with the makeup transfer task. The de-identification methods based on adversarial examples are often difficult to robust to various face recognition models. DeepPrivacy [13] (line-(c)) replaces the original face area by randomly generating faces, which has a better generation effect and higher privacy level, but it does not consider the similarity preservation and the generated results may have unnatural expressions. CIAGAN [21] (line-(d)) has poor image quality with obvious artifacts and low resolution. Gu et al. [10] (line-(e)) proposed a face identity transformer conditioned on passwords to enable anonymization and de-anonymization, where there may exist blob-like artifacts. IdentityDP [34] (line-(f)) retains more similarity by adding noise to the disentangled identity features, which

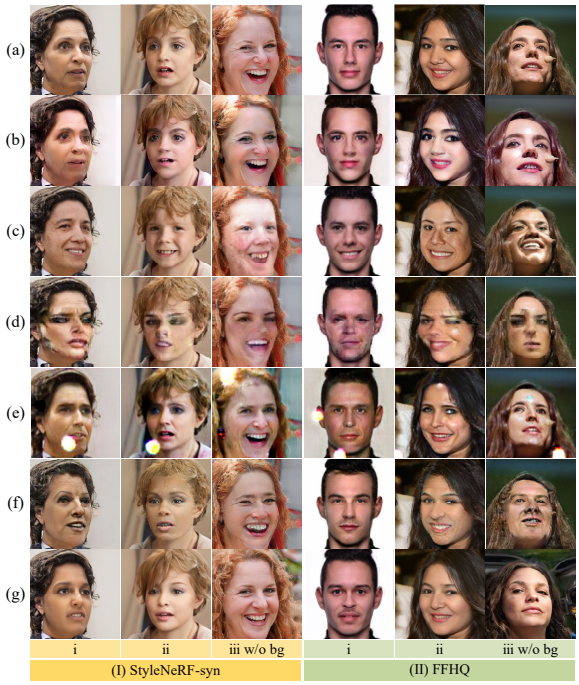

Figure 5: Single-View de-identification results, where (a) original image (b) AMT-GAN (c) DeepPrivacy (d) CIAGAN (e) Gu et al. (f) IdentityDP($\epsilon = 0.5$) and (g) ours.

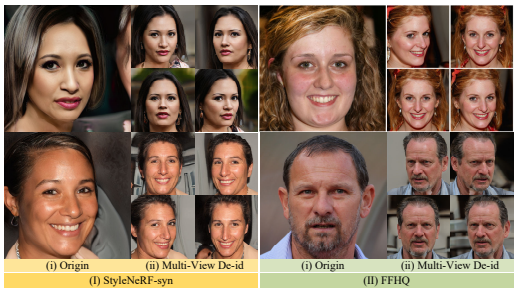

Figure 6: Multi-view de-identification results.

performs effectively for frontal images but it may fail when the face is viewed at a large angle. Our method achieves superior de-identification results, particularly with side or large-angle faces. It should be noted that we mainly focus on the face region and introduce some randomness in rendering the background to enhance naturalness so that the same background replacement strategy as CIAGAN [21] is applied based on facial masks.

5.2.2 Multi-View De-identification. The qualitative results are displayed in Figure 6. It can be seen that our method can generate multi-view de-identification results while maintaining identity consistency when varying directions. The ability to generate de-identification results from multiple viewpoints is relevant to privacy issues such as avatar generation.

5.2.3 Group De-identification. We generate a series of images with the same identity from different views and the results are presented in Figure 7. In other methods, the same de-identification conditions are typically applied across all images in each group in order to

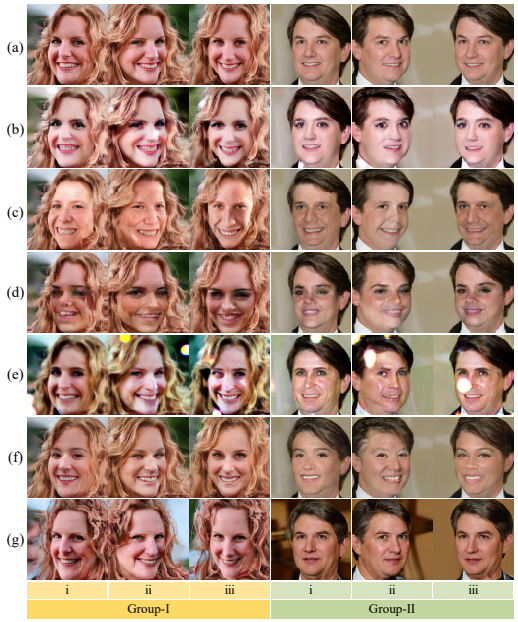

Figure 7: Group de-identification results, where (a) original image (b) AMT-GAN (c) DeepPrivacy (d) CIAGAN (e) Gu et al. (f) IdentityDP($\epsilon = 0.5$) and (g) ours.

maximize the identity consistency of de-identified results. Specifically, AMT-GAN [12] uses the same reference images for makeup transfer, Gu et al. [10] applies the same password as condition, and IdentityDP [34] adds constant noise to each image in the same set.

DeepPrivacy [13] generates a new face for each image for replacement, and there is a large identity variation. Even if the same noise is added, IdentityDP [34] generates results with more significant identity differences. While AMT-GAN [12], CIAGAN [21] and Gu et al. [10] have better control of identity, the image quality is less satisfactory due to the presence of artifacts. In our approach, we employ a novel technique that involves utilizing the same latent code to guide the generation process. This unique method allows us to maintain a remarkable level of consistency in terms of identity, even when dealing with complex and challenging scenarios such as large head rotations. When we de-identify video sequences of the same person such as talking videos, maintaining the identity consistency between frames is important for video coherence.

5.3 Quantitative Evaluation

To the best of our knowledge, there are not yet universally acknowledged evaluation criteria for face de-identification. Based on relevant studies, we evaluate our approach in terms of both privacy protection effectiveness and image utility preservation.

(1) Privacy Protection Effectiveness: The determining factor for whether two images possess the same identity information is the distance of identity embedding. We respectively utilized various state-of-the-art tools including ArcFace [5], Face Recognition Library (FR)[2], FaceNet [29] to calculate the **identity distance** between de-identification results and the original images.

[2]Code available at https://github.com/ageitgey/face_recognition

Table 1: Privacy protection effectiveness evaluation and comparison with other methods. The red one represents the best and the blue one indicates the second.

Id-distance	ArcFace↓	FR↑	FaceNet↑	
			VGGFace2	CASIA
AMT-GAN [12]	0.668	0.529	0.935	0.914
DeepPrivacy [13]	0.555	0.764	1.215	1.113
CIAGAN [21]	0.475	0.723	1.152	1.037
Gu et al. [10]	0.592	0.851	1.232	1.179
IdentityDP [34]	0.421	0.793	1.246	1.218
Ours	0.562	0.872	1.259	1.235

Table 2: Identity consistency evaluation and comparison with other methods. Contrary to Table 1, the higher similarity among a group of images, the greater level of consistency.

Id-consistency	ArcFace↑	FR↓	FaceNet↓	
			VGGFace2	CASIA
AMT-GAN [12]	0.716	0.351	0.573	0.582
DeepPrivacy [13]	0.606	0.526	0.926	0.898
CIAGAN [21]	0.708	0.434	0.774	0.763
Gu et al. [10]	0.675	0.473	0.837	0.809
IdentityDP [34]	0.409	0.632	1.013	1.107
Ours	0.765	0.410	0.462	0.494

(2) Image Utility Preservation: To measure the utility of computer vision tasks, we define **face detectability (*FD*)** as the proportion of de-identified faces that can still be detected by a face detector. Additionally, we detect the face region to determine the **pixel-level difference (*PD*)** from the original image.

Furthermore, we measure the similarity of de-identification results to the original using several metrics. We use **PSNR** (peak signal-to-noise ratio) and **SSIM** (structure similarity) to measure image similarity at the pixel level. Since these indicators primarily focus on objective image quality, we incorporated **LPIPS** (Learned perceptual image patch similarity) distance to measure visual similarity, which has been demonstrated to be more correlated with human perceptual similarity than traditional metrics.

We randomly selected 500 images from both the synthesized datasets and FFHQ datasets for testing, and the comparison of privacy protection effectiveness is shown in Table 1. ArcFace calculates the cosine similarity between identity embedding while all others are Euclidean distance. In comparison, AMT-GAN [12] offers less protection for identity. Compared with the de-identification methods based on entire face synthesis like DeepPrivacy [13], Gu et al. [10], IdentityDP [34] and our approach are more specific to identity features and thus can achieve better protection.

We also compute the identity distance between the results of multi-view image de-identification. We randomly select 1,000 images from the synthesized datasets, corresponding to 125 identities and 8 different viewpoints for each identity (including at least one frontal image). The identity distance is between multi-view results and the frontal de-identified result, where a closer distance indicates a better consistency. The results shown in Table 2 prove that our method can achieve effective de-identification while preserving identity consistency in multi-view de-identification.

Table 3: Utility comparison under different metrics.

	FD ↑	*PD* ↓	PSNR ↑	SSIM ↑	LPIPS ↓
AMT-GAN [12]	0.984	2.345	19.413	0.752	0.268
DeepPrivacy [13]	0.998	3.125	21.503	0.775	0.391
CIAGAN [21]	0.975	4.396	17.236	0.512	0.469
Gu et al. [10]	0.917	3.148	18.526	0.677	0.432
IdentityDP [34]	0.942	1.842	23.664	0.822	0.285
Ours	0.996	2.014	21.236	0.718	0.311

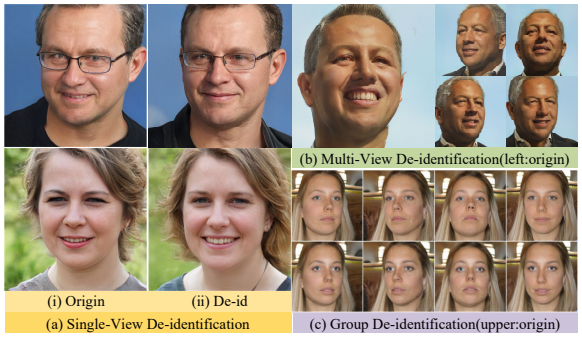

Figure 8: De-identification results using EG3D [2].

The utility evaluation results are shown in Table 3. Although IdentityDP [34] can retain a greater degree of similarity between the original and de-identified images, it may not be as effective in generating certain images where facial features cannot be accurately detected. Upon further investigation, we found out that most of these failed results were for faces with large side angles. Additionally, we analyzed that the reason for the lower similarity in our approach mainly lies in the changes in the background. In order to satisfy the naturalness of multi-view generation, we pay more attention to face area and there exists randomness in rendering.

6 CONCLUSION

In this paper, we propose an advanced framework for face de-identification, leveraging 3D-aware generative models, that offers superior performance across diverse applications. In *Single-View De-identification*, it can generate corresponding results based on the input image. In *Multi-View De-identification*, it can generate multi-view results with a single input image. In *Group De-identification*, a series of anonymized results can be generated for a set of input images. Through extensive experiments, we demonstrate that our approach is highly effective in protecting privacy while maintaining multi-view identity consistency. Furthermore, our framework is highly adaptable, with the flow capable of being applied to different 3D-aware generators (such as EG3D in Figure 8). Our approach has the potential to extend the application of de-identification and has important implications for privacy issues such as avatar generation and video sequence processing. Overall, our framework represents a significant advancement in the field, offering unparalleled performance and versatility for various applications.

ACKNOWLEDGMENTS

This work was supported by the Fundamental Research Funds for the Central Universities, STCSM under Grant 22DZ2229005, 111 project BP0719010.

REFERENCES

[1] Jingyi Cao, Bo Liu, Yunqian Wen, Rong Xie, and Li Song. 2021. Personalized and invertible face de-identification by disentangled identity information manipulation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 3334–3342.

[2] Eric Chan, Connor Z. Lin, Matthew A. Chan, Koki Nagano, Boxiao Pan, Shalini De Mello, Orazio Gallo, Leonidas J. Guibas, Jonathan Tremblay, S. Khamis, Tero Karras, and Gordon Wetzstein. 2021. Efficient Geometry-aware 3D Generative Adversarial Networks. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2021), 16102–16112.

[3] Eric R Chan, Marco Monteiro, Petr Kellnhofer, Jiajun Wu, and Gordon Wetzstein. 2021. pi-gan: Periodic implicit generative adversarial networks for 3d-aware image synthesis. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 5799–5809.

[4] Bo Dai and Dahua Lin. 2017. Contrastive Learning for Image Captioning. In *NIPS*.

[5] Jiankang Deng, J. Guo, and Stefanos Zafeiriou. 2018. ArcFace: Additive Angular Margin Loss for Deep Face Recognition. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2018), 4685–4694.

[6] Yu Deng, Jiaolong Yang, Jianfeng Xiang, and Xin Tong. 2022. Gram: Generative radiance manifolds for 3d-aware image generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 10673–10683.

[7] S Galanakis, Baris Gecer, Alexandros Lattas, and Stefanos Zafeiriou. 2022. 3DMM-RF: Convolutional Radiance Fields for 3D Face Modeling. *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)* (2022), 3525–3536.

[8] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron C. Courville, and Yoshua Bengio. 2014. Generative Adversarial Nets. In *NIPS*.

[9] Jiatao Gu, Lingjie Liu, Peng Wang, and Christian Theobalt. 2022. StyleNeRF: A Style-based 3D Aware Generator for High-resolution Image Synthesis. In *Tenth International Conference on Learning Representations*. OpenReview. net, 1–25.

[10] Xiuye Gu, Weixin Luo, Michael S. Ryoo, and Yong Jae Lee. 2019. Password-conditioned Anonymization and Deanonymization with Face Identity Transformers. In *European Conference on Computer Vision*.

[11] Yang Hong, Bo Peng, Haiyao Xiao, Ligang Liu, and Juyong Zhang. 2021. HeadNeRF: A Realtime NeRF-based Parametric Head Model. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2021), 20342–20352.

[12] Shengshan Hu, Xiaogeng Liu, Yechao Zhang, Minghui Li, Leo Yu Zhang, Hai Jin, and Libing Wu. 2022. Protecting Facial Privacy: Generating Adversarial Identity Masks via Style-robust Makeup Transfer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 15014–15023.

[13] Håkon Hukkelås, Rudolf Mester, and Frank Lindseth. 2019. Deepprivacy: A generative adversarial network for face anonymization. In *Advances in Visual Computing: 14th International Symposium on Visual Computing, ISVC 2019, Lake Tahoe, NV, USA, October 7–9, 2019, Proceedings, Part I 14*. Springer, 565–578.

[14] Yonghyun Jeong, Jooyoung Choi, Sungwon Kim, Youngmin Ro, Tae-Hyun Oh, Doyeon Kim, Heonseok Ha, and Sungroh Yoon. 2021. FICGAN: facial identity controllable GAN for de-identification. *arXiv preprint arXiv:2110.00740* (2021).

[15] Minguk Kang and Jaesik Park. 2020. ContraGAN: Contrastive Learning for Conditional Image Generation. In *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (Eds.), Vol. 33. Curran Associates, Inc., 21357–21369. https://proceedings.neurips.cc/paper/2020/file/f490c742cd8318b8ee6dca10af2a163f-Paper.pdf

[16] Tero Karras, Samuli Laine, and Timo Aila. 2019. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 4401–4410.

[17] Zhenzhong Kuang, Huigui Liu, Jun Yu, Aikui Tian, Lei Wang, Jianping Fan, and Noboru Babaguchi. 2021. Effective de-identification generative adversarial network for face anonymization. In *Proceedings of the 29th ACM International Conference on Multimedia*. 3182–3191.

[18] Jeong-gi Kwak, Yuanming Li, Dongsik Yoon, Donghyeon Kim, David Han, and Hanseok Ko. 2022. Injecting 3D Perception of Controllable NeRF-GAN into StyleGAN for Editable Portrait Image Synthesis. In *Computer Vision – ECCV 2022*, Shai Avidan, Gabriel Brostow, Moustapha Cissé, Giovanni Maria Farinella, and Tal Hassner (Eds.). Springer Nature Switzerland, Cham, 236–253.

[19] Yongxiang Li, Qianwen Lu, Qingchuan Tao, Xingbo Zhao, and Yanmei Yu. 2021. SF-GAN: Face De-Identification Method Without Losing Facial Attribute Information. *IEEE Signal Processing Letters* 28 (2021), 1345–1349. https://doi.org/10.1109/LSP.2021.3067517

[20] Tianxiang Ma, Dongze Li, Wei Wang, and Jing Dong. 2021. CFA-Net: Controllable Face Anonymization Network with Identity Representation Manipulation. *arXiv preprint arXiv:2105.11137* (2021).

[21] Maxim Maximov, Ismail Elezi, and Laura Leal-Taixé. 2020. Ciagan: Conditional identity anonymization generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 5447–5456.

[22] Lars M. Mescheder, Andreas Geiger, and Sebastian Nowozin. 2018. Which Training Methods for GANs do actually Converge?. In *International Conference on Machine Learning*.

[23] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. 2020. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. In *European Conference on Computer Vision*.

[24] Thu Nguyen-Phuoc, Chuan Li, Lucas Theis, Christian Richardt, and Yong-Liang Yang. 2019. Hologan: Unsupervised learning of 3d representations from natural images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 7588–7597.

[25] Michael Niemeyer and Andreas Geiger. 2021. Giraffe: Representing scenes as compositional generative neural feature fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 11453–11464.

[26] Roy OrEl, Xuan Luo, Mengyi Shan, Eli Shechtman, Jeong Joon Park, and Ira Kemelmacher-Shlizerman. 2022. StyleSDF: High-Resolution 3D-Consistent Image and Geometry Generation. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 13493–13503. https://doi.org/10.1109/CVPR52688.2022.01314

[27] Slobodan Ribaric, Aladdin Ariyaeeinia, and Nikola Pavesic. 2016. De-Identification for Privacy Protection in Multimedia Content. *Image Commun.* 47, C (sep 2016), 131–151. https://doi.org/10.1016/j.image.2016.05.020

[28] Elad Richardson, Yuval Alaluf, Or Patashnik, Yotam Nitzan, Yaniv Azar, Stav Shapiro, and Daniel Cohen-Or. 2021. Encoding in Style: A StyleGAN Encoder for Image-to-Image Translation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2287–2296.

[29] Florian Schroff, Dmitry Kalenichenko, and James Philbin. 2015. FaceNet: A unified embedding for face recognition and clustering. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 815–823. https://doi.org/10.1109/CVPR.2015.7298682

[30] Katja Schwarz, Yiyi Liao, Michael Niemeyer, and Andreas Geiger. 2020. Graf: Generative radiance fields for 3d-aware image synthesis. *Advances in Neural Information Processing Systems* 33 (2020), 20154–20166.

[31] Keqiang Sun, Shangzhe Wu, Zhaoyang Huang, Ning Zhang, Quan Wang, and Hongsheng Li. 2022. Controllable 3D Face Synthesis with Conditional Generative Occupancy Fields. *ArXiv* abs/2211.13251 (2022).

[32] Qianru Sun, Liqian Ma, Seong Joon Oh, Luc Van Gool, Bernt Schiele, and Mario Fritz. 2017. Natural and Effective Obfuscation by Head Inpainting. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2017), 5050–5059.

[33] Ayush Tewari, Mohamed Elgharib, Gaurav Bharaj, Florian Bernard, Hans-Peter Seidel, Patrick Pérez, Michael Zollhofer, and Christian Theobalt. 2020. Stylerig: Rigging stylegan for 3d control over portrait images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 6142–6151.

[34] Yunqian Wen, Bo Liu, Ming Ding, Rong Xie, and Li Song. 2022. Identitydp: Differential private identification protection for face images. *Neurocomputing* 501 (2022), 197–211.

[35] Yifan Wu, Fan Yang, Yong Xu, and Haibin Ling. 2019. Privacy-Protective-GAN for Privacy Preserving Face De-Identification. *Journal of Computer Science and Technology* 34 (2019), 47–60.

[36] Yuan Zhao, Bo Liu, Tianqing Zhu, Ming Ding, and Wanlei Zhou. 2022. Private-encoder: Enforcing privacy in latent space for human face images. *Concurrency and Computation: Practice and Experience* 34, 3 (2022), e6548.

[37] Peiye Zhuang, Liqian Ma, Oluwasanmi Koyejo, and Alexander Schwing. [n. d.]. Controllable Radiance Fields for Dynamic Face Synthesis. *3DV* ([n. d.]). https://par.nsf.gov/biblio/10387045

[38] P. Zhuang, L. Ma, O. Koyejo, and A. G. Schwing. 2022. Controllable Radiance Fields for Dynamic Face Synthesis. In *Proc. 3DV*.

[39] Yiyu Zhuang, Hao Zhu, Xusen Sun, and Xun Cao. 2022. MoFaNeRF: Morphable Facial Neural Radiance Field. In *European Conference on Computer Vision*.