ITC 2/55
Information Technology and Control
Vol. 55 / No. 2/ 2026
pp. 531-552
DOI 10.5755/j01.itc.55.2.42886

Generative Adversarial Network Facial Privacy Protection Model Combined with Facial Structure Modeling

Received 2025/06/26 | Accepted after revision 2025/12/23

HOW TO CITE: Sun, Z., Li, L. (2026) Generative Adversarial Network Facial Privacy Protection Model Combined with Facial Structure Modeling. *Information Technology and Control*, 55(2), 531-552. https://doi.org/10.5755/j01.itc.55.2.42886

Generative Adversarial Network Facial Privacy Protection Model Combined with Facial Structure Modeling

Zhiyong Sun
Information Network Center, The Second Affiliated Hospital of Heilongjiang University of Chinese Med-icine, Harbin, 150001, China

Liheng Li*
School of Medical Information Engineering, Heilongjiang University of Chinese Medicine, Harbin, 150001, China

Corresponding author: lilihenng@outlook.com

With the continuous development of intelligent vision technology, facial image anonymization plays an increasingly important role in public privacy protection and data compliance sharing. However, traditional methods have a contradiction between identity feature removal and image naturalness preservation, and are prone to distortion when dealing with semantic consistency issues such as posture. To this end, a structure identity separation guided facial anonymous generation method is constructed on the framework of generative adversarial networks. A structural perturbation module guided by differential privacy mechanisms is designed, and an attention guided feature weakening mechanism is introduced. Additionally, a style guided and structure preserving conditional generation network is integrated. In performance testing, the proposed method achieves identity consistency scores of 0.814 and 0.759 for frontal and lateral postures, respectively, with corresponding natural scores of 4.08 and 3.95. The inference delay, number of generated graphs per unit energy consumption, and peak memory of this method are 114.8 ms, 19.2 IpJ, and 1286 MB, respectively. The experimental results show that this method balances visual consistency and structural rationality while enhancing identity concealment, providing a feasible path and technical support for the secure generation and compliant application of private images.

KEYWORDS: Face anonymity; Differential privacy; Identity suppression; Generative adversarial networks; Structure modeling

1. Introduction

With the widespread popularity of social networks, video surveillance, and smart terminals, facial images have been deeply embedded in human daily life and social interactions, becoming an important medium for tasks such as identity authentication, behavior analysis, and human-computer interaction [17]. Additionally, the surge in image acquisition and storage behavior also raises high concerns about facial privacy breaches, especially in the context of rapid development of big data and artificial intelligence technology. How to effectively protect user identity security without affecting image availability has become one of the research hotspots in the fields of computer vision and information security [8, 19]. Compared with traditional methods such as occlusion and blurring, various privacy protection methods that integrate image synthesis, identity decoupling, and differential perturbation mechanisms have emerged in recent years, making initial progress in improving image quality and diversity control capabilities [13]. However, from the perspective of real-world scenarios and policy guidance, the demand for facial privacy protection has grown exponentially, with complaints of facial recognition misuse on social media platforms alone increasing by about 37% year-on-year. Unauthorized facial feature collection has become a potential source of risk in fields such as video surveillance, intelligent access control, and financial risk management. Meanwhile, existing anonymization methods still face issues such as insufficient anonymity strength, reversibility risks, and semantic distortion in practical deployment, making it difficult to meet the dual requirements of multi-source data sharing and model interpretability [10].

He et al. proposed a diffusion model method that integrates multi-scale image inversion and energy scheduling to address the problem of inconsistent goals and difficulty in unified modeling between anonymization and visual identity hiding tasks in facial images [5]. By constructing stable diffusion embeddings and energy functions, the unified modeling of the two privacy protection tasks was achieved [5]. Osorio Roig et al. pointed out that there are weak links in the attack modeling of existing privacy enhanced facial recognition technologies, and proposed an attack method based on facial attribute similarity, using black box models and small annotated databases for attack inference [18]. Wang et al. proposed an identity hider model that combines appearance transformation and style generation [26]. This method achieved a balance between identity hiding and usability by manipulating the generation of network latent space and attribute replacement [26]. Ghani et al. focused on the privacy leakage risk brought by facial recognition technology in intelligent systems and proposed a comprehensive privacy protection framework that combines Generative Adversarial Network (GAN), blockchain, and distributed computing [3]. The experimental results showed that the model achieved an accuracy of 93.84% on high-resolution images [3]. Sivalakshmi et al. designed a deep learning model NDLM for real-world scenarios [25]. This model integrated two stages of face detection and privacy protection, using a hybrid convolutional network and an improved Fire Eagle algorithm to improve face recognition accuracy while preserving service quality [25].

Some studies have further expanded the research path of facial privacy protection from multiple feature fusion, anti-counterfeiting mechanism optimization, personalized adversarial strategies, and attack robustness evaluation, and proposed more diverse and practical solutions. Xiao et al. were concerned about the privacy risks of facial images being abused by deep learning recognition systems on social platforms [28]. They proposed a protection scheme based on multi-feature fusion tensors, which combined the complementarity of deep features and manual features to improve camouflage effects while maintaining visual similarity [28]. Morris et al. pointed out that existing image privacy protection mechanisms ignore the problem of background personnel and location leakage in images [15]. Therefore, a location aware image access control system was proposed, which could recognize sensitive areas in social networks in real time and replace them with synthetic faces to prevent user privacy exposure [15]. Zhong et al. proposed a one person one mask universal protection method to address the privacy threat posed by malicious identification systems to ordinary users on social media [33]. This method constructed class masks for individuals in the feature subspace and was suit-

able for black box recognition models with different loss functions and network structures [33]. Gong et al. proposed a method for replacing and restoring variational autoencoders, which decoupled identity factors from identity independent attributes and optimized visual fusion through image captioning [4]. Rot et al. proposed a privacy detector framework to restore attribute information such as gender and age from privacy enhanced images, in response to the lack of restoration attack evaluation in existing soft biometric suppression methods [4].

In addition to privacy protection mechanisms, recent research has also focused on the deep modeling of facial expressions and emotional features. Wang et al. proposed a Federated Long Tail Learning Framework Based on Multiple Negotiation Calibration (FedMDD) [21]. Through the feature fusion and hierarchical calibration mechanism of multiple clients, it effectively alleviated the feature bias problem caused by uneven sample distribution, providing transferable ideas for face anonymization and feature consistency learning in cross domain scenarios [27]. Meena et al. used deep Convolutional Neural Networks (CNN) for multi sentiment recognition of facial expressions [14]. The model significantly improved its robustness to complex lighting and pose changes through multi-scale convolutional layers and residual feature mapping, providing a feature extraction basis for emotion preservation and semantic consistency in anonymization tasks [14]. Qi et al. proposed the Mapped Binary Patterns (MBP) based on cognitive mechanisms, which introduces feature mapping and perceptual weights on the basis of traditional Local Binary Patterns (LBP), achieving high-precision recognition of facial expressions and structural information encoding [20].

In summary, although the current facial image anonymization technology has achieved a certain balance between privacy protection and visual quality, there are still several key issues that need to be addressed urgently. Firstly, most methods lack fine-grained modeling of facial structural information, making it difficult to achieve high-quality anonymization while ensuring consistency between facial expressions and postures. Secondly, most existing identity removal strategies are based on global feature substitution and lack an explicit mechanism for weakening identity salience, which can lead to residual identity information and affect anonymity strength [2, 29, 7]. To this end, based on the traditional GAN framework, a structure identity decoupling guided anonymous face generation method (SID-GAF) is proposed, which constructs a two-stage anonymous generation framework with structural perturbation and identity saliency suppression as the core, providing effective technical support for trustworthy face image publishing and privacy protection.

Compared with mainstream diffusion models and Variational Autoencoder (VAE) based anonymization methods, the innovation of SID-GAF lies in its redefinition of the concept of "structurally controllable anonymization". The diffusion model achieves privacy enhancement by gradually adding and removing noise, but lacks explicit constraints on the consistency of facial geometry and posture. Although the VAE method can partially decouple identity and semantics in the latent space, the latent variables lack interpretability and are difficult to accurately control the anonymity strength at the specific structural level. In contrast, SID-GAF adopts a collaborative mechanism of "structural modeling+identity saliency suppression+differential privacy perturbation", which explicitly separates structural and identity features during the generation process, making the anonymous process controllable and interpretable.

The innovation of the research lies in: Firstly, it designs a geometric structure perturbation method incorporating differential privacy-inspired noise injection to enhance the structural protection strength and control ability of anonymous images. Secondly, it introduces a mechanism to suppress identity salience and guide the network to explicitly weaken identity features. Thirdly, it integrates multi-source style information and structural features to guide the generation network to achieve natural and personalized anonymous face synthesis. Compared to the prevailing mainstream methods, the proposed approach stands out with a fundamental distinction. Rather than merely attaining privacy preservation through appearance style transformation, it also introduces perturbations and reconstructs facial geometric relationships at the structural level. This is achieved by harmoniously integrating "structural modeling + identity suppression + privacy-inspired perturbation," thereby enabling controllable levels

of anonymity while upholding the naturalness of the image. This difference makes the model significantly superior to previous methods in terms of semantic consistency, posture preservation, and cross scene robustness, and has higher theoretical interpretability and generality.

2. Methods and Materials

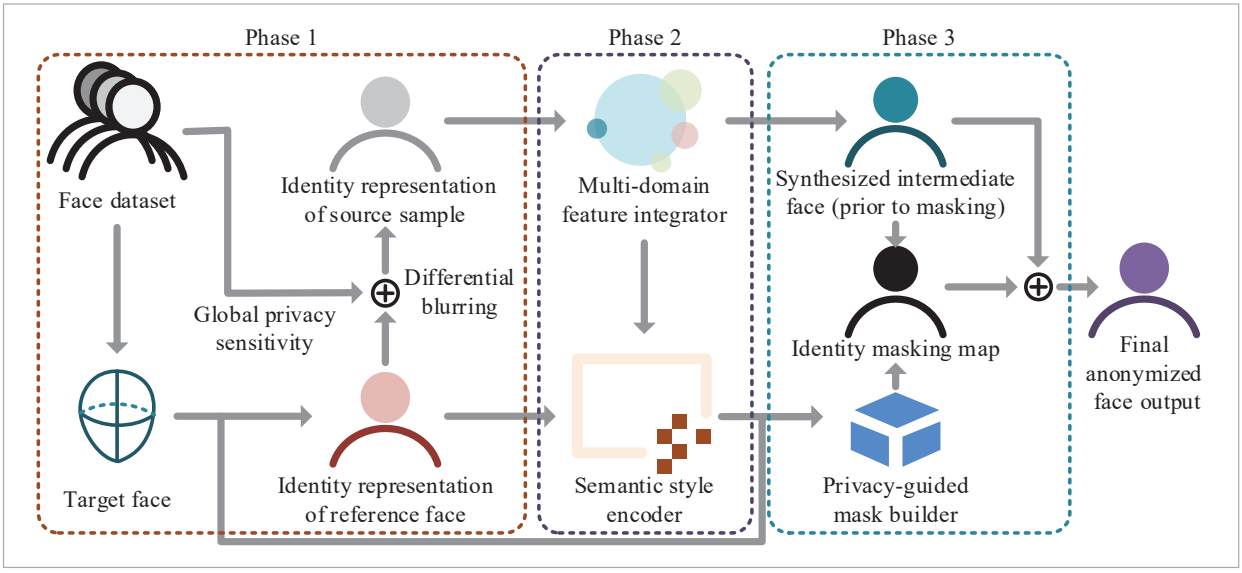
To achieve controllable and privacy preserving facial anonymization processing, a structure identity separation guided facial anonymization generation method SID-GAF was developed. Firstly, a geometric structure representation with anonymization characteristics was constructed through differential perturbation to achieve structural level privacy protection. Subsequently, an identity saliency suppression module and a style guided generation network were introduced to synergistically weaken identity information and improve visual quality.

2.1. Facial Structure Driven Facial Anonymous Generation Method

To improve the identity privacy protection effect in the process of facial image anonymization, while preserving the non identity visual attributes of the original image as much as possible, a facial structure oriented anonymous image generation method is proposed. This method takes geometric structure modeling as the core, combines differential privacy mechanism to perturb structural information, and synthesizes anonymous images with realism and diversity through style encoding and feature fusion to enhance identity unidentifiable while maintaining image naturalness. The overall model structure consists of three parts, and the specific process is shown in Figure 1. Among them, the "geometric structure" mentioned is not a traditional facial template, but a high-dimensional geometric representation vector composed of facial keypoint coordinates and their spatial topological relationships, used to describe facial proportions, contour shapes, and local structural constraints. The structural information is automatically extracted by the keypoint detection model and can be formally represented as a set of structural vectors containing coordinates and shape descriptors. Its function is to provide controllable structural constraint inputs for subsequent anonymous generation, enabling the model to reconstruct anonymous faces with consistent expressions and poses without relying on specific appearance textures.

As shown in Figure 1, the method can be divided into three stages. In the first stage, the input image is structurally modeled and differentially perturbed to generate a geometric representation with privacy

Figure 1
Overall architecture of anonymous face generation method for structural modeling.



preserving properties. In the second stage, a semantic style encoder is used to extract the appearance features of the image and integrate them with anonymous structural information in a multi domain feature integrator. In the final stage, a privacy guided mask builder is used to synthesize anonymous facial images, and the authenticity and identity removal effect of the images are further optimized through adversarial discrimination mechanisms.

In the first stage, a geometric representation vector R_i of the face is constructed using keypoint detection and image modeling techniques to describe the facial features arrangement and contour shape in the image. To enhance privacy protection without compromising availability, Laplace mechanism is applied to differential perturbation of structural feature R_i to generate anonymous structural representation $\tilde{R}_i$. In the standard Laplace mechanism of differential privacy, the perturbation is calibrated such that the difference in output distribution between any two adjacent inputs is controlled by the privacy budget parameter ε, as expressed in Equation (1) [31].

$$\tilde{R}_i = R_i + Laplace(0, \Delta / \varepsilon). \quad (1)$$

In Equation (1), Δ represents the global sensitivity of structural features. The smaller ε, the stronger the disturbance, and the higher the anonymity. The processed structural feature $\tilde{R}_i$ are used as structural constraint inputs for subsequent generation modules to guide the structural consistency reconstruction of anonymous images.

In the second stage, the system synthesizes anonymous images based on the perturbed structural feature $\tilde{R}_i$ and input image X_r through a conditional generation model. This process relies on a collaborative mechanism of structure driven and style guided to construct an end-to-end anonymous image generation network. The network consists of four modules, namely semantic style encoder, multi-domain feature integrator, privacy guided mask builder, and discriminator, each responsible for style extraction, identity removal, structure reconstruction, and authenticity evaluation tasks. Its overall modular design is shown in Figure 2.

As shown in Figure 2, the semantic style encoder takes the original image as input, relies on multiple layers of ResBlocks to extract global semantic and style features, and obtains style vector representations through fully connected layers for subsequent conditional guidance. The multi-domain feature integrator combines perturbed structural features and style vectors, and fuses semantic and spatial information at different levels through skip connections to achieve suppression and structural preservation of identity features. Subsequently, the privacy-guid-

Figure 2

Modular structure diagram of anonymous generation network.

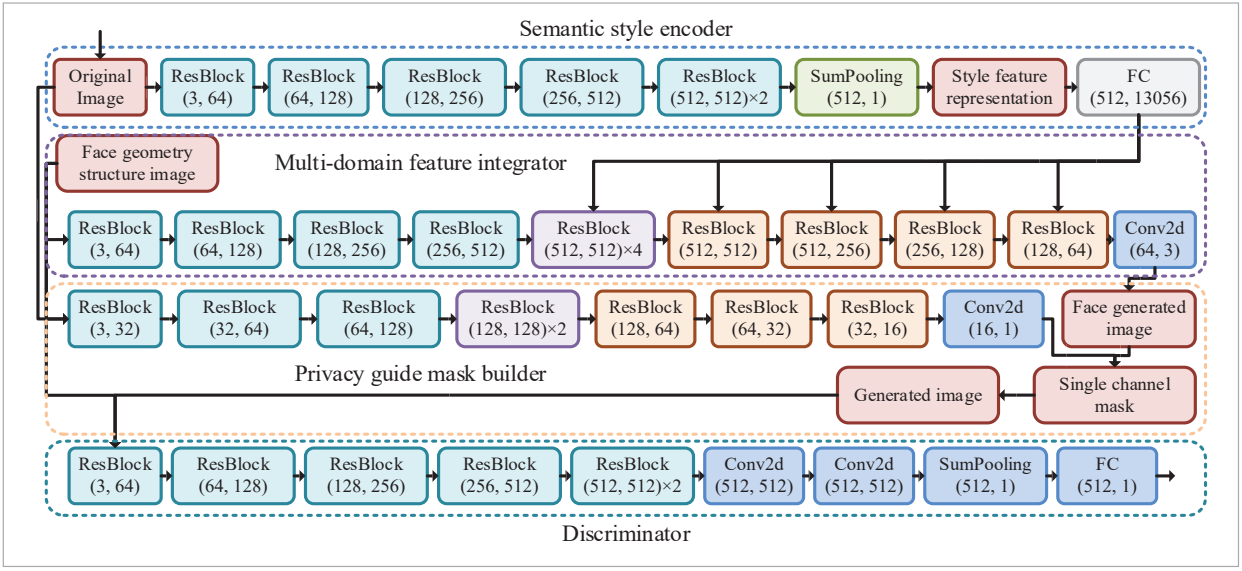
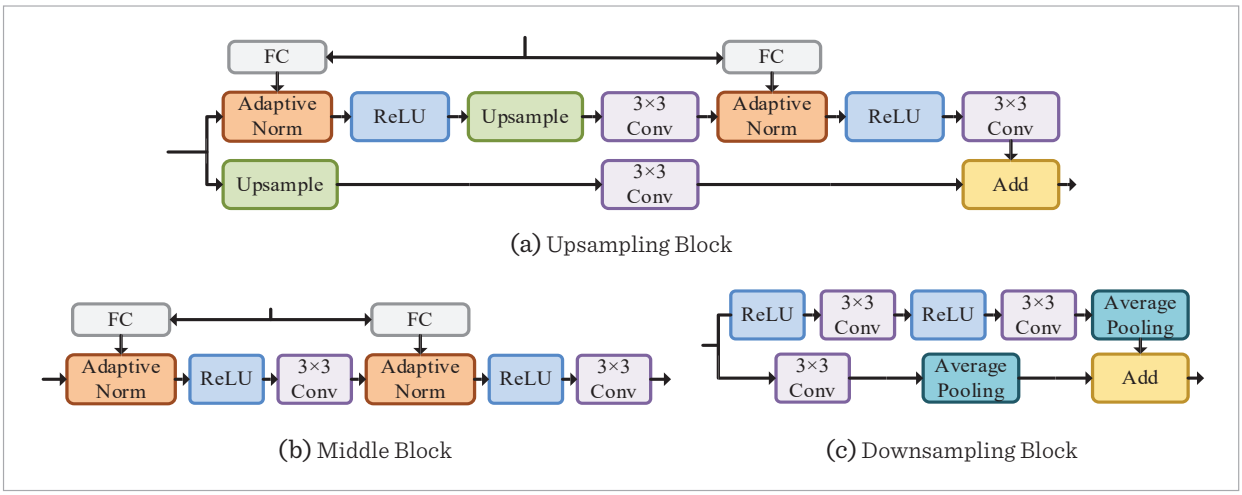


Figure 3
The same type of residual unit structure.



ed mask builder constructs facial structural perturbation regions based on local perception, outputs binary mask images through convolution and upsampling operations, and reconstructs anonymous images under the guidance of this mask. Finally, the discriminator simultaneously receives the original image and the synthesized result, and determines whether the facial image output by the network has authenticity or unidentifiable through feature comparison and classification.

Furthermore, to achieve multi-scale modeling of semantics, structure, and local style in each submodule, various residual structural units are widely used within the network for feature representation and contextual modeling. According to their different positions and functions, residual blocks can be divided into three categories: downsampling blocks for spatial compression, intermediate blocks for semantic enhancement, and upsampling blocks for image restoration, as shown in Figure 3.

Figures 3 (a)-(c) are used for upsampling, feature transformation, and downsampling processes, respectively. The upsampling block improves spatial restoration accuracy through interpolation and normalization, and introduces residual connections to alleviate gradient vanishing. It also introduces fully connected layers and adaptive normalization in the middle block enhances nonlinear modeling capabilities. The downsampling block combines convolution operation and average pooling to extract local semantics. To enhance the adaptive expression ability of the fusion module for style features, Adaptive Instance Normalization (AdaIN) is introduced, whose mathematical form is shown in Equation (2) [9, 16].

$$AdaIN(x,y)=\sigma(y)\cdot\left(\frac{x-\mu(x)}{\sigma(x)}\right)+\mu(y). \tag{2}$$

In Equation (2), x represents the style feature map, γ is the conditional vector, $\mu(\cdot)$ and $\sigma(\cdot)$ represent the channel dimension mean and standard deviation. This mechanism standardizes and reconstructs input features, allowing the generator to dynamically adjust the distribution of feature maps based on different style features, thereby enhancing the ability to control differences between samples.

In the third stage, the model introduces a differential privacy mechanism for structural information related to identity features in facial images, and constructs a structural level anonymous processing flow. Firstly, using a structure detection function $f(\cdot)$, a facial structure vector is obtained from the input image m-dimension, which includes geometric quantities such as eyebrow width, eye distance, and nasal wing width. To calibrate the structural noise following the Laplace mechanism in differential privacy, it is necessary to evaluate the maximum sensitivity of different local structural vectors. The definition of global sensitivity is shown in Equation (3).

$$\Delta f = \max_{x,x' \in U} \left\| f(x) - f(x') \right\|_1 . \quad (3)$$

In Equation (3), U represents the image dataset and $\|\cdot\|_1$ represents the norm L_1. Subsequently, structural anonymization is achieved by adding Laplacian noise to the structural vector, with the perturbation form shown in Equation (4).

$$\tilde{R} = f(X) + Lap\left(\frac{\Delta f}{\varepsilon}\right). \quad (4)$$

In Equation (4), $\tilde{R}$ is the anonymous structural vector, ε is the privacy budget, and controls the noise intensity. $Lap(\cdot)$ is the Laplace distribution with zero mean and scale parameter $\frac{\Delta f}{\varepsilon}$. Under the classical Laplace mechanism of differential privacy, the following Equation (5) characterizes the relationship between the noise scale and the global sensitivity.

$$P_r[\tilde{R} \mid f(X)] \le e^{\varepsilon} \cdot P_r[\tilde{R} \mid f(X')]. \quad (5)$$

Equation (5) can be expanded to obtain Equation (6).

$$P_r[\tilde{R} \mid f(X)] = \prod_{i=1}^{m} exp\left(- \frac{\varepsilon \cdot \left| R_{(i)} - f(X)_i \right|}{\Delta f} \right). \quad (6)$$

By applying this mechanism independently to each structural dimension, the perturbation follows the same per-coordinate privacy constraint at the level of facial geometry, and limits the influence of any single structural change on the perturbed representation. The anonymous structure vector $\tilde{R}$ after disturbance can jointly drive the generation process of anonymous face images with the style encoding vector of the reference image. The mapping function is defined in Equation (7).

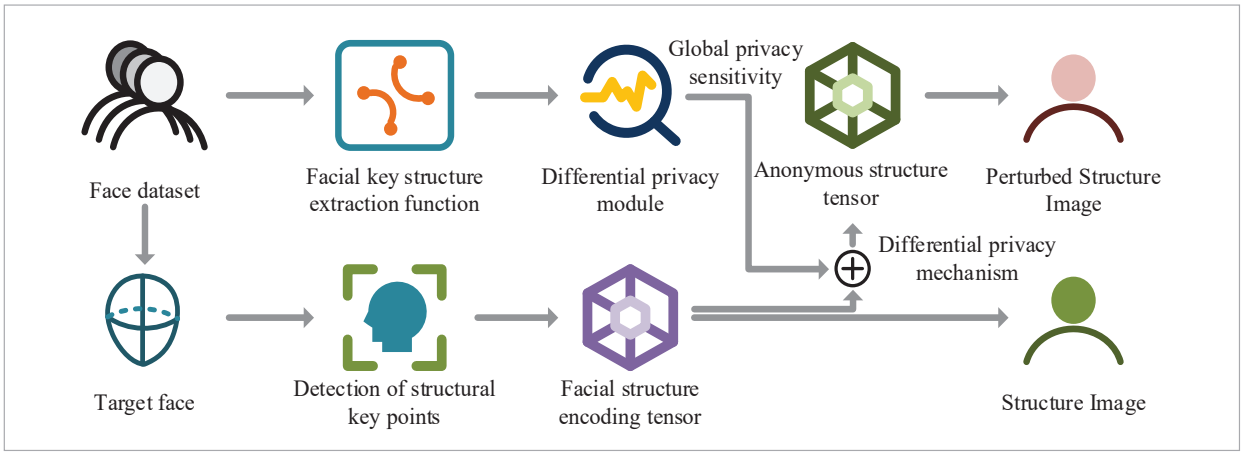
$$X_f = h(\tilde{R}, R_s(X_T)). \quad (7)$$

In Equation (7), $h(\cdot)$ is the conditional generator. X_f is the output anonymous image, which maintains the same style as the reference image, but has been desensitized in terms of structural features. The detailed steps for anonymous processing are shown in Figure 4.

Figure 4 shows the modules of structural keypoint detection, privacy sensitivity estimation, Laplace perturbation, and anonymous structural mapping. After extracting the structural vector from the input image and linking it with the noise module, the generated anonymous structure effectively shields the highly recognizable structural dimensions while retaining some geometric consistency, providing structural protection constraints for the generation stage and further enhancing the controllability and security of anonymous images. It is worth noting that, in this work, the term "differential privacy-based" only refers to the design of local structural perturbations via Laplace and exponential

Figure 4

Flowchart of the facial structure anonymous perturbation protection mechanism.



mechanisms. The identity encoder and generator are trained in a standard (non-DP) manner without DP-SGD or compositional analysis.

2.2. Facial Anonymity Enhancement Method with Deactivation Control

The method of combining facial geometric structure modeling with differential privacy perturbation mechanism can achieve anonymization of structural features. However, in some complex scenarios, there may still be weak identity information residue, which affects the overall anonymity strength. To further enhance the ability to suppress individual identity features and improve the semantic consistency of generated images in dimensions such as expression and posture, a structure identity separation guided face anonymous generation method SID-GAF is proposed based on this. It introduces a face generation network with identity saliency suppression mechanism and structure preservation, and models individual identifiable factors at multiple scales. At the same time, by using style guided branches to enhance the realism and consistency of generated images, while balancing anonymity and visual quality, more diverse and controllable image generation can be achieved while maintaining the above structural perturbation strategy inspired by differential privacy. The system framework diagram of this method is shown in Figure 5.

As shown in Figure 5, the system consists of an identity encoder, a structural perturbation module, and a style guidance generator. The identity encoder extracts potential identity features through channel attention mechanism and combines them with classification tasks to locate salient regions, and then uses deactivation projection to eliminate identity information. The structural perturbation module adds noise to facial keypoint vectors within a differential-privacy-inspired framework, generating anonymous structural maps to replace the original geometric input and effectively masking sensitive features. The style guided generator integrates identity representation, anonymous structure, and target style encoding to output personalized anonymous images that preserve expressions and postures, enhancing naturalness and background consistency.

To weaken the discriminative ability of identity sensitive areas in images, an identity saliency suppression mechanism based on Gradient weighted Class Activation Mapping (Grad-CAM) is designed. The identity attention map is generated through gradient backpropagation, which explicitly suppressed the feature response of specific regions, thereby achieving the weakening and anonymization of identity related information [22, 24]. To visually demonstrate the processing flow, Figure 6 shows the operational flowchart of the identity saliency suppression mechanism. To ensure that the Grad-CAM-based identity suppression does not bias the evaluation outcome, the identity

Figure 5
System architecture of SID-GAF model.

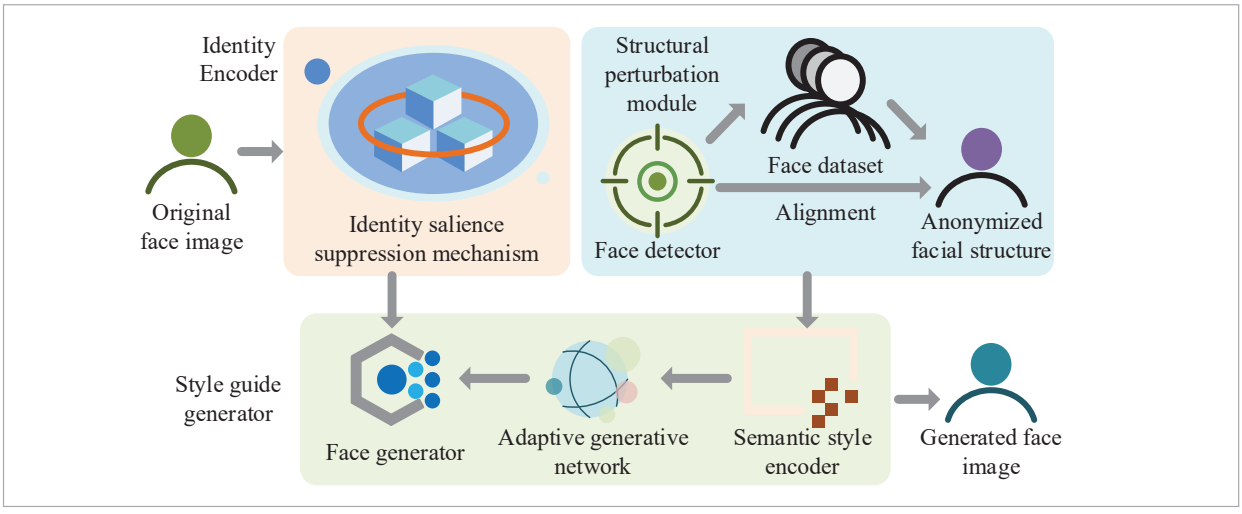
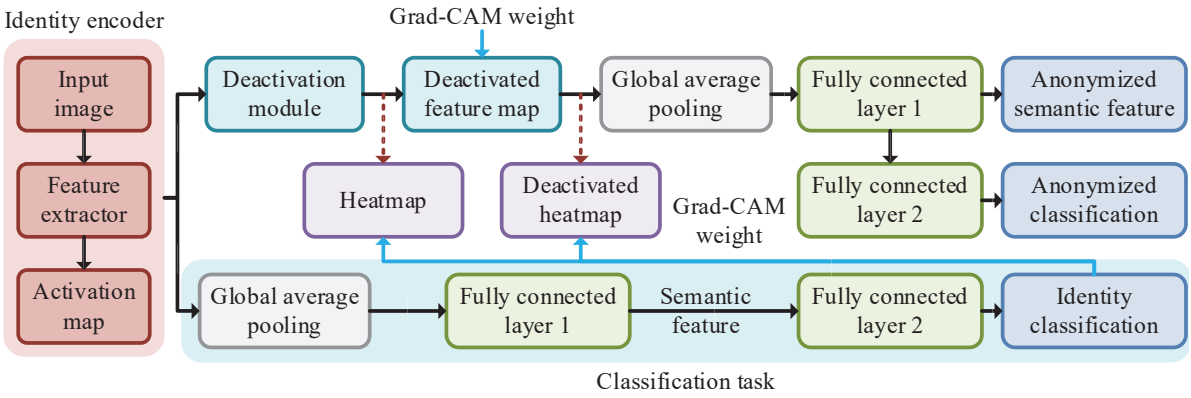


Figure 6
Workflow of identity suppression through attention mapping.



classifier used to compute the saliency map was independently trained using a ResNet-50 backbone on identity labels from the training split of CelebA-HQ. It does not share parameters or samples with the re-identification network employed for Attack Success Rate (ASR) and Re-identification Failure Rate (RFR) evaluation, thereby preventing self-evaluation effects and ensuring adversarial independence.

Figure 6 shows the process of extracting initial feature A from the input image, calculating channel weight a^k through gradient calculation, constructing the attention map M^k, and generating final anonymous feature $\hat{A}$ using a perturbation function. Assuming the input image is I, the feature map extracted by the convolutional network is $A \in R^{C\times H\times W}$, where C represents the number of channels, and H and W are the height and width of the feature map, respectively. To quantify the importance of each channel in a certain category k, a gradient weighting strategy is used to define its response weight, as shown in Equation (8).

$$a_c^k = \frac{1}{Z}\sum_{i=1}^{H}\sum_{j=1}^{W}\frac{\partial y^k}{\partial A_{c,i,j}}. \tag{8}$$

In Equation (8), a_c^k represents the influence weight of the c-th channel on the category k, y^k is the classifier's prediction score for the category k, and $Z = H \times W$ is the normalization coefficient. This weight is used to weight and sum up each channel, and obtain the final attention map through the ReLU function, as shown in Equation (9).

$$M^k = \mathrm{ReLU}\left(\sum_{c=1}^{C} a_c^k A^c\right). \tag{9}$$

In Equation (9), M^k can be regarded as the most representative region of the category k in the image, used to guide the feature deactivation process. To further complete the feature weakening process, a perturbation function is introduced to adjust the feature map A, as shown in Equation (10).

$$A' = A + a_c^k \otimes y_c^k. \tag{10}$$

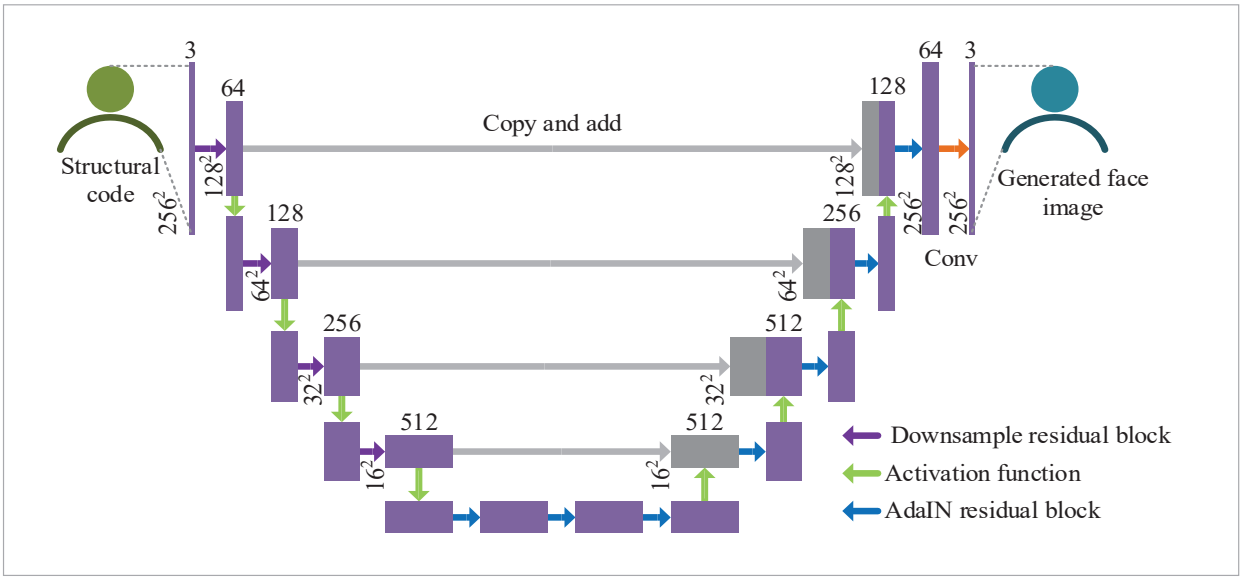
In Equation (10), y_c^k is the learnable offset term, and the symbol $\otimes$ represents the element wise multiplication operation. Considering that masking processing may affect the global discriminative ability, further regulatory factor γ is introduced to define the perturbation features after fusing multiple types of responses, as shown in Equation (11).

$$A = A + \gamma\sum_{k=1}^{K}(a^k)^{\top}\Phi_k a^k \tag{11}$$

In Equation (11), Φ_k is the weight mapping matrix of the category k, $\hat{A}$ represents the anonymized feature map after deactivation adjustment. Finally, anonymous feature $\hat{A}$ is input into the feature extraction layer and fully connected layer FC1, generating identity weakening representation vectors for

Figure 7

Design of adaptive generation network with style and structure fusion.



subsequent synthesis or judgment of facial images. Furthermore, the anonymous face synthesis of the model is implemented by an adaptive generative network. To enhance the consistency of the model in background transfer and local structure, the generator adopts an improved U-Network (U-Net) structure and integrates facial structural features and identity style information to improve expression ability. The network structure is shown in Figure 7.

As shown in Figure 7, the generator consists of two input branches. A receiver structure encoding and background feature Z_b are used to extract local geometric information through residual downsampling. The other method concatenates identity Z_{id} and style vector Z_s, and generates affine parameters through a fully connected layer. The two branch features are fused through the AdaIN module during the upsampling stage, which controls the consistency of the generated style and preserves the facial contours and background information of the original image with the assistance of structural branches.

Furthermore, to enhance the anonymization effect of facial geometry and improve the privacy protection capability of the method, an exponential mechanism inspired by differential privacy is adopted to replace the traditional direct replacement of geometric labels. By constructing a structural similarity measurement function and probability sampling strategy, identity sensitive information can be protected by utilizing noise disturbance while ensuring structural similarity [23]. Firstly, the geometric structure in the original image is represented as s, which includes a set of facial keypoints, such as eye, eyebrow, mouth, and other facial contours. To construct a geometric structure database E from all candidate structural samples $\{x_i\}_{i=1}^{N}$, it is necessary to select the most suitable anonymous alternative from this set. To measure the similarity between s and each sample x_i, a normalized distance function is designed as shown in Equation (12).

$$u(s,x_i)=\frac{\|s-x_i\|_2-\min\limits_{x_f\in E}\|s-x_f\|_2}{\max\limits_{x_f\in E}\|s-x_f\|_2-\min\limits_{x_f\in E}\|s-x_f\|_2}. \tag{12}$$

In Equation (12), $\|\cdot\|_2$ represents the Euclidean distance. This function compresses all structural distances into the [0,1] interval to measure the similarity between each candidate structure and the target structure.

Next, to implement the structural replacement process of differential privacy mechanism, an exponential sampling strategy based on this scoring function is constructed. If the privacy budget parameter ε is set, the probability of each sample $x_i \in E$ being sam-

pled as a replacement result is defined as shown in Equation (13).

$$P(x_i) = \frac{\exp(-\varepsilon \cdot u(s, x_i) / 2\Delta)}{\sum_{x_f \in E} \exp(-\varepsilon \cdot u(s, x_j) / 2\Delta)}. \quad (13)$$

In Equation (13), Δ represents the sensitivity boundary of the function u, which is used to control the maximum fluctuation range of similarity score perturbation. Higher value of ε indicates an increased tendency towards similar structures, while smaller value of ε enhances anonymity and increases uncertainty in structure selection, thereby improving the overall level of privacy protection. The selected geometric structure will replace the original structure, and combined with the facial expression and posture features in the original image, a highly privacy preserving facial geometric structure image will be reconstructed.

Finally, a comprehensive multi-objective loss function is designed to achieve coordinated optimization of identity anonymity, image fidelity, and structural rationality. The total loss is expressed as Equation (14).

$$L_{total} = \lambda_1 L_{adv} + \lambda_2 L_{id} + \lambda_3 L_{geo} + \lambda_4 L_{rec} + \lambda_5 L_{feat} + \lambda_6 L_{pseudo}. \quad (14)$$

In Equation (14), L_{adv} is the counteractive loss, which guides the generated image to be more realistic. L_{id} is the identity suppression loss, Kullback Leibler Divergence (KL) is used tow constrain the classifier output to tend towards a uniform distribution, to weaken the distinguish ability of real identities. L_{geo} is a geometric consistency item, which limits the offset between the anonymous structure and the original structure in the dimensions of expression and posture. L_{rec} is the image reconstruction item, and the fidelity between the composite image and the input image is constrained by the pixel space. L_{feat} is the loss of feature compatibility, and features are used to maintain the harmony between image style and content. L_{pseudo} is the loss of pseudo identity confrontation, and an auxiliary discriminator is introduced to improve the consistency and controllability of pseudo identity features. Weights λ_1-λ_6 are used to balance the objectives of each subtask.

3. Results

3.1. SID-GAF Face Anonymization Basic Performance Test

The experiment was conducted on a facial image generation platform built on Python and PyTorch, with a hardware environment including an Intel Core i9-12900K processor, 64 GB memory, Ubuntu 20.04, and NVIDIA RTX 4090 GPU for model training and inference. The overall architecture of the system was based on the proposed SID-GAF method, consisting of an identity deactivation module, a differential perturbation structure modeling unit, and a style-guided generator. Three publicly available facial datasets were adopted for different evaluation perspectives: CelebA-HQ to assess structural restoration and visual fidelity under standard imaging conditions; FaceForensics++ to evaluate anonymity robustness in forgery–real mixed contexts; and WIDER Face, covering significant variations in race, age, pose, illumination, and occlusion, to verify cross-domain generalization performance.

To ensure methodological clarity and reproducibility, the datasets were divided using a 70/10/20 train–validation–test split, with all identities in the test set strictly isolated from those used during model training and structural perturbation processing. Unless otherwise specified, a representative subset covering 250 disjoint identities was used for evaluation, drawn from approximately 8,000 anonymized samples to ensure fairness and consistency across experiments. Re-identification-related metrics were computed using a pretrained ArcFace (ResNet-100) model that remained entirely independent from SID-GAF. Cosine similarity was applied as the matching function with a decision threshold of 0.3. The identity encoder and style encoder follow a ResNet-like structure with four residual blocks (64–256 channels), and the generator adopts a U-Net-style symmetric encoder–decoder architecture with skip connections and AdaIN-based modulation. All models were trained using the Adam optimizer (learning rate = 0.0003, $\beta_1 = 0.9$, $\beta_2 = 0.999$) for 120 epochs, with early stopping applied if the validation FID did not improve for 10 consecutive epochs. The loss weights follow the configuration given in Table 1 and were kept fixed across all experiments to avoid hyperparameter-driven performance fluctuations.

Table 1
Parameter sensitivity test results.

Combination	Learning rate	Identity vector dimension	Perturbation strength	Identity categories κ	FID ↓	mIoU ↑	ID retention ↓
P1	0.0005	256	0.05	40	23.6	62.8	9.1
P2	0.0003	128	0.08	40	24.2	61.9	10.3
P3	0.0003	256	0.10	50	22.8	63.5	8.7
P4	0.0001	128	0.10	60	25.4	60.2	11.5
P5	0.0001	256	0.12	50	23.1	63.0	9.3
P6	0.0005	128	0.12	60	24.6	61.2	10.7
P7	0.0003	192	0.06	50	22.9	63.8	9.0
P8	0.0001	192	0.08	40	24.0	62.5	9.8

In all experimental processes, the weight parameters of the loss function were uniformly set as follows: The weight λ_1 of the identity elimination loss term was 1, the λ_2 of image reconstruction loss term was 10, the λ_3 of the structural consistency loss was 1, the λ_4 of the pose preservation loss was 5, the λ_5 of the background preservation loss was 0.1, and the λ_6 of the style regularization loss was 0.01. To verify the performance stability of the proposed model under the parameter settings of each submodule, parameter sensitivity testing was first conducted to explore the effects of loss function weight factors and privacy disturbance intensity on the performance of anonymous images. The specific results are shown in Table 1.

From the experimental results in Table 1, parameter combination P3 performed well in multiple key indicators, achieving the lowest Fréchet Inception Distance (FID) of 22.8 and the highest Mean Intersection over Union (mIoU) of 63.5, while maintaining a low identity retention rate of 8.7. This combination can effectively suppressed the recognizability of identity features while improving the quality of image structure restoration. Further comparison between P2 and P4 revealed that moderately increasing the disturbance intensity and identity vector dimension could enhance the model's expressive power, but setting too many identity categories might introduce discriminative confusion. P3 was ultimately determined as the optimal parameter combination, and subsequent experiments were trained and evaluated based on a learning rate of 0.0003, an identity vector dimension of 256, a perturbation intensity of 0.10, and 50 identity categories.

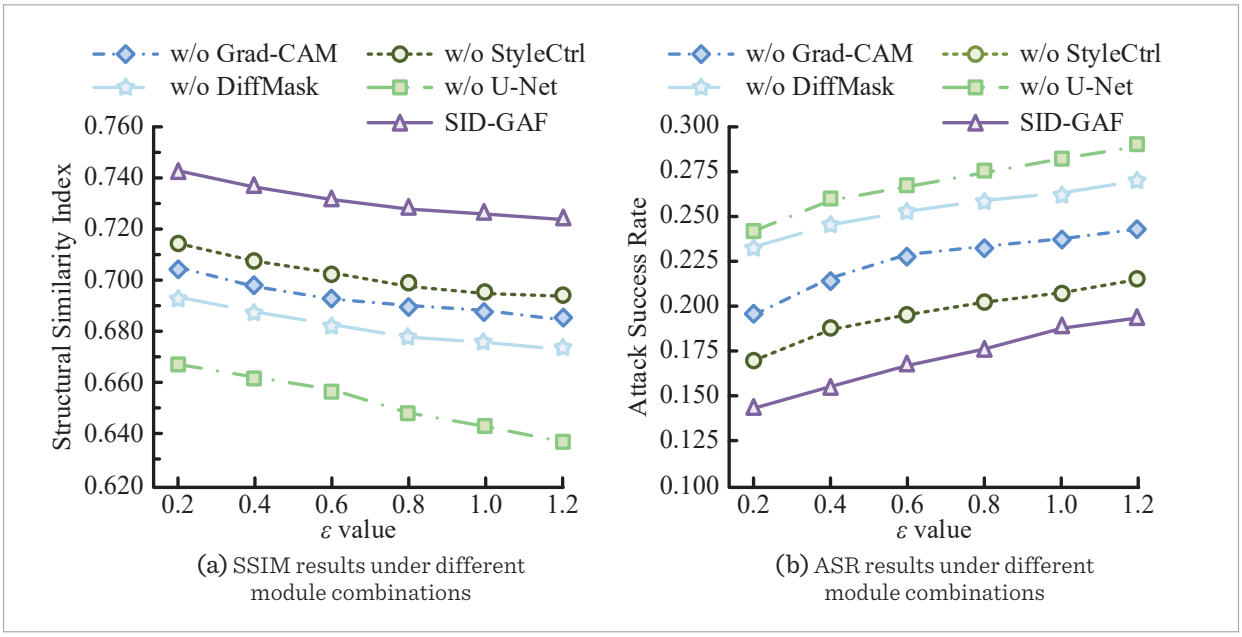
Further module ablation experiments were conducted, with four key sub modules removed and a control group constructed, including w/o Grad-CAM for removing the identity feature response regulation module (without using Grad-CAM for deactivation guidance), w/o DiffMask for removing the structural perturbation branch (directly encoding the original structure without differential perturbation), w/o StyleCtrl for removing the style control network (without introducing style guidance, only retaining the basic decoding process), and w/o U-Net for removing the multi-scale reconstruction path (using a symmetric convolution structure without skip connections). Using privacy budget ε as the control variable, the research observed the performance fluctuation characteristics of different module combinations under different privacy budget intensities, as shown in Figure 8.

In Figure 8(a), when the privacy budget parameter was set to 1.2, the Structural Similarity Index (SSIM) after removing Grad-CAM, DiffMask, StyleCtrl, and U-Net modules were 0.684, 0.672, 0.693, and 0.637, respectively, all significantly lower than the 0.723 of the complete SID-GAF model. ASR measures the proportion of anonymized samples whose identity can still be matched by the ArcFace adversarial recognizer under cosine similarity with threshold 0.3:

$$ASR = \frac{1}{N}\sum_{i=1}^{N}\mathbf{1}(\sin(f(z_i), f(x_i)) \geq 0.3).$$

Figure 8(b) shows that the results of the five models in terms of ASR for identity recognition attacks were 0.239, 0.270, 0.216, and 0.284, respectively,

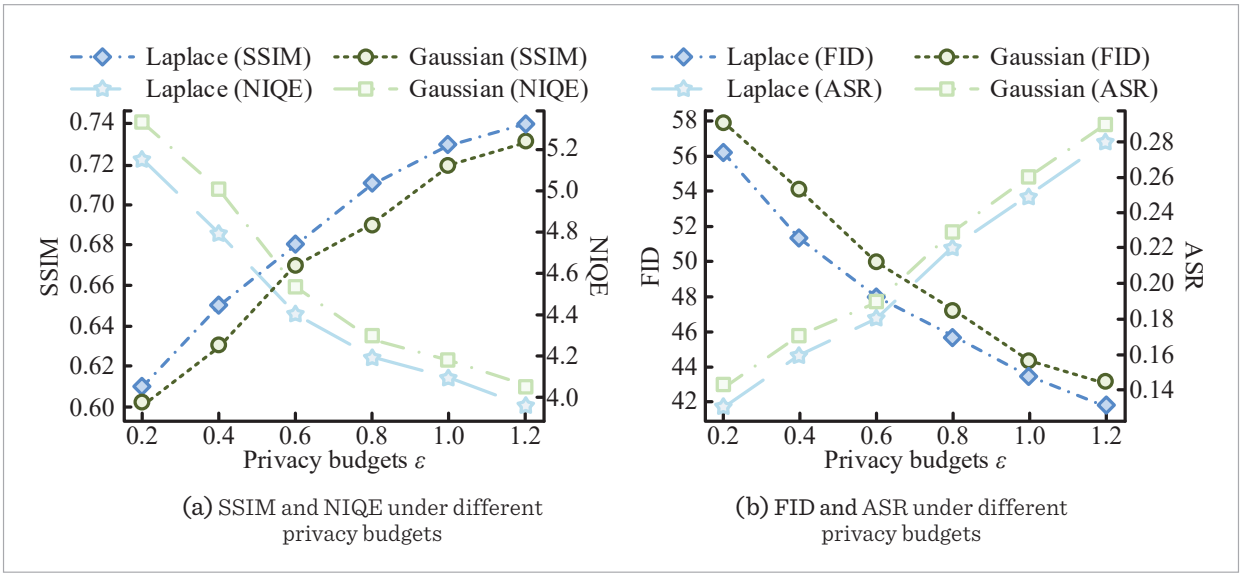
Figure 8
Module ablation test results.



(a) SSIM results under different module combinations

(b) ASR results under different module combinations

Figure 9
Comparison of perturbation mechanisms under different privacy budget parameters.



(a) SSIM and NIQE under different privacy budgets

(b) FID and ASR under different privacy budgets

all higher than the complete model's 0.195. Overall, removing Grad-CAM and DiffMask weakened the model's ability to suppress identity sensitive areas and structural perturbations, resulting in a decrease in anonymity performance. The lack of U-Net could lead to insufficient information fusion in the decoding path, resulting in a decrease in image restoration quality and an increase in attack risk. Although StyleCtrl had a relatively small impact on ASR, the semantic coordination and background consistency of the generated image were significantly weakened after the absence.

To investigate the impact of differential privacy budget on model generation performance and anonymity strength, the experiment set privacy budget parameter $\varepsilon \in \{0.2, 0.4, 0.6, 0.8, 1.0, 1.2\}$, and used Gaussian and Laplace mechanisms to perturb structural features with noise. SSIM, Natural Image Quality Evaluator (NIQE), FID, and ASR were evaluated separately, and the results are shown in Figure 9.

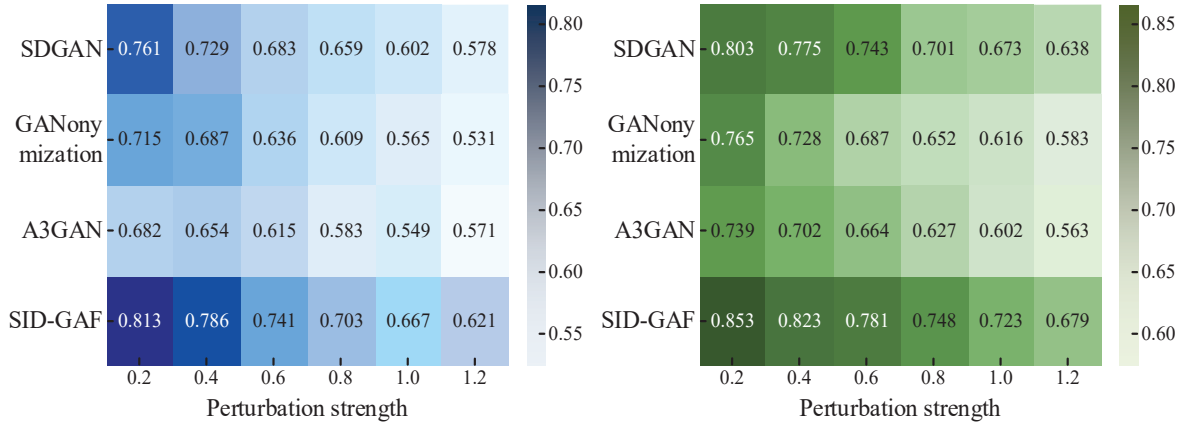
In Figure 9(a), as the privacy budget ε increased, SSIM showed an overall upward trend, indicating that the fidelity of image structure improved with the weakening of noise disturbance intensity. Meanwhile, NIQE gradually decreased, reflecting an improvement in visual naturalness. In contrast, the Laplacian mechanism had higher SSIM and lower NIQE in most ε settings, indicating its advantage in maintaining image realism. This was because the sparse perturbations generated by the Laplace mechanism caused less damage to the local structure and preserved more geometric details. In Figure 9(b), as the ε increased, the overall FID decreased, and the distribution of image features became more consistent with the real data. However, ASR increased accordingly, indicating that attackers were more likely to recover effective identity features from the generated image. The Gaussian mechanism exhibited low ASR within the range of $\varepsilon \leq 0.6$, reflecting its suppression ability under strong privacy constraints. However, in high ε settings, ASR rapidly increased, indicating that it was more likely to expose sensitive information in high signal-to-noise scenarios. Based on comprehensive evaluation, when $\varepsilon = 0.8$, the Laplacian mechanism achieved a better compromise among various indicators, balancing anonymity and image quality.

Further performance tests of generated images under different disturbance intensities were designed, and background preservation rate and pose preservation were used as evaluation indicators to compare with existing mainstream anonymous generation methods. The comparative methods included Semantic-aware De-identification Generative Adversarial Network (SDGAN) [11], Anonymization Framework with Facial Expression Preserving Ability (GANonymization) [6], and Attribute-aware Anonymization Network (A3GAN) [32]. The results are shown in Figure 10.

Background Retention Rate (BRR) evaluates how well non-facial regions are preserved after anonymization, computed as $BRR = \frac{1}{N}\sum_{i=1}^{N}\frac{|B_i \cap B_i'|}{|B_i|}$, where B_i and B_i' denote the background pixel sets before and after anonymization, respectively. From Figure 10(a), in terms of BRR, the proposed method SID-GAF outperformed the comparison method at all perturbation intensities, with a maximum of 0.813. The performance of SDGAN and GANonymization fluctuated greatly under strong disturbance conditions, which was speculated to be

Figure 10

Comparison of background retention rate and posture preservation under different disturbance intensities.



(a) Background retention test results (b) Pose Consistency test results

related to the lack of clear background modeling mechanisms, and the emphasis on facial expression camouflage in GANonymization led to a decrease in the ability to express background features. A3GAN exhibited relatively stable performance under moderate disturbances, but lacked specific structural support during disturbance enhancement, resulting in rough image boundary processing and a decrease in retention rate. In Figure 10(b), the attitude preservation results showed that SID-GAF, with the constraint of attitude loss term and structural alignment mechanism, maintained high consistency under different disturbance intensities, up to 0.853. Although GANonymization had the advantage of preserving facial expressions, it lacked a clear posture modeling strategy, resulting in a decrease in performance under high disturbances.

To further verify the performance differences of different anonymization methods in preserving facial visual features and removing identities, all methods were inferred under the same input samples to maintain consistency in test conditions. The results are shown in Figure 11.

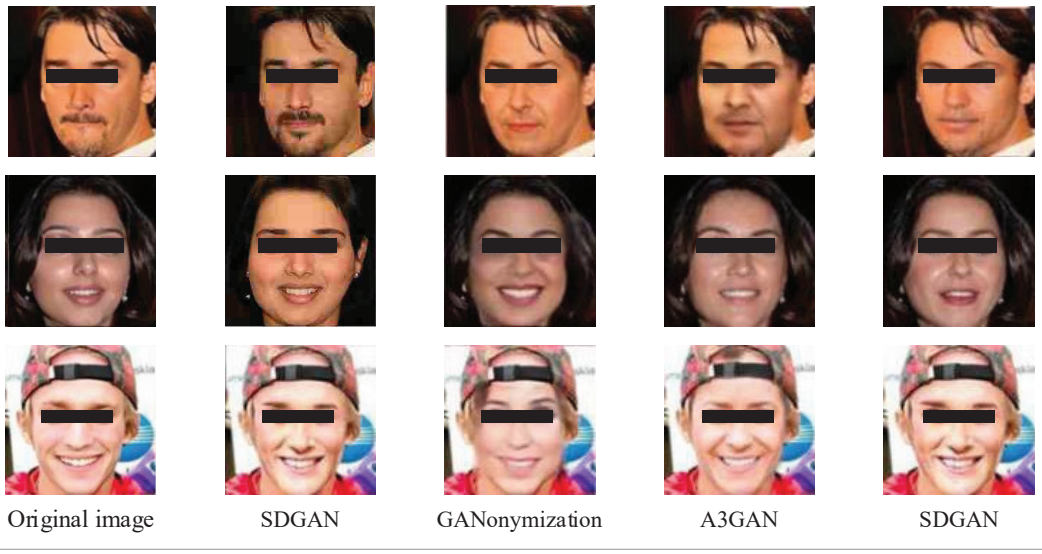
Figure 11 shows significant differences between the various methods in facial feature reconstruction and identity perturbation. SDGAN could preserve local facial structure to a certain extent, but some identity remained. GANonymization focused more on expression consistency, resulting in natural and smooth generated results, but some samples exhibited slight feature oversmoothing. A3GAN achieved more significant facial feature changes under attribute space perturbation, with more pronounced differences in appearance characteristics such as gender and age, but its structural preservation was weaker. In contrast, the proposed SID-GAF model, by introducing a structure-identity separation mechanism, only deflected features in prominent areas such as the eyes, nose, and mouth, maintaining the original facial contour and lighting consistency. The generated results achieved a better balance between visual naturalness, structural stability, and identity removal strength.

3.2. Actual Testing of SID-GAF in Facial Anonymity Applications

To further validate the generalization and adaptation performance of the proposed structure identity separation guided facial anonymization generation model in real-world application environments, four types of testing tasks were constructed around typical scenarios, covering dimensions such as different posture conditions, social image environments, edge device deployment, and judicial review evaluation, to comprehensively evaluate the practicality, stability, and interpretability of the model in the process

Figure 11

Comparison of visual effects of different anonymization methods.



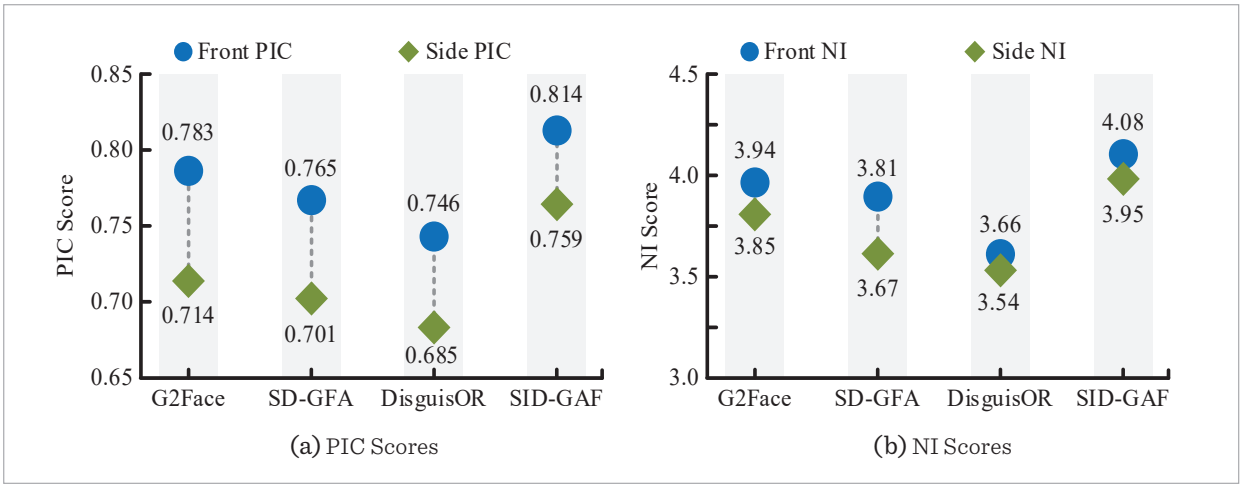
of facial anonymization. The comparison method selected the current mainstream anonymity generation frameworks, including the high fidelity reversible anonymity generation network G2Face [30], the Style Diversity-based Generative Face Anonymization (SD-GFA) framework [34], and the structure preserving anonymity scheme DisguisOR [1] suitable for medical scenarios, representing three research directions: identity preservation control, style diversity, and local feature masking. To test the anonymity consistency and natural expression ability of the model under different posture conditions, a cross posture consistency evaluation experiment was first conducted, and the results are shown in Figure 12.

Pose-based Identity Consistency (PIC) measures cross-view anonymization coherence via $PIC = sim(f(z_{\text{front}}), f(z_{\text{profile}}))$, where $f(\cdot)$ is the ArcFace-based embedding function and $sim(\cdot)$ denotes cosine similarity between frontal and profile anonymized samples. Figure 12(a) shows the PIC scores for both frontal and lateral poses, with SID-GAF reaching 0.814 and 0.759, respectively. In contrast, although G2Face had a better design in terms of reversibility, it was prone to introducing facial geometric distortion during complex pose switching, resulting in decreased consistency. SD-GFA and DisguisOR limited modeling capabilities for style and structure, making it difficult to fully maintain identity neutral representations across postures. Naturalness Index (NI) is obtained through human perceptual scoring on a five-level realism scale. Figure 12(b) reflects the NI of the generated image, with SID-GAF reaching 4.08 and 3.95 in frontal and lateral poses, respectively. Its style regularization and structural consistency joint optimization strategy enhanced the coordination and semantic coherence between the anonymous area and the surrounding background. Due to the lack of a fine reconstruction module, DisguisOR was prone to artifacts and blurring during the structural restoration stage, which affected the natural perception of images.

To verify the privacy preservation ability in social image scenarios, the experiment introduced three indicators: RFR, Intra-group Feature Separability (IFS), and Target Anonymization Confusion Rate (TACR). RFR measures the proportion of anonymized samples that cannot be matched back to their original identity: $RFR = \frac{1}{N}\sum_{i=1}^{N}\mathbf{1}(\sin(f(z_i), f(x_i)) < 0.3)$, where x_i and z_i are the original and anonymized images and the threshold 0.3 is applied to cosine similarity. IFS assesses embedding dispersion: $IFS = \frac{1}{M}\sum_{p\neq q} dist(g(p), g(q))$, where $g(\cdot)$ denotes the latent feature mapping and $dist(\cdot)$ is Euclidean distance. TACR indicates identity misclassification: $TACR = \frac{1}{N}\sum_{i=1}^{N}\mathbf{1}(\arg\max_j sim(f(z_i), f(x_j)) \neq i)$, where a misclassification is counted if the anonymized sample is most similar to the embedding of an incorrect identity. The research observed the changes in

Figure 12

Results of face anonymity consistency and naturalness test across poses.



(a) PIC Scores (b) NI Scores

target identity recognition and group distribution characteristics after anonymous intervention at three levels of image occlusion ratios: low occlusion (10%), medium occlusion (25%), and high occlusion (40%). The results are shown in Table 2.

As shown in Table 2, in terms of RFR indicators, as the image occlusion ratio increased from low to high, SID-GAF reached 27.2% in high occlusion scenes, which was better than G2Face's 18.9%, and had stronger identity hiding ability in complex multi-person images. IFS display showed that SID-GAF always maintained the lowest value, dropping to 0.261 under high occlusion conditions, reflecting its minimal interference from anonymous targets and non-targets in a group background, and maintaining reasonable local structural consistency. TACR revealed the probability of misidentifying the target as someone else after anonymous processing, with SID-GAF reaching 25.9%, 29.4%, and 31.0% under three levels of occlusion, respectively, reflecting its high semantic fusion and identity suppression capabilities in anonymous camouflage, and exhibiting stronger anonymity adaptability in the context of multi-target interference.

To test the practical usability of generated images in high reliability applications such as legal review, the experiment selected four typical judicial image scenarios: natural light, side face occlusion, low resolution, and facial expression changes. The distribution stability and performance of each anonymization model were analyzed on two indicators: facial readability score (FRS) and semantic retention rate (SRR). FRS is assigned by expert evaluators based on visual interpretability of critical facial attributes. SRR measures the consistency of non-identity attributes (e.g., expression) in anonymized samples: $SRR = \frac{1}{N}\sum_{i=1}^{N}\left(1 - dist(s(x_i), s(z_i))\right)$, where $s(x_i)$ and are $s(z_i)$ semantic attribute feature vectors before and after anonymization. The results are shown in Figure 13.

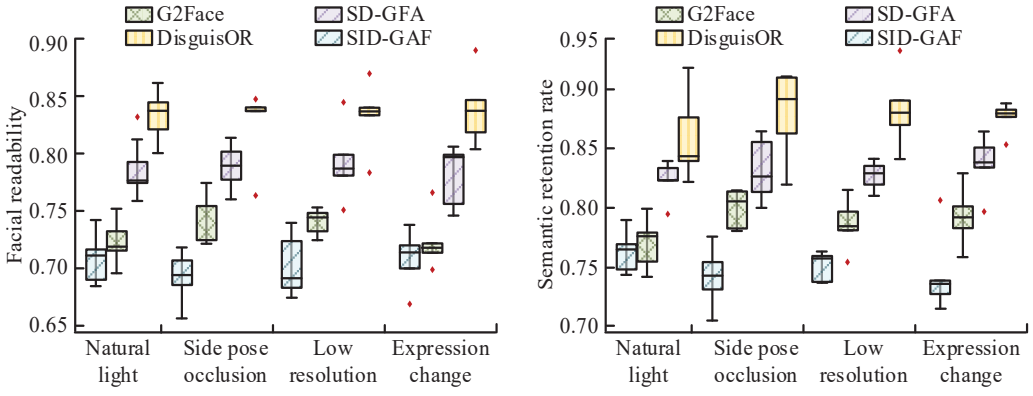
Figure 13(a) shows that under the FRS index, SID-GAF exhibits overall higher median and smaller discrete intervals in four types of judicial scenarios, demonstrating strong stability in maintaining the clarity of local facial structures. Among them, the median scores of SID-GAF in natural light and facial expression change scenes were 0.734 and 0.719, respectively, demonstrating good image detail restoration ability. In Figure 13(b), the box line height of SID-GAF on the SRR index was concentrated, and there were few outliers, reflecting its robustness in the semantic information transmission chain. In comparison, the SRR fluctuation of G2Face under side face occlusion was large, reflecting the limitations of its geometric alignment mechanism under

Table 2
Privacy preservation and group distribution test results under different occlusion ratios.

Occlusion Level	Method	RFR (%)	IFS	TACR (%)
Low (10%)	G2Face	22.4	0.341	18.7
	SD-GFA	25.6	0.324	21.3
	DisguisOR	27.1	0.310	23.5
	SID-GAF	30.8	0.292	25.9
Medium (25%)	G2Face	20.1	0.356	20.6
	SD-GFA	23.3	0.337	23.9
	DisguisOR	25.7	0.319	26.2
	SID-GAF	28.9	0.276	29.4
High (40%)	G2Face	18.9	0.370	22.2
	SD-GFA	21.4	0.352	25.1
	DisguisOR	24.2	0.331	27.4
	SID-GAF	27.2	0.261	31.0

Figure 13

Results of face readability and semantic preservation performance in different judicial scenarios.



(a) Facial readability under different judicial scenarios

(b) Semantic retention rate under different judicial scenarios

complex occlusion conditions. Although DisguisOR maintained a certain advantage over FRS, its style perturbation had a significant impact on semantic consistency, resulting in a lower SRR.

In addition to the GAN-based anonymization baselines, two diffusion-driven approaches were included: Diff-Privacy [5], which follows the diffusion-based facial privacy protection method of He et al., and Face Anonymization Made Simple [12], a recent diffusion-based anonymization model requiring only reconstruction loss without landmarks or masks. The comparison is summarized in Table 3.

As shown in Table 3, the SID-GAF proposed by the research performed well overall in five indicators, especially in terms of inference delay, number of generated images per unit energy consumption, and peak memory, showing significant advantages of 114.8 ms, 19.2 IpJ, and 1286 MB, respectively. In comparison, DisguisOR had a complex multi-scale structure, resulting in a delay of up to 142.3 ms, and was at a disadvantage in terms of energy efficiency ratio and memory overhead. SDGAN and A3GAN improved model compression ratio and energy efficiency through feature sharing and attribute guid-

Table 3

Performance comparison results in a simulated edge device deployment environment.

Method	Inference Latency (ms)	Compression Ratio	Images per joule (IpJ)	Model Params (M)	Peak Memory (MB)
G2Face	137.5	3.1	16.7	24.3	1478
SD-GFA	128.9	3.4	17.5	20.7	1392
DisguisOR	142.3	2.9	15.9	26.2	1525
SDGAN	119.6	3.7	18.9	18.4	1327
GANonymization	133.8	3.2	16.4	22.0	1409
A3GAN	125.7	3.6	18.1	19.2	1355
Diff-Privacy	265.4	2.8	12.3	28.6	1584
Face Anonymization Made Simple	247.8	2.9	13.1	27.5	1557
SID-GAF (proposed)	114.8	3.8	19.2	17.1	1286

ance mechanisms, respectively, but they still lagged behind in inference speed. Diffusion-based models (Diff Privacy and Face Anonymization Made Simple) performed well in image fidelity, but their multi-step sampling process resulted in significantly higher inference latency and memory consumption compared to other methods, with inference latency reaching 265.4 ms and 247.8 ms, respectively, and IpJ only 12.3 and 13.1. In contrast, SID-GAF achieved significant optimization of inference speed and energy efficiency under the collaborative framework of structural modeling and identity suppression, with a latency of only 114.8 ms and an IpJ of 19.2, demonstrating efficient adaptability and real-time potential on resource constrained devices.

To support the quantitative comparisons, statistical testing was carried out using five independent runs per model configuration with different random seeds. Each run generated 1,000 anonymized images, and per-image metric values were treated as independent samples for the calculation of FID and SSIM. For re-identification-related metrics such as ASR, results were computed per identity over the 250 identity groups. Two-sample t-tests were applied between SID-GAF and each baseline, and Bonferroni correction was used to account for multiple pairwise comparisons. In addition, to mitigate potential distributional bias of FID, a non-parametric Mann–Whitney U test was also performed and yielded consistent significance trends. The results are summarized in Table 4.

The statistical results indicated that SID-GAF consistently achieved lower FID and ASR and higher SSIM compared to evaluated baselines under the specified testing configuration. While the performance differences remained statistically significant after correction, the conclusions were reported conservatively to reflect robustness trends rather than absolute guarantees.

In summary, to verify the core features of SID-GAF proposed by the research, the experiment conducted a systematic evaluation from four dimensions: structural consistency, identity suppression ability, visual naturalness, and inference efficiency. Figures 9-10 validate the structural preservation and attitude controllability of the model, where the SSIM and attitude preservation results directly reflected the effectiveness of structural modeling and geometric perturbation mechanisms at the semantic level. The ablation experiment in Figure 8 and the ASR index verified the weakening effect of identity saliency inhibition mechanism on the risk of identity leakage. The analysis of naturalness and semantic preservation in Figures 11-12 shows that SID-GAF could still maintain high visual quality and content consistency under differential perturbations, reflecting the model's advantages in naturalness and semantic coordination. The edge deployment results demonstrated the advantages of the model in inference latency, energy efficiency, and memory usage, demonstrating its good scalability and practical adaptability under resource limited conditions.

4. Conclusion

Aiming at the risk of privacy leakage of identity information in facial images, a novel anonymous image generation method SID-GAF was developed, which integrated structural modeling, style guidance, and saliency suppression mechanisms. In the ablation test, when the privacy budget parameter

Table 4

Statistical Significance Analysis of Key Evaluation Metrics.

Method	FID (Mean ± SD)	SSIM (Mean ± SD)	ASR (Mean ± SD)	p-value	Significance
G2Face	26.8 ± 0.9	0.693 ± 0.012	0.239 ± 0.007	0.0023	**
SD-GFA	24.1 ± 0.8	0.702 ± 0.009	0.221 ± 0.005	0.0018	**
Diff-Privacy	22.9 ± 0.7	0.714 ± 0.011	0.213 ± 0.006	0.0009	**
SID-GAF	21.7 ± 0.6	0.723 ± 0.010	0.195 ± 0.004	/	/

Note: * indicates "p<0.05 (significant difference)"; ** indicates "p<0.01 (highly significant difference)"; *** indicates "p<0.001 (highly significant difference)"
Non-parametric Mann–Whitney U testing indicated consistent result trends.

value was 1.2, the SSIMs of the complete model, removed Grad-CAM, DiffMask, StyleCtrl, and U-Net modules were 0.723, 0.684, 0.672, 0.693, and 0.637, respectively, with corresponding ASR values of 0.195, 0.239, 0.270, 0.216, and 0.284, respectively. When the disturbance intensity was 0.2, the BRR of SID-GAF was 0.813 and the attitude retention was 0.853. In application testing, the PIC of the proposed method was 0.814 and 0.759, and the NI was 4.08 and 3.95, respectively, in the frontal and lateral postures. In the 40% occlusion ratio of the high-occlusion image, the RFR of the proposed method was 27.2%, IFS was 0.261, and TACR was 31.0%. Under natural light and facial expression changes, the median scores of SID-GAF were 0.734 and 0.719, respectively, and the median scores of SRR were 0.767 and 0.738, respectively. The comprehensive evaluation results indicated that SID-GAF achieved a good balance between identity hiding and image quality in multiple task scenarios. Although the SID-GAF proposed by the research achieved excellent performance in identity suppression, structural preservation, and visual consistency, there are still certain limitations. Under complex lighting, extreme posture, or local occlusion conditions, the robustness of structural keypoint extraction still needs to be improved. There was an imbalance in the distribution of privacy-enhancing noise among different populations and expression categories, and the model's cross dataset generalization ability still needs further validation. Future research will focus on exploring adaptive structural perturbations and dynamic privacy budget learning mechanisms to enhance the anonymity strength control and stability of the model in multiple scenarios. Meanwhile, the research will consider extending SID-GAF to video level anonymous consistency modeling, combining temporal constraints and cross frame optimization strategies to achieve higher-level continuous privacy protection and application scalability.

Funding

The research is supported by Heilongjiang Province Traditional Chinese Medicine Scientific Research Project, "Research on the Construction of a Visualized Knowledge Graph for the Diagnosis and Treatment of Stroke from the Ancient Classic 'Great Compendium of Acupuncture and Moxibustion' Based on AI Learning", ZHY2024-114; Heilongjiang Province Postdoctoral Research Startup Foundation, "Research on Healthcare Information Evolution Analysis Technology in Cold Region Contexts", LBH-Q18115.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

1. Bastian, L., Wang, T. D., Czempiel, T., Busam, B., Navab, N. DisguisOR: Holis-tic Face Anonymization for the Operat-ing Room. International Journal of Computer Assisted Radiology and Sur-gery, 2023, 18(7), 1209-1215. https://doi.org/10.1007/s11548-023-02939-6
2. Bastian, L., Wang, T. D., Czempiel, T., Busam, B., Navab, N. DisguisOR: Holis-tic Face Anonymization for the Operat-ing Room. International Journal of Computer Assisted Radiology and Sur-gery, 2023, 18(7), 1209-1215. https://doi.org/10.1007/s11548-023-02939-6
3. Chen, C. Y., Jhong, S. Y., Hsia, C. H. A Domain Generalized Face Anti-Spoofing System Using Domain Adversarial Learning. International Journal of Engi-neering and Technology Innovation, 2024, 14(4), 378-388. https://doi.org/10.46604/ijeti.2024.13314
4. Ghani, M. A. N. U., She, K., Rauf, M. A., Khan, S., Khan, J. A., Aldakheel, E. A., Khafaga, D. S. Enhancing Security and Privacy in Distributed Face Recognition Systems Through Blockchain and GAN Technologies. Computers, Materials & Continua, 2024, 79(2), 2609-2623. https://doi.org/10.32604/cmc.2024.049611
5. Gong, M., Liu, J., Li, H., Xie, Y., Tang, Z. Disentangled Representation Learning for Multiple Attributes Preserving Face Deidentification. IEEE Transactions on Neural Networks and Learning Systems, 2022, 33(1), 244-256. https://doi.org/10.1109/TNNLS.2020.3027617
6. He, X., Zhu, M., Chen, D., Wang, N., Gao, X. Diff-Privacy: Diffusion-Based Face Privacy Protection. IEEE Transactions on Circuits and Systems for Video Tech-nology, 2024, 34(12), 13164-13176. https://doi.org/10.1109/TCSVT.2024.3449290
7. Hellmann, F., Mertes, S., Benouis, M., Hustinx, A., Hsieh, T. C., Conati, C., An-dré, E. Ganonymization:

A GAN-Based Face Anonymization Framework for Preserving Emotional Expressions. ACM Transactions on Multimedia Compu-ting, Communications, and Applica-tions, 2024, 21(1), 1-27. https://doi.org/10.1145/3641107

8. Jeremiah, S. R., Castro, O. E. L., Sharma, P. K., Park, J. H. CCTV Footage De-Identification for Privacy Protection: A Comprehensive Survey. Journal of In-ternet Technology, 2024, 25(3), 379-386. https://doi.org/10.53106/160792642024052503004
9. Kavoliūnaitė-Ragauskienė, E. Right to Privacy and Data Protection Concerns Raised by the Development and Usage of Face Recognition Technologies in the European Union. Journal of Human Rights Practice, 2024, 16(2), 658-674. https://doi.org/10.1093/jhuman/huad065
10. Khan, S., Huh, J., Ye, J. C. Switchable and Tunable Deep Beamformer Using Adap-tive Instance Normalization for Medical Ultrasound. IEEE Transactions on Medi-cal Imaging, 2022, 41(2), 266-278. https://doi.org/10.1109/TMI.2021.3110730
11. Khan, W., Topham, L., Khayam, U., Or-tega-Martorell, S., Heather, P., Ansell, D., Hussain, A. Person De-Identification: A Comprehensive Review of Methods, Datasets, Applications, and Ethical As-pects Along with New Dimensions. IEEE Transactions on Biometrics, Behavior, and Identity Science, 2024, 7(3), 293-312. https://doi.org/10.1109/TBIOM.2024.3485990
12. Kim, H., Pang, Z., Zhao, L., Su, X., Lee, J. S. Semantic-Aware Deidentification Generative Adversarial Networks for Identity Anonymization. Multimedia Tools and Applications, 2023, 82(10), 15535-15551. https://doi.org/10.1007/s11042-022-13917-6
13. Kung, H. W., Varanka, T., Saha, S., Sim, T., Sebe, N. Face Anonymization Made Simple. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2025, 1040-1050. https://doi.org/10.1109/WACV61041.2025.00110
14. Li, R., Gao, M., Jin, X. Recognize Me If You Can: Two-Stream Adversarial Transfer for Facial Privacy Protection Using Fine-Grained Makeup. The Visual Computer, 2025, 41(9), 6943-6954. https://doi.org/10.1007/s00371-025-04015-3
15. Meena, G., Mohbey, K. K., Indian, A., Khan, M. Z., Kumar, S. Identifying Emo-tions from Facial Expressions Using a Deep Convolutional Neural Network-Based Approach. Multimedia Tools and Applications, 2024, 83(6), 15711-15732. https://doi.org/10.1007/s11042-023-16174-3
16. Morris, J., Newman, S., Palaniappan, K., Fan, J., Lin, D. Do You Know You Are Tracked by Photos That You Didn't Take: Large-Scale Location-Aware Multi-Party Image Privacy Protection. IEEE Transactions on Dependable and Secure Computing, 2023, 20(1), 301-312. https://doi.org/10.1109/TDSC.2021.3132230
17. Mulay, S., Ram, K., Sivaprakasam, M. Attention Adaptive Instance Normaliza-tion Style Transfer for Vascular Segmen-tation Using Deep Learning. Applied In-telligence, 2023, 53(24), 29638-29655. https://doi.org/10.1007/s10489-023-05033-1
18. Osorio-Roig, D., Rathgeb, C., Drozdowski, P., Busch, C. Stable Hash Generation for Efficient Privacy-Preserving Face Identification. IEEE Transactions on Biometrics, Behavior, and Identity Science, 2022, 4(3), 333-348. https://doi.org/10.1109/TBIOM.2021.3100639
19. Osorio-Roig, D., Rathgeb, C., Drozdowski, P., Terhörst, P., Štruc, V., Busch, C. An Attack on Facial Soft-Biometric Privacy Enhancement. IEEE Transactions on Biometrics, Behavior, and Identity Science, 2022, 4(2), 263-275. https://doi.org/10.1109/TBIOM.2022.3172724
20. Preethi, P., Mamatha, H. R. Region-Based Convolutional Neural Network for Segmenting Text in Epigraphical Im-ages. Artificial Intelligence Applica-tions, 2023, 1(2), 119-127. https://doi.org/10.47852/bonviewAIA2202293
21. Qi, C., Li, M., Wang, Q., Zhang, H., Xing, J., Gao, Z., Zhang, H. Facial Expression Recognition Based on Cognition and Mapped Binary Patterns. IEEE Access, 2018, 6, 18795-18803. https://doi.org/10.1109/ACCESS.2018.2816044
22. Rot, P., Grm, K., Peer, P., Štruc, V. Priva-cyProber: Assessment and Detection of Soft-Biometric Privacy-Enhancing Tech-niques. IEEE Transactions on Dependa-ble and Secure Computing, 2024, 21(4), 2869-2887. https://doi.org/10.1109/TDSC.2023.3319500
23. Sano, T. Facial Attractiveness Factors and Their Learning Processes by Com-paring Class Activation Mapping-Based Visualization Methods Using Convolu-tional Neural Networks. International Journal of Affective Engineering, 2023, 22(3), 201-208. https://doi.org/10.5057/ijae.IJAE-D-22-00013
24. Shen, J., Tian, J., Wang, Z., Zhu, Q. Pre-serving Social Relationship Privacy via the Exponential Mechanism of Personal-ized Differential Privacy. IEEE Transac-tions on Computational Social Systems, 2025, 12(3), 1164-1180. https://doi.org/10.1109/TCSS.2024.3508744

25. Sivalakshmi, M., Prasad, K. R., Bindu, C. S. Convolutional-Based Variational Au-toencoders for Face Privacy Protection in Video Surveillance. Journal of Dis-crete Mathematical Sciences and Cryp-tography, 2024, 27(4), 1205-1214. https://doi.org/10.47974/JDMSC-1974

26. Wang, T., Zhang, Y., Yang, Z., Xiao, X., Zhang, H., Hua, Z. Seeing Is Not Believ-ing: An Identity Hider for Human Vi-sion Privacy Protection. IEEE Transactions on Biometrics, Behavior, and Iden-tity Science, 2025, 7(2), 170-181. https://doi.org/10.1109/TBIOM.2024.3449849

27. Wang, Y., Li, J., Zhang, H., Zhang, J., Wan, F., Qiu, A., Gao, Z. FedMDD: Mul-ti-Deliberation Based Calibration for Federated Long-Tailed Learning. Knowledge-Based Systems, 2025, 283, 113741. https://doi.org/10.1016/j.knosys.2025.113741

28. Xiao, Y., Han, J. J., Cen, J., Fang, Z. Ten-sor-Based Privacy Protection Scheme with Multifeature Fusion for Facial Recognition. IEEE Transactions on Sys-tems, Man, and Cybernetics: Systems, 2025, 55(6), 3964-3975. https://doi.org/10.1109/TSMC.2025.3547887

29. Xie, T., Yuan, L., Zhang, Q., Wu, J., Ren, F. Ciphertext Fuzzy Retrieval Mecha-nism with Bidirectional Verification and Privacy Protection. IEEE Internet of Things Journal, 2024, 11(24), 41061-41083. https://doi.org/10.1109/JIOT.2024.3458457

30. Yang, H., Xu, X., Xu, C., Zhang, H., Qin, J., Wang, Y., He, S. G2Face: High-Fidelity Reversible Face Anonymization via Generative and Geometric Priors. IEEE Transactions on Information Forensics and Security, 2024, 19, 8773-8785. https://doi.org/10.1109/TIFS.2024.3449104

31. Yang, X., Li, R., Yang, X., Zhou, Y., Liu, Y., Han, J. D. J. Coordinate-Wise Mono-tonic Transformations Enable Privacy-Preserving Age Estimation With 3D Face Point Cloud. Science China Life Sciences, 2024, 67(7), 1489-1501. https://doi.org/10.1007/s11427-023-2518-8

32. Zhai, L., Guo, Q., Xie, X., Ma, L., Wang, Y. E., Liu, Y. A3GAN: Attribute-Aware Anonymization Networks for Face De-Identification. Proceedings of the 30th ACM International Conference on Mul-timedia, 2022, 5303-5313. https://doi.org/10.1145/3503161.3547872

33. Zhong, Y., Deng, W. OPOM: Customized Invisible Cloak Towards Face Privacy Protection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(3), 3590-3603. https://doi.org/10.1109/TPAMI.2022.3175602

34. Zhu, M., He, P., Zhang, Y., Li, J., Qiu, Y. Against Linkage: A Novel Generative Face Anonymization Framework with Style Diversification. IET Image Pro-cessing, 2024, 18(13), 4114-4126. https://doi.org/10.1049/ipr2.13237