This WACV paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

Reverse Personalization

Han-Wei Kung[1] Tuomas Varanka[2] Nicu Sebe[1]
[1]University of Trento [2]University of Oulu
hanwei.kung@unitn.it

Figure 1. Our reverse personalization method removes identity-specific features while preserving both the original facial attributes and surrounding scene—without requiring subject finetuning. It also provides intuitive user control over which attributes are retained or modified, enabling flexible and customizable anonymization for downstream applications.

Abstract

Recent text-to-image diffusion models have demonstrated remarkable generation of realistic facial images conditioned on textual prompts and human identities, enabling creating personalized facial imagery. However, existing prompt-based methods for removing or modifying identity-specific features rely either on the subject being well-represented in the pre-trained model or require model finetuning for specific identities. In this work, we analyze the identity generation process and introduce a reverse personalization framework for face anonymization. Our approach leverages conditional diffusion inversion, allowing direct manipulation of images without using text prompts. To generalize beyond subjects in the model's training data, we incorporate an identity-guided conditioning branch. Unlike prior anonymization methods, which lack control over facial attributes, our framework supports attribute-controllable anonymization. We demonstrate that our method achieves a state-of-the-art balance between identity removal, attribute preservation, and image quality. Source code and data are available at https://github.com/hanweikung/reverse-personalization.

1. Introduction

Given an image of a human face, how can we remove identity-specific features while preserving non-identity attributes? Recent advances in diffusion models have enabled the creation of realistic visual content [65, 68], including face synthesis [8, 32, 44, 63]. Several studies [25, 52, 78] have demonstrated the effectiveness of these models in generating faces from prompts and identities. Additionally, research [9, 28, 33, 60] has shown that manipulating attention weights on specific tokens within prompts can influence semantic alignment and provide fine-grained control.

Building on this insight, we observe that adjusting attention weights on celebrity names within prompts controls

the likeness of generated images. However, this approach presents one limitation. If the input image shows a non-celebrity, their identity may not exist in the model's learned feature space. In such cases, modifying attention weights has a negligible effect, as the model lacks a representation of the individual.

To overcome this limitation, we first employ *diffusion inversion* techniques [33, 75], which map an input image into the latent space of a pre-trained model. This enables us to synthesize facial images with specific traits using conditioning methods. Beyond text prompts, recent advances have introduced expressive conditioning techniques, such as face embeddings [77, 82, 85] and semantic masks [18, 32, 87]. Inspired by developments in personalization [22, 78, 85], we adopt identity-conditioned generation using face embeddings, allowing us to extract identity features from arbitrary input images.

Our approach offers several key advantages. As an inversion-based method, it avoids model retraining, thus preserving the original generative capabilities of the diffusion model. This also ensures compatibility with other identity-conditioned generation methods and enables flexible control over high-level facial attributes. Importantly, our method supports image-only inputs, eliminating reliance on textual instructions.

A challenge then arises: how can we modulate how generated faces reflect the identity in the input image? By analyzing the generation process, we find that increasing the classifier-free guidance [30] scale amplifies identity-defining features. The model appears to first synthesize a neutral face, then progressively injects characteristic traits. This insight led us to experiment with guidance scales, hypothesizing that they could suppress identity-specific features and produce a *reverse* identity. Our experiments confirm this hypothesis, motivating the development of a novel mechanism termed *reverse personalization*.

Reverse personalization offers compelling benefits in *face anonymization*, where the goal is to protect personal identity while preserving utility. This capability is important across sectors such as healthcare [59, 84] and security [24, 41]. Anonymization techniques help address ethical concerns about surveillance and individual autonomy, while ensuring compliance with data privacy regulations including GDPR [2] and CCPA [1].

Despite progress in GAN- and diffusion-based anonymization, current methods face persistent challenges. Many struggle to strike a balance between removing identity-specific features and preserving non-identity attributes and realism [12, 57]. Furthermore, research has shown that individuals' willingness to disclose personal information—particularly sensitive demographic attributes such as age, race, and gender—is context-dependent [64]. For instance, in workplace settings, individuals may withhold such details to avoid bias or misunderstandings, especially in environments with status differences or limited trust [64]. Conversely, in healthcare contexts, disclosing demographic information can lead to improved outcomes by enabling more personalized and effective treatment [21]. However, existing face anonymization methods lack control over whether such attributes are retained or altered [47]. In contrast, our reverse personalization framework addresses this limitation, enabling users to flexibly control facial attributes in anonymized outputs.

We analyze identity-conditioned generation and apply our findings to face anonymization. We demonstrate that reverse personalization removes identifiable facial features while maintaining realism and attribute consistency. We contribute:

- *Conditional inversion:* we present a conditioning strategy that guides the inversion process to facilitate identity manipulation.
- *Reverse personalization:* we propose a guidance mechanism that steers the generative process away from identity-defining features, enabling anonymization while maintaining realism. Our method achieves an optimal balance between identity obfuscation, attribute preservation, and visual quality.
- *Attribute-controllable anonymization:* our approach includes intuitive controls for adjusting facial features like age, gender, and ethnicity, making it easy to generate customizable, anonymized results.

2. Related Work

Personalization. The field of personalized text-to-image generation [14, 22, 48, 70, 74] focuses on adapting diffusion models to synthesize images of specific subjects, styles, or concepts that are meaningful to individual users. This line of research enables models to learn visual concepts from a few example images, allowing for tailored image generation beyond general models.

Textual Inversion [22] optimizes token embeddings linked to a placeholder token, allowing the model to incorporate new concepts into text-driven generation with a few subject images. Similarly, DreamBooth [70] introduces a fine-tuning approach that binds an identifier to a specific subject. It leverages the semantic priors of pre-trained diffusion models while employing a class-specific prior preservation loss to maintain visual consistency with the broader data distribution.

Despite effectiveness, these methods suffer from computational costs, requiring several minutes to hours for fine-tuning. To overcome this limitation, parameter-efficient fine-tuning methods have emerged. HyperDreamBooth [71] introduces a hypernetwork-based architecture that generates personalized weights from a single image, offering a speedup—as much as 25 times faster than DreamBooth [70]

and 125 times faster than Textual Inversion [22]. JeDi [86] reduces computational demands by learning the joint distribution of multiple text-image pairs of a common subject, enabling finetuning-free personalization. IP-Adapter [85] adopts an alternative strategy based on image prompt adaptation, allowing users to guide image generation using reference images while retaining controllable text prompts.

These personalization techniques have unlocked a variety of applications, including identity-preserving face generation [25, 52, 54, 73, 78, 80, 82], virtual try-on [81], and customized content creation [61]. However, while most efforts focus on preserving identity, style, or subject appearance, we repurpose these principles toward the opposite goal—removing identifiable facial traits. Our work extends the technical foundations of personalization to enable attribute-controllable face anonymization.

Face anonymization. Face anonymization, or face de-identification, refers to concealing or altering facial traits to protect individual privacy in images. Traditional methods, such as blurring, pixelation, and masking, are simple and effective at preventing identification by humans. However, these approaches degrade the visual quality of images, destroying important contextual information like facial expressions, pose, and background—limiting their usefulness for downstream computer vision tasks [12, 57].

In response, modern face anonymization techniques strive for balance between privacy protection and data utility. Rather than obscuring faces, these methods aim to prevent identity recognition while preserving identity-agnostic attributes crucial for tasks such as emotion analysis [37, 42] and estimating head pose [53] and gaze [79]. Among these, GAN-based methods [11, 15, 27, 56] have been explored. DP [36] and its successor DP2 [34] employ conditional GANs [58] to generate synthetic faces that maintain pose and background context. GANonymization [26] further advances this idea by focusing on preserving facial expressions during anonymization. Methods like FALCO [6] and RiDDLE [51] exploit the latent space of StyleGAN2 [40] to generate realistic anonymized faces while retaining non-identity features.

Driven by recent progress in generative modeling, diffusion models have emerged as promising for face anonymization. For example, LDFA [45] combines face detection with a latent diffusion model to generate in-painted faces. FAMS [49] uses only reconstruction loss, eliminating dependence on identity losses derived from face recognition models or the use of facial landmarks and masks, which are prone to inaccuracies.

Despite these advancements, challenges remain. Achieving an balance between anonymization and preservation of image utility is difficult [35]. Additionally, most methods offer limited controllability over which sensitive attributes are preserved or concealed during anonymization. To address the limitations, our reverse personalization framework leverages the precise reconstruction and generative power of diffusion models. Our method enables user-controlled anonymization by allowing selective preservation or removal of specific facial attributes, while maintaining realism and utility for identity-agnostic tasks.

3. Method

3.1. Preliminary

DDPM Inversion. Denoising Diffusion Probabilistic Models (DDPMs) [31] define a forward process that gradually adds Gaussian noise to data and a learned reverse process that denoises it step by step. The reverse process reconstructs x_{t-1} from a noisy input x_t as follows:

$$x_{t-1} = \hat{\mu}_t(x_t) + \sigma_t z_t, \quad z_t \sim \mathcal{N}(0, I), \tag{1}$$

where $\hat{\mu}_t(x_t)$ is the predicted mean, σ_t is the variance schedule, and z_t is standard Gaussian noise. To recover the noise used at each step, we compute:

$$z_t = \frac{x_{t-1} - \hat{\mu}_t(x_t)}{\sigma_t}. \tag{2}$$

IP-Adapter. The IP-Adapter [85] is a lightweight module that enables diffusion models to use image prompts by modifying their attention mechanisms. Specifically, it integrates visual information without requiring model retraining. The modified attention is computed as:

$$\begin{aligned} \mathbf{Z}^{\text{new}} &= \text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) \\ &\quad + \lambda_{ipa} \cdot \text{Attention}(\mathbf{Q}, \mathbf{K}', \mathbf{V}'), \end{aligned} \tag{3}$$

where $\mathbf{Q}, \mathbf{K}, \mathbf{V}$ are the original queries, keys, and values, and $\mathbf{K}', \mathbf{V}'$ are keys and values derived from the image prompt. The IP-Adapter [85] scale λ_{ipa} controls the influence of the visual prompt. This mechanism facilitates flexible multimodal conditioning while preserving the pretrained model's structure.

3.2. Motivation

Given an image of a person, our goal is to modify defining facial features while preserving non-identifying features and the surrounding context. Achieving balance between data utility and identity protection is central to face anonymization, which supports scientific progress [27] and privacy [62].

An intuitive approach leverages inversion techniques in diffusion models, which map real images into latent representations compatible with a pretrained generator. One such method, Null-text Inversion [60], employs DDIM inversion [75] to extract a sequence of latent codes. These are

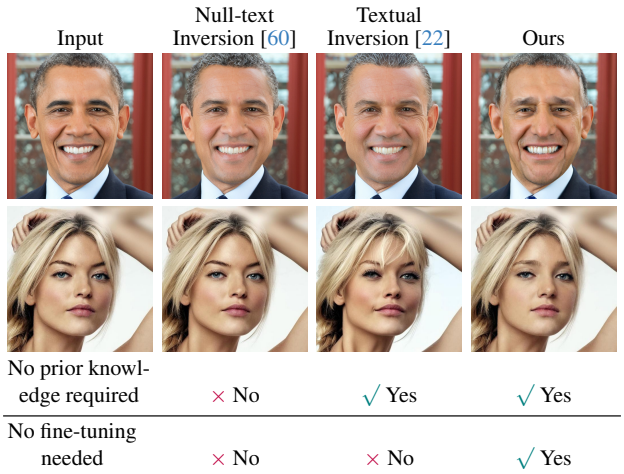


Figure 2. Prompt-based attention reweighting methods such as Null-text Inversion [60] can modify facial identity only when the diffusion model knows the subject (e.g., well-known figures, such as Obama). Personalization methods like Textual Inversion [22] require fine-tuning with multiple reference images. Unlike these methods, our reverse personalization approach can modify facial identity without prior model knowledge or fine-tuning.

then used to optimize the null-text embedding in classifier-free guidance [30], allowing the model to reconstruct the input image. This optimized embedding enables methods like Prompt-to-Prompt [28] to reweight attention maps during generation, offering fine-grained control. For instance, as shown in Fig. 2, attenuating the influence of the token "Obama" in the prompt "a photo of Obama" alters identity-specific facial features while preserving non-identity-relevant attributes like expression and the background context.

However, this strategy has a limitation: it relies on the model's knowledge of the subject's identity. When the individual is poorly represented in the training data—the woman in Fig. 2—adjusting prompt weights has little effect on the output.

To address this, model personalization techniques, such as Textual Inversion [22], are applied. Textual Inversion [22] fine-tunes the model's word embeddings on few images of a concept (e.g., the women in Figs. 1 and 2) and associates with a unique token S_*. This token can be used in prompts (e.g., "a photo of S_*") to trigger identity-aware generation.

Although this approach moves beyond the model's prior knowledge, it has challenges. It demands computational resources to fine-tune and may underperform when limited reference images are available, failing to capture the identity.

3.3. Reverse personalization

To address the limitations of prompt-based reweighting and fine-tuning methods, we propose an approach to modify identity-specific facial features without relying on prompt reweighting or computationally intensive personalization. Our method leverages identity embeddings extracted from facial images and integrates them through identity-conditioned adapters to modulate the diffusion process. A schematic overview of our approach is illustrated in Fig. 3.

In standard diffusion model inversion, the denoising network uses prompt embeddings to guide the recovery of noise trajectories, ensuring that the reconstructed image matches the input prompt semantically and visually. This enables prompt-based image editing, where modifying the prompt can yield targeted changes in the output.

However, in our reverse personalization setting—where only one image is provided and the goal is to alter identity—prompt substitution is not applicable. Instead, we introduce *null-identity embeddings* to guide the inversion process without identity-specific conditioning.

We further improve inversion efficiency using DPM-Solver++ [55], a higher-order solver that surpasses DDPM-based methods in speed [9]. This also ensures perfect reconstruction of the input image, preserving structural and contextual details while enabling modification of identity-specific features.

Our inversion process is formulated as:

$$z_t = \frac{x_{t-1} - \hat{\mu}_t(x_t, x_{t+1}, \varnothing_{id})}{\sigma_t}, \qquad t = T, \dots, 1, \quad (4)$$

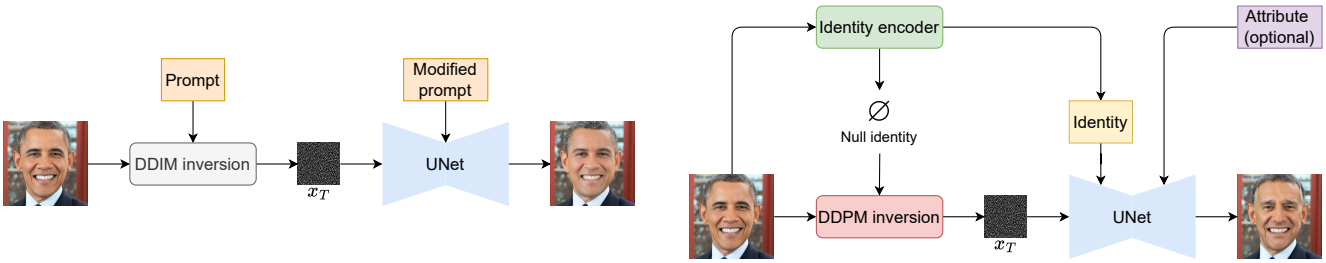
where $\hat{\mu}_t(x_t, x_{t+1}, \varnothing_{id})$ is the second-order estimate of the denoised sample, $\varnothing_{id}$ denotes the null identity embedding, and σ_t is the noise variance at step t.

Text-to-image diffusion models employ classifier-free guidance [30] to improve text fidelity by interpolating between a conditional and an unconditional prediction. In contrast, we reinterpret classifier-free guidance [30] to control identity alignment. Specifically, we condition one forward pass on the identity embedding and leave the other unconditioned. Our reverse personalization guidance is defined as:

$$\begin{aligned} \hat{\epsilon}_\theta(x_t, t, c_{id}) = \lambda_{cfg} \cdot \epsilon_\theta(x_t, t, c_{id}) \\ + (1 - \lambda_{cfg}) \cdot \epsilon_\theta(x_t, t, \varnothing_{id}), \end{aligned} \quad (5)$$

where λ_{cfg} is the guidance scale, c_{id} is the identity embedding, and $\varnothing_{id}$ is the null identity.

While standard personalization approaches use a positive guidance scale to reinforce identity, we instead apply a *negative guidance scale* to steer generation away from the provided identity. Our motivation comes from an observation: as the guidance scale increases, the model tends to enhance and exaggerate facial characteristics. This suggests



(a) Conventional prompt-based reweighting method.

(b) Our identity-based method.

Figure 3. Our reverse personalization framework. Unlike prior prompt-based reweighting approaches that guide both inversion and generation using textual prompts, our method leverages identity information to condition these processes. The framework also supports intuitive user control, enabling selective retention or modification of semantic facial attributes.

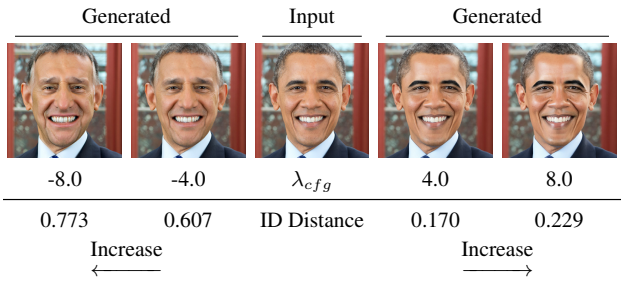


Figure 4. We visualize how varying the classifier-free guidance [30] scale affects identity preservation. Identity distance from the input increases with scale—especially in the negative direction—confirming a divergence from the input identity.

that negating the scale may attenuate those features. We explore this hypothesis and visualize the results in Fig. 4.

3.4. Attribute-controllable anonymization

Most anonymization methods do not allow users to choose which facial features to keep or change, limiting an important part of user control. However, research has shown that individuals' privacy preferences are shaped by a balance between perceived risks and potential benefits [64]. In some cases, retaining certain personal attributes can serve societal and individual goals—from improving healthcare [21] and workplace equity [10] to advancing research [4] and social justice [19]. To address this, our approach enables intuitive and consistent control over these attributes through natural language prompts. This is possible through our integration of image prompt adapters, which condition the diffusion model on visual identity cues without disrupting its interpretation of textual instructions.

To control facial attributes while anonymizing an image, we first invert the image to its latent noise representation x_T. During reverse diffusion process, we guide the model using a new textual prompt $\hat{c}_{attr}$ that specifies the desired facial attributes (e.g., "a young woman" or "an elderly man").

We modify the sampling equation by replacing the original attribute prompt c_{attr} with the updated prompt $\hat{c}_{attr}$, while reusing the original noise trajectory z_t. The sampling step becomes:

$$\hat{x}_{t-1} = \hat{\mu}_t(\hat{x}_t, \hat{x}_{t+1}, c_{id}, \hat{c}_{attr}) + \sigma_t z_t. \tag{6}$$

By substituting the previously estimated noise z_t, we obtain:

$$\begin{aligned}\hat{x}_{t-1} = \hat{\mu}_t(\hat{x}_t, \hat{x}_{t+1}, c_{id}, \hat{c}_{attr}) + x_{t-1} \\ - \hat{\mu}_t(x_t, x_{t+1}, \varnothing_{id}, c_{attr}).\end{aligned} \tag{7}$$

This formulation ensures that the anonymized output respects both the conditioning identity and the desired semantic attributes, allowing precise and interpretable control over the anonymization process.

4. Experiments

We evaluate our method across benchmarks, comparing it with prior face anonymization approaches in terms of identity removal, attribute preservation, image quality, and controllability. We also present ablation studies to analyze key design choices.

4.1. Hyperparameter analysis

Figure 5 presents a quantitative analysis of hyperparameters—classifier-free guidance [30] scale and IP-Adapter [85] scale—evaluated on CelebA-HQ [38] and FFHQ [39] datasets.

Figure 5a illustrates how the classifier-free guidance [30] scale influences re-identification rates, computed using two state-of-the-art face recognition models: SwinFace [66] and AdaFace [43]. When the guidance scale approaches zero, the generated faces closely resemble the original inputs, leading to higher re-identification rates. Conversely, as the guidance scale becomes more negative, the generated faces diverge further from the originals, resulting in lower re-identification rates.

While reducing the guidance scale lowers re-identification rates, excessively negative values degrade

	Re-ID (%) ↓				Attribute preservation ↓						Image quality			
	SwinFace		AdaFace		Expression		Gaze		Pose		FID ↓		Face IQA ↑	
	CHQ	FHQ	CHQ	FHQ	CHQ	FHQ	CHQ	FHQ	CHQ	FHQ	CHQ	FHQ	CHQ	FHQ
Ours	2.622	4.800	0.783	2.029	9.119	9.353	**0.152**	0.177	**0.050**	0.052	**4.809**	**8.651**	0.856	**0.921**
NullFace [50]	**0.489**	**0.844**	**0.157**	**0.358**	10.025	9.856	0.165	0.187	0.055	0.058	8.426	8.932	0.758	0.796
FAMS [49]	7.467	24.111	3.309	13.912	10.012	**8.847**	0.165	**0.176**	0.054	**0.049**	17.253	11.381	0.815	0.808
FALCO [6]	1.889	-	0.179	-	10.206	-	0.263	-	0.088	-	39.501	-	**0.875**	-
RiDDLE [51]	-	2.044	-	0.512	-	10.049	-	0.214	-	0.080	-	65.141	-	0.674
LDFA [45]	19.578	21.000	11.495	11.369	**8.653**	10.392	0.265	0.346	0.093	0.114	8.303	10.390	0.785	0.857
DP2 [34]	3.889	6.644	0.901	1.927	9.931	10.141	0.263	0.295	0.163	0.163	17.544	18.809	0.599	0.629

Table 1. Quantitative comparison of facial anonymization methods on CHQ (CelebA-HQ [38]) and FHQ (FFHQ [39]). The best result is marked in **bold**, and the second-best is underlined.

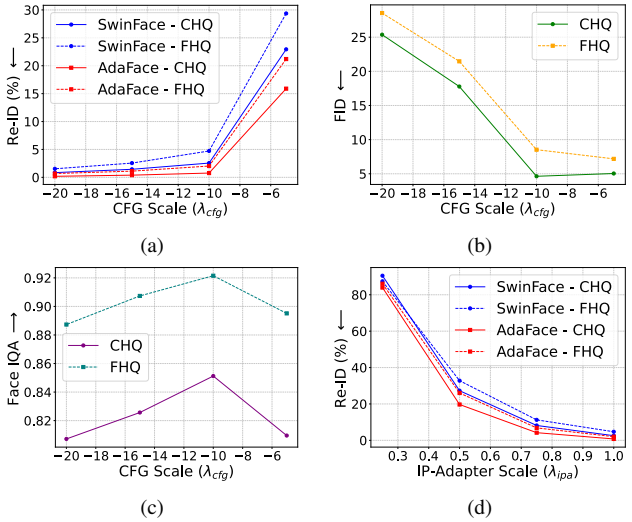


Figure 5. Hyperparameter analysis of classifier-free guidance [30] scale and IP-Adapter [85] scale on CHQ (CelebA-HQ [38]) and FHQ (FFHQ [39]).

image quality. As shown in Figs. 5b and 5c, overly negative guidance scales lead to higher Fréchet Inception Distance (FID) [29] and reduced scores from a face-specific image quality assessment (IQA) model [13].

Figure 5d examines the impact of the IP-Adapter [85] scale under a fixed negative guidance setting. This parameter controls the strength of identity embeddings. IP-Adapter [85] scale of 0.0 disables identity conditioning, leading to poor anonymization and high re-identification rates. As the scale increases, the model increasingly leverages the identity embeddings, enabling more effective modification of identity-specific features and a drop in re-identification rates.

4.2. Comparison with existing methods

We compare our method against six state-of-the-art facial anonymization approaches: NullFace [50], FAMS [49], FALCO [6], RiDDLE [51], LDFA [45], and DP2 [34], using CelebA-HQ [38] and FFHQ [39] datasets. We evaluated 4,500 subjects from each dataset. Our method was implemented on Stable Diffusion XL (SDXL) [65] for its superior image quality, with hyperparameters set to an IP-Adapter [85] scale of 1.0 and a classifier-free guidance [30] scale of –10.0. On an A100 GPU, generating a 1024×1024 image took approximately 13 seconds.

Evaluation was conducted across three dimensions: identity removal, attribute preservation, and image quality. Identity removal was measured by re-identification rates using SwinFace [66] and AdaFace [43]. To avoid bias, a different face recognition model [16] was used to extract identity embeddings during generation. Attribute preservation was assessed via three metrics: expression difference using a 3D face reconstruction model [17], pose difference using a head pose estimation network [69], and gaze difference using a gaze estimation model [3]. For image quality, we adopted the approach from previous studies [6, 15, 27, 56] by calculating FID [29]. Additionally, we utilized a face-specific IQA model [13].

As summarized in Tab. 1, our method and NullFace [50] exhibit balanced performance without weaknesses. FAMS [49] and LDFA [45] suffer from high re-identification rates, indicating poor anonymization. Despite leveraging StyleGAN2 [40], known for producing photorealistic images, FALCO [6] and RiDDLE [51] exhibit highest FID [29] scores—due to their inability to preserve background consistency, resulting in generated images that deviate from the originals. DP2 [34] struggles with pose preservation and image quality, likely because its inpainting strategy relies on pose estimation, which can be inaccurate, leading to generated faces misaligned with the original orientation. We also include privacy–utility trade-off plots in the supplementary material (Sec. 6) to visualize the relationship between identity removal, attribute preservation, and image quality across methods.

While NullFace [50] achieves the lowest re-identification rates, it underperforms in attribute preservation and image quality compared to our method. NullFace [50] samples

Figure 6. Qualitative comparison of anonymization results on CelebA-HQ [38].

Figure 7. Qualitative comparison of anonymization results on FFHQ [39].

anonymized identities in the embedding space of an identity encoder, whereas our approach operates in the latent space of diffusion models. A drawback of the former is that sampled embeddings may not always correspond to valid human faces, leading to lower Face IQA [13] scores.

Figures 6 and 7 present qualitative comparisons on CelebA-HQ [38] and FFHQ [39], respectively. Additional examples are provided in the supplementary material (Sec. 11).

4.3. Attribute-controllable anonymization

An advantage of our method is its ability to control high-level facial attributes in generated images. To demonstrate this capability, we conduct experiments on controlling three types of basic demographic information—age, sex, and race—so that the generated anonymized faces match those of the original images.

Among existing methods, DP2 [34] is the only one that supports attribute-guided anonymization. It does this by adapting StyleMC [46], which manipulates images along semantically meaningful directions in the GAN [23] latent space using a CLIP-based [67] loss guided by textual prompts. While enabling attribute control improves DP2 [34]'s accuracy in matching target attributes, its overall performance remains below that of our method.

	Age ↓		Sex (%) ↑		Race (%) ↑	
	CHQ	FHQ	CHQ	FHQ	CHQ	FHQ
Ours*	**3.744**	**4.314**	**99.485**	**98.638**	**87.016**	76.548
Ours†	4.931	5.854	92.663	82.720	60.882	49.479
NullFace [50]	5.461	6.017	87.113	77.023	62.784	56.048
FAMS [49]	6.293	6.598	29.802	58.704	32.749	44.621
FALCO [6]	5.016	-	89.138	-	66.014	-
RiDDLE [51]	-	6.144	-	73.437	-	54.039
LDFA [45]	4.426	5.888	88.713	77.401	67.572	59.238
DP2 [34]*	3.890	4.468	98.124	96.674	82.433	**83.033**
DP2 [34]†	4.822	5.627	87.523	79.750	63.951	58.865

* w/ attribute control † w/o attribute control

Table 2. Accuracy of preserving high-level facial attributes (age, sex, race) in anonymized faces.

In Tab. 2, our attribute-controlled anonymization achieves the highest accuracy in preserving all three attributes on CelebA-HQ [38], and two of the three attributes on FFHQ [39]. For the race attribute on FFHQ [39], DP2 [34] with attribute control slightly outperforms our method. Figure 8 presents qualitative examples that illustrate the effectiveness of attribute control, highlighting differences between preserving and altering these facial features.

4.4. Ablation study

To assess key design choices, we conducted an ablation study with two alternative implementations. First, we replaced our DDPM [31] inversion with DDIM [75]. Second, we substituted the SDXL [65] model with InstantID [78], a generation model designed for identity conditioning.

Quantitative results are in Tab. 3. The DDIM inversion [75] underperforms across all metrics. This is primarily due to DDIM's inability to reconstruct the original image [33, 60], leading to degraded attribute preservation and image quality. Furthermore, DDIM [75] exhibits poor performance to large adjustments in the classifier-free guidance [30] scale, where excessive changes introduce artifacts and cause image failures. Consequently, this implementation suffers from higher re-identification rates due to its limited capacity to support aggressive anonymization.

The InstantID-based implementation achieves lower re-identification rates than SDXL [65]. However, this results in reduced attribute preservation and higher FID [29]. This degradation is due to the ID embedding in InstantID, which entangles identity features with other facial attributes [78], making it difficult to preserve details unrelated to identity.

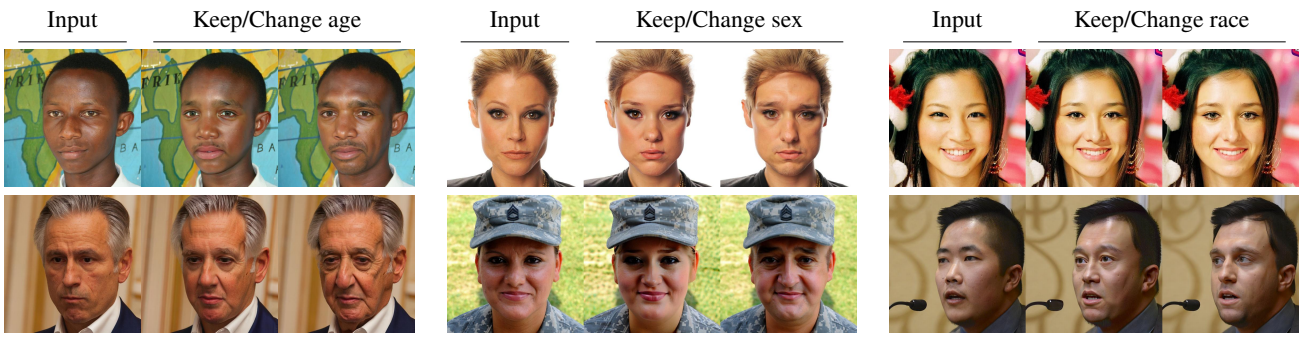


Figure 8. Qualitative comparison demonstrating the effect of attribute control on anonymized faces. Examples show the difference between preserving and altering age, sex, and race attributes.

	Re-ID (%) ↓				Attribute preservation ↓						Image quality			
	SwinFace		AdaFace		Expression		Gaze		Pose		FID ↓		Face IQA ↑	
	CHQ	FHQ	CHQ	FHQ	CHQ	FHQ	CHQ	FHQ	CHQ	FHQ	CHQ	FHQ	CHQ	FHQ
Ours	2.622	4.800	0.783	2.029	**9.119**	**9.353**	**0.152**	**0.177**	**0.050**	**0.052**	**4.809**	**8.651**	0.856	0.921
DDIM [75]	36.889	39.533	36.206	37.698	9.319	10.275	0.454	0.458	0.154	0.171	47.194	38.873	0.708	0.733
InstantID [78]	**0.489**	**0.911**	**0.202**	**0.781**	11.532	11.802	0.185	0.204	0.057	0.054	30.334	18.347	**0.889**	**0.947**

Table 3. Ablation study comparing our method with alternatives using DDIM inversion [75] and InstantID [78] as the generation backbone.

Qualitative comparisons in Fig. 9 further support these findings. The DDIM-based model fails to preserve facial attributes such as expression, gaze, and pose, and introduces artifacts. Meanwhile, the InstantID-based model tends to produce facial outputs appearing oversaturated and lacking realism.

We also evaluated the attribute control of these alternative implementations, results in Tab. 4. Neither approach matches our SDXL-based implementation in preserving high-level facial attributes.

	Age ↓		Sex (%) ↑		Race (%) ↑	
	CHQ	FHQ	CHQ	FHQ	CHQ	FHQ
Ours*	**3.744**	**4.314**	**99.485**	**98.638**	**87.016**	**76.548**
Ours†	4.931	5.854	92.663	82.720	60.882	49.479
DDIM [75]	4.257	5.413	78.701	71.016	62.285	55.889
InstantID [78]	16.009	15.160	83.038	78.038	19.600	28.517

* w/ attribute control † w/o attribute control

Table 4. Attribute control accuracy (age, sex, race) for alternative implementations. Neither DDIM inversion [75] nor InstantID [78] achieves attribute consistency comparable to our SDXL-based method.

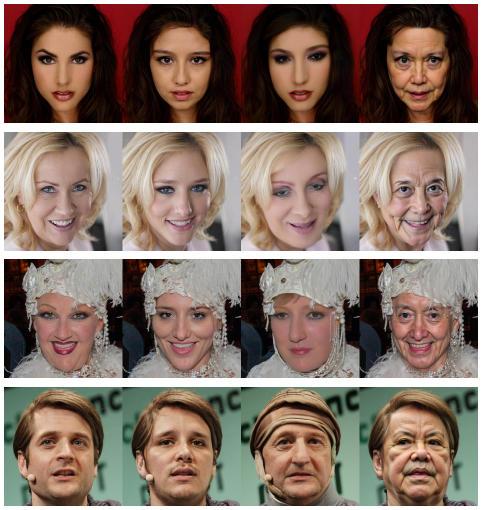


Figure 9. Qualitative results from the ablation study. Our method maintains both realism and accurate attribute preservation.

5. Conclusion

We proposed a reverse personalization framework for face anonymization that removes identity-specific features while preserving non-identity attributes. Our approach does not require the diffusion model to have prior exposure to the subject or any model fine-tuning, making it broadly applicable to arbitrary individuals. In addition to offering flexible anonymization, our method strikes an optimal balance among identity obfuscation, attribute preservation, and image quality.

Acknowledgments and Disclosure of Funding

This work was supported by the EU Horizon project "ELIAS - European Lighthouse of AI for Sustainability" (No. 101120237) and the FIS project GUIDANCE (Debugging Computer Vision Models via Controlled Cross-modal Generation) (No. FIS2023-03251). We acknowledge ISCRA for awarding this project access to the LEONARDO supercomputer, owned by the EuroHPC Joint Undertaking, hosted by CINECA (Italy), and thank the Finnish Foundation for Technology Promotion.

References

[1] California Consumer Privacy Act (CCPA). https://www.oag.ca.gov/privacy/ccpa/. 2

[2] General Data Protection Regulation (GDPR) Compliance Guidelines. https://gdpr.eu/. 2

[3] Ahmed A Abdelrahman, Thorsten Hempel, Aly Khalifa, Ayoub Al-Hamadi, and Laslo Dinges. L2cs-net: Fine-grained gaze estimation in unconstrained environments. In *2023 8th International Conference on Frontiers of Signal Processing (ICFSP)*, pages 98–102. IEEE, 2023. 6

[4] Claudia Acciai, Jesper W Schneider, and Mathias W Nielsen. Estimating social bias in data sharing behaviours: an open science experiment. *Scientific Data*, 10(1):233, 2023. 5

[5] Nouar AlDahoul, Talal Rahwan, and Yasir Zaki. Ai-generated faces influence gender stereotypes and racial homogenization. *Scientific Reports*, 15(1):14449, 2025. 2

[6] Simone Barattin, Christos Tzelepis, Ioannis Patras, and Nicu Sebe. Attribute-preserving face dataset anonymization via latent code optimization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8001–8010, 2023. 3, 6, 7, 1, 2, 4, 5

[7] Andreas Blattmann, Tim Dockhorn, Sumith Kulal, Daniel Mendelevitch, Maciej Kilian, Dominik Lorenz, Yam Levi, Zion English, Vikram Voleti, Adam Letts, et al. Stable video diffusion: Scaling latent video diffusion models to large datasets. *arXiv preprint arXiv:2311.15127*, 2023. 2

[8] Fadi Boutros, Jonas Henry Grebe, Arjan Kuijper, and Naser Damer. Idiff-face: Synthetic-based face recognition through fizzy identity-conditioned diffusion model. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 19650–19661, 2023. 1

[9] Manuel Brack, Felix Friedrich, Katharia Kornmeier, Linoy Tsaban, Patrick Schramowski, Kristian Kersting, and Apolinário Passos. Ledits++: Limitless image editing using text-to-image models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8861–8870, 2024. 1, 4

[10] Irene Browne and Joya Misra. The intersection of gender and race in the labor market. *Annual Review of Sociology*, 29(1):487–513, 2003. 5

[11] Zikui Cai, Zhongpai Gao, Benjamin Planche, Meng Zheng, Terrence Chen, M Salman Asif, and Ziyan Wu. Disguise without disruption: Utility-preserving face de-identification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 918–926, 2024. 3

[12] Jingyi Cao, Xiangyi Chen, Bo Liu, Ming Ding, Rong Xie, Li Song, Zhu Li, and Wenjun Zhang. Face de-identification: State-of-the-art methods and comparative studies. *arXiv preprint arXiv:2411.09863*, 2024. 2, 3

[13] Chaofeng Chen, Jiadi Mo, Jingwen Hou, Haoning Wu, Liang Liao, Wenxiu Sun, Qiong Yan, and Weisi Lin. Topiq: A top-down approach from semantics to distortions for image quality assessment. *IEEE Transactions on Image Processing*, 2024. 6, 7, 1

[14] Wenhu Chen, Hexiang Hu, Yandong Li, Nataniel Ruiz, Xuhui Jia, Ming-Wei Chang, and William W Cohen. Subject-driven text-to-image generation via apprenticeship learning. *Advances in Neural Information Processing Systems*, 36:30286–30305, 2023. 2

[15] Nicola Dall'Asen, Yiming Wang, Hao Tang, Luca Zanella, and Elisa Ricci. Graph-based generative face anonymisation with pose preservation. In *International Conference on Image Analysis and Processing*, pages 503–515. Springer, 2022. 3, 6

[16] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4690–4699, 2019. 6

[17] Yu Deng, Jiaolong Yang, Sicheng Xu, Dong Chen, Yunde Jia, and Xin Tong. Accurate 3d face reconstruction with weakly-supervised learning: From single image to image set. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019. 6

[18] Yingying Deng, Xiangyu He, Fan Tang, and Weiming Dong. Z-magic: Zero-shot multiple attributes guided image creator. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 18390–18400, 2025. 2

[19] Frank Edwards, Hedwig Lee, and Michael Esposito. Risk of being killed by police use of force in the united states by age, race–ethnicity, and sex. *Proceedings of the National Academy of Sciences*, 116(34):16793–16798, 2019. 5

[20] Emilio Ferrara. Genai against humanity: Nefarious applications of generative artificial intelligence and large language models. *Journal of Computational Social Science*, 7(1):549–569, 2024. 2

[21] Amelia Fiske, Sarah Blacker, Lester Darryl Geneviève, Theresa Willem, Marie-Christine Fritzsche, Alena Buyx, Leo Anthony Celi, and Stuart McLennan. Weighing the benefits and risks of collecting race and ethnicity data in clinical settings for medical artificial intelligence. *The Lancet Digital Health*, 7(4):e286–e294, 2025. 2, 5

[22] Rinon Gal, Yuval Alaluf, Yuval Atzmon, Or Patashnik, Amit H Bermano, Gal Chechik, and Daniel Cohen-Or. An image is worth one word: Personalizing text-to-image generation using textual inversion. *arXiv preprint arXiv:2208.01618*, 2022. 1, 2, 3, 4

[23] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in Neural Information Processing Systems*, 27, 2014. 7

[24] Mislav Grgic, Kresimir Delac, and Sonja Grgic. Scface–surveillance cameras face database. *Multimedia Tools and Applications*, 51:863–879, 2011. 2

[25] Zinan Guo, Yanze Wu, Zhuowei Chen, Lang Chen, Peng Zhang, and Qian He. Pulid: Pure and lightning id customization via contrastive alignment. *arXiv preprint arXiv:2404.16022*, 2024. 1, 3

[26] Fabio Hellmann, Silvan Mertes, Mohamed Benouis, Alexander Hustinx, Tzung-Chien Hsieh, Cristina Conati, Peter Krawitz, and Elisabeth André. Ganonymization: A gan-based face anonymization framework for preserving emotional expressions. *ACM Transactions on Multimedia Computing, Communications and Applications*, 21(1):1–27, 2024. 3

[27] Majed El Helou, Doruk Cetin, Petar Stamenkovic, and Fabio Zund. Vera: Versatile anonymization fit for clinical facial images. *arXiv preprint arXiv:2312.02124*, 2023. 3, 6

[28] Amir Hertz, Ron Mokady, Jay Tenenbaum, Kfir Aberman, Yael Pritch, and Daniel Cohen-Or. Prompt-to-prompt image editing with cross attention control. *arXiv preprint arXiv:2208.01626*, 2022. 1, 4

[29] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in Neural Information Processing Systems*, 30, 2017. 6, 7, 1

[30] Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. *arXiv preprint arXiv:2207.12598*, 2022. 2, 4, 5, 6, 7, 1

[31] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33:6840–6851, 2020. 3, 7

[32] Ziqi Huang, Kelvin CK Chan, Yuming Jiang, and Ziwei Liu. Collaborative diffusion for multi-modal face generation and editing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6080–6090, 2023. 1, 2

[33] Inbar Huberman-Spiegelglas, Vladimir Kulikov, and Tomer Michaeli. An edit friendly ddpm noise space: Inversion and manipulations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12469–12478, 2024. 1, 2, 7

[34] Håkon Hukkelås and Frank Lindseth. Deepprivacy2: Towards realistic full-body anonymization. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1329–1338, 2023. 3, 6, 7, 1, 2, 4, 5, 8

[35] Håkon Hukkelås and Frank Lindseth. Does image anonymization impact computer vision training? In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 140–150, 2023. 3

[36] Håkon Hukkelås, Rudolf Mester, and Frank Lindseth. Deepprivacy: A generative adversarial network for face anonymization. In *International Symposium on Visual Computing*, pages 565–578. Springer, 2019. 3, 2

[37] Jing Jiang and Weihong Deng. Disentangling identity and pose for facial expression recognition. *IEEE Transactions on Affective Computing*, 13(4):1868–1878, 2022. 3

[38] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196*, 2017. 5, 6, 7, 1, 2, 3, 4

[39] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4401–4410, 2019. 5, 6, 7, 1, 2, 8

[40] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8110–8119, 2020. 3, 6

[41] Furkan Kasım, Terrance E Boult, Rensso Mora, Bernardo Biesseck, Rafael Ribeiro, Jan Schlueter, Tom áš Repák, R Vareto, David Menotti, William Robson Schwartz, et al. Watchlist challenge: 3 rd open-set face detection and identification. In *2024 IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–10. IEEE, 2024. 2

[42] Daeha Kim and Byung Cheol Song. Optimal transport-based identity matching for identity-invariant facial expression recognition. *Advances in Neural Information Processing Systems*, 35:18749–18762, 2022. 3

[43] Minchul Kim, Anil K Jain, and Xiaoming Liu. Adaface: Quality adaptive margin for face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18750–18759, 2022. 5, 6, 1

[44] Minchul Kim, Feng Liu, Anil Jain, and Xiaoming Liu. Dcface: Synthetic face generation with dual condition diffusion model. In *Proceedings of the ieee/cvf Conference on Computer Vision and Pattern Recognition*, pages 12715–12725, 2023. 1

[45] Marvin Klemp, Kevin Rösch, Royden Wagner, Jannik Quehl, and Martin Lauer. Ldfa: Latent diffusion face anonymization for self-driving applications. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3199–3205, 2023. 1, 3, 6, 7, 2, 4, 5, 8

[46] Umut Kocasari, Alara Dirik, Mert Tiftikci, and Pinar Yanardag. Stylemc: Multi-channel based fast text-guided image generation and manipulation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 895–904, 2022. 7

[47] Zhenzhong Kuang, Xiaochen Yang, Yingjie Shen, Chao Hu, and Jun Yu. Facial identity anonymization via intrinsic and extrinsic attention distraction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12406–12415, 2024. 2

[48] Nupur Kumari, Bingliang Zhang, Richard Zhang, Eli Shechtman, and Jun-Yan Zhu. Multi-concept customization of text-to-image diffusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1931–1941, 2023. 2

[49] Han-Wei Kung, Tuomas Varanka, Sanjay Saha, Terence Sim, and Nicu Sebe. Face anonymization made simple. In *Proceedings of the Winter Conference on Applications of Com-*

puter Vision (WACV), pages 1040–1050, 2025. 3, 6, 7, 1, 2, 4, 5, 8

[50] Han-Wei Kung, Tuomas Varanka, Terence Sim, and Nicu Sebe. Nullface: Training-free localized face anonymization. *arXiv preprint arXiv:2503.08478*, 2025. 6, 7, 1, 2, 3, 4, 5, 8

[51] Dongze Li, Wei Wang, Kang Zhao, Jing Dong, and Tieniu Tan. Riddle: Reversible and diversified de-identification with latent encryptor. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8093–8102, 2023. 1, 3, 6, 7, 2, 8

[52] Zhen Li, Mingdeng Cao, Xintao Wang, Zhongang Qi, Ming-Ming Cheng, and Ying Shan. Photomaker: Customizing realistic human photos via stacked id embedding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8640–8650, 2024. 1, 3

[53] Hai Liu, Tingting Liu, Zhaoli Zhang, Arun Kumar Sangaiah, Bing Yang, and Youfu Li. Arhpe: Asymmetric relation-aware representation learning for head pose estimation in industrial human–computer interaction. *IEEE Transactions on Industrial Informatics*, 18(10):7107–7117, 2022. 3

[54] Renshuai Liu, Bowen Ma, Wei Zhang, Zhipeng Hu, Changjie Fan, Tangjie Lv, Yu Ding, and Xuan Cheng. Towards a simultaneous and granular identity-expression control in personalized face generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2114–2123, 2024. 3

[55] Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. Dpm-solver++: Fast solver for guided sampling of diffusion probabilistic models. *Machine Intelligence Research*, pages 1–22, 2025. 4

[56] Maxim Maximov, Ismail Elezi, and Laura Leal-Taixé. Ciagan: Conditional identity anonymization generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5447–5456, 2020. 3, 6

[57] Blaž Meden, Peter Rot, Philipp Terhörst, Naser Damer, Arjan Kuijper, Walter J Scheirer, Arun Ross, Peter Peer, and Vitomir Štruc. Privacy–enhancing face biometrics: A comprehensive survey. *IEEE Transactions on Information Forensics and Security*, 16:4147–4183, 2021. 2, 3

[58] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014. 3

[59] Hiroyuki Mishima, Hisato Suzuki, Michiko Doi, Mutsuko Miyazaki, Satoshi Watanabe, Tadashi Matsumoto, Kanako Morifuji, Hiroyuki Moriuchi, Koh-ichiro Yoshiura, Tatsuro Kondoh, et al. Evaluation of face2gene using facial images of patients with congenital dysmorphic syndromes recruited in japan. *Journal of Human Genetics*, 64(8):789–794, 2019. 2

[60] Ron Mokady, Amir Hertz, Kfir Aberman, Yael Pritch, and Daniel Cohen-Or. Null-text inversion for editing real images using guided diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6038–6047, 2023. 1, 3, 4, 7

[61] Jisu Nam, Soowon Son, Zhan Xu, Jing Shi, Difan Liu, Feng Liu, Seungryong Kim, and Yang Zhou. Visual persona: Foundation model for full-body human customization. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 18630–18641, 2025. 3

[62] Elaine M Newton, Latanya Sweeney, and Bradley Malin. Preserving privacy by de-identifying face images. *IEEE transactions on Knowledge and Data Engineering*, 17(2): 232–243, 2005. 3

[63] Foivos Paraperas Papantoniou, Alexandros Lattas, Stylianos Moschoglou, Jiankang Deng, Bernhard Kainz, and Stefanos Zafeiriou. Arc2face: A foundation model for id-consistent human faces. In *European Conference on Computer Vision*, pages 241–261. Springer, 2024. 1

[64] Katherine W Phillips, Nancy P Rothbard, and Tracy L Dumas. To disclose or not to disclose? status distance and self-disclosure in diverse environments. *Academy of Management Review*, 34(4):710–732, 2009. 2, 5

[65] Dustin Podell, Zion English, Kyle Lacey, Andreas Blattmann, Tim Dockhorn, Jonas Müller, Joe Penna, and Robin Rombach. Sdxl: Improving latent diffusion models for high-resolution image synthesis. *arXiv preprint arXiv:2307.01952*, 2023. 1, 6, 7, 2

[66] Lixiong Qin, Mei Wang, Chao Deng, Ke Wang, Xi Chen, Jiani Hu, and Weihong Deng. Swinface: A multi-task transformer for face recognition, expression recognition, age estimation and attribute estimation. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(4):2223–2234, 2023. 5, 6, 1

[67] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International Conference on Machine Learning*, pages 8748–8763. PMLR, 2021. 7

[68] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10684–10695, 2022. 1

[69] Nataniel Ruiz, Eunji Chong, and James M Rehg. Fine-grained head pose estimation without keypoints. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 2074–2083, 2018. 6

[70] Nataniel Ruiz, Yuanzhen Li, Varun Jampani, Yael Pritch, Michael Rubinstein, and Kfir Aberman. Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 22500–22510, 2023. 2

[71] Nataniel Ruiz, Yuanzhen Li, Varun Jampani, Wei Wei, Tingbo Hou, Yael Pritch, Neal Wadhwa, Michael Rubinstein, and Kfir Aberman. Hyperdreambooth: Hypernetworks for fast personalization of text-to-image models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6527–6536, 2024. 2

[72] Seyedmorteza Sadat, Otmar Hilliges, and Romann M Weber. Eliminating oversaturation and artifacts of high guidance scales in diffusion models. In *The Thirteenth International Conference on Learning Representations*, 2024. 1

[73] Kaede Shiohara and Toshihiko Yamasaki. Face2diffusion for fast and editable face personalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6850–6859, 2024. 3

[74] Kihyuk Sohn, Lu Jiang, Jarred Barber, Kimin Lee, Nataniel Ruiz, Dilip Krishnan, Huiwen Chang, Yuanzhen Li, Irfan Essa, Michael Rubinstein, et al. Styledrop: Text-to-image synthesis of any style. *Advances in Neural Information Processing Systems*, 36:66860–66889, 2023. 2

[75] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020. 2, 3, 7, 8

[76] Rebecca Umbach, Nicola Henry, Gemma Faye Beard, and Colleen M Berryessa. Non-consensual synthetic intimate imagery: Prevalence, attitudes, and knowledge in 10 countries. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, pages 1–20, 2024. 2

[77] Dani Valevski, Danny Lumen, Yossi Matias, and Yaniv Leviathan. Face0: Instantaneously conditioning a text-to-image model on a face. In *SIGGRAPH Asia 2023 Conference Papers*, pages 1–10, 2023. 2

[78] Qixun Wang, Xu Bai, Haofan Wang, Zekui Qin, Anthony Chen, Huaxia Li, Xu Tang, and Yao Hu. Instantid: Zero-shot identity-preserving generation in seconds. *arXiv preprint arXiv:2401.07519*, 2024. 1, 2, 3, 7, 8

[79] Yaoming Wang, Yangzhou Jiang, Jin Li, Bingbing Ni, Wenrui Dai, Chenglin Li, Hongkai Xiong, and Teng Li. Contrastive regression for domain adaptation on gaze estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19376–19385, 2022. 3

[80] Guangxuan Xiao, Tianwei Yin, William T Freeman, Frédo Durand, and Song Han. Fastcomposer: Tuning-free multi-subject image generation with localized attention. *International Journal of Computer Vision*, pages 1–20, 2024. 3

[81] Zhenyu Xie, Haoye Dong, Yufei Gao, Zehua Ma, and Xiaodan Liang. Dreamvton: Customizing 3d virtual try-on with personalized diffusion models. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pages 10784–10793, 2024. 3

[82] Jianqing Xu, Shen Li, Jiaying Wu, Miao Xiong, Ailin Deng, Jiazhen Ji, Yuge Huang, Guodong Mu, Wenjie Feng, Shouhong Ding, et al. Id^3: Identity-preserving-yet-diversified diffusion models for synthetic face recognition. *Advances in Neural Information Processing Systems*, 37: 77777–77798, 2024. 2, 3

[83] Haoxin Yang, Xuemiao Xu, Cheng Xu, Huaidong Zhang, Jing Qin, Yi Wang, Pheng-Ann Heng, and Shengfeng He. G 2 face: High-fidelity reversible face anonymization via generative and geometric priors. *IEEE Transactions on Information Forensics and Security*, 2024. 2

[84] Yahan Yang, Junfeng Lyu, Ruixin Wang, Quan Wen, Lanqin Zhao, Wenben Chen, Shaowei Bi, Jie Meng, Keli Mao, Yu Xiao, et al. A digital mask to safeguard patient privacy. *Nature Medicine*, 28(9):1883–1892, 2022. 2

[85] Hu Ye, Jun Zhang, Sibo Liu, Xiao Han, and Wei Yang. Ip-adapter: Text compatible image prompt adapter for text-to-image diffusion models. *arXiv preprint arXiv:2308.06721*, 2023. 2, 3, 5, 6

[86] Yu Zeng, Vishal M Patel, Haochen Wang, Xun Huang, Ting-Chun Wang, Ming-Yu Liu, and Yogesh Balaji. Jedi: Joint-image diffusion models for finetuning-free personalized text-to-image generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6786–6795, 2024. 3

[87] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3836–3847, 2023. 2