# HIPP: Protecting Image Privacy via High-Quality Reversible Protected Version

**Xi Ye**, **Lina Wang***, **Run Wang***, **Jiatong Liu** and **Geying Yang**

Key Laboratory of Aerospace Information Security and Trusted Computing, Ministry of Education, School of Cyber Science and Engineering, Wuhan University, China

{xixiye, lnwang, wangrun, liujiatong, yanggeying}@whu.edu.cn,

## Abstract

With the rapid development of the internet, sharing photos through Social Network Platforms (SNPs) has become a new way for people to socialize, which poses serious threats to personal privacy. Recently, a thumbnail-preserving image privacy protection technique has emerged and garnered widespread attention. However, the existing schemes based on this technique often introduce noticeable noise into the protected image, resulting in poor visual quality. Motivated by the observation that a latent vector can be decoupled into the detail and contour components, in this paper, we propose HIPP, a thumbnail-preserving image privacy protection scheme that decouples the detail and contour information contained in the latent vector corresponding to the original image and reconstructs details by generation model. As a result, the generated protected image appears natural and has a thumbnail similar to the original one. Moreover, the protected images can be restored to versions that are indistinguishable from the original images. Experiments on CelebA, Helen, and LSUN datasets show that the SSIM between the restored and original images achieves 0.9899. Furthermore, compared to the previous works, HIPP achieves the lowest runtime and file expansion rate, with values of 0.07 seconds and 1.1046, respectively.

## 1 Introduction

In recent years, with the rapid development of the internet, more and more people tend to share various images about their lives on Social Network Platforms (SNPs). In 2023, there were 5.07 billion social media users worldwide [Petrosyan, 2024], with approximately 2 billion users uploading photos on Instagram monthly [Dixon, 2024] and around $243,055$ new photos being uploaded to Facebook every minute [Kevin, 2024]. Although SNPs provide convenient sharing and communication, bringing users closer together, they have introduced a range of privacy issues as images often contain users' personal private information such as identity and location [Such and Criado, 2018; Morales *et al.*, 2021; Fan, 2019; Zeng *et al.*, 2015; Li *et al.*, 2024]. For one thing, malicious users may analyze the content of images on SNPs to obtain private information of users [Barnes, 2006; Narayanan and Shmatikov, 2009]. For another, SNPs themselves may be curious about the user information contained in the uploaded images [Isaak and Hanna, 2018; Mazzarolo *et al.*, 2021; Ra *et al.*, 2013]. As traditional image encryption ensures that all information in the original image is invisible, providing maximum privacy protection but limiting the sharing experience of non-sensitive information [Singh and Singh, 2022; Krishna *et al.*, 2022; Deng *et al.*, 2023]. Partial image protection makes detected sensitive areas invisible while keeping the resting unchanged for original visual effect preservation [Beugnon *et al.*, 2019; Morris *et al.*, 2023; Zhang *et al.*, 2023a; He *et al.*, 2024a]. Unfortunately, the sensitive areas in images are prone to omission, posing a risk of privacy leakage. Another popular technology for privacy protection is named face anonymization[Maxim *et al.*, 2020; Li *et al.*, 2023; He *et al.*, 2024b], which works by modifying facial features to alter the appearance. However, this technol-

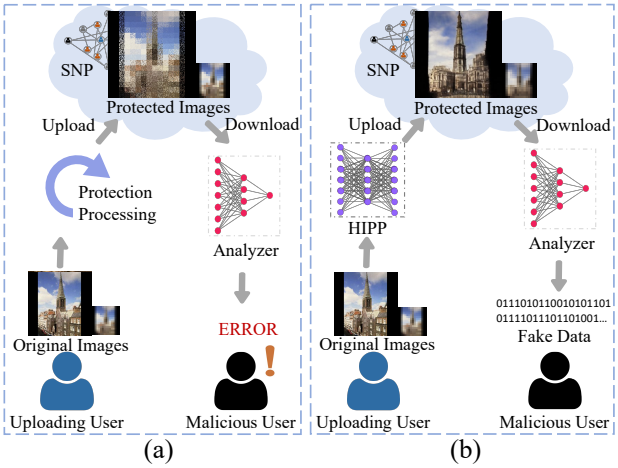


Figure 1: The overview of the difference between previous works and HIPP. (a) Previous works. The protected image processed by these schemes appears noisy and cannot be analyzed for valid data. (b) HIPP. The generated protected image appears natural, misleading malicious users into obtaining fake data during analyzing.

*Lina Wang and Run Wang are corresponding authors

ogy only focus on faces, ignoring that other content in shared images may reveal sensitive personal information such as current location and home address.

In view of this, a novel thumbnail-preserving image privacy protection technique has been proposed, which is usually categorized into two types: 1) complete thumbnail preservation [Wright *et al.*, 2015; Tajik *et al.*, 2019; Chai *et al.*, 2022; Zhang *et al.*, 2022c]; 2) approximate thumbnail preservation [Marohn *et al.*, 2017; Zhang *et al.*, 2022b; Ye *et al.*, 2023; Zhang *et al.*, 2023b; Chowdhury *et al.*, 2024; Zhao *et al.*, 2024]. On the one hand, the technique protects all image details within thumbnail blocks by disturbing the specific pixel values to prevent malicious users from extracting privacy information from the details of images displayed on SNPs. On the other hand, it preserves the overall visual effect of the original image for online browsing and management by ensuring that the protected image has a same or approximate thumbnail to the original one. As shown in Figure 1(a), the uploading user, i.e., the image owner, processes the original images locally and uploads the protected images into SNP. The malicious user, who is curious about the privacy information of other users, may collect a large number of images displayed on SNP but cannot analyze the detail privacy information within them. Although the existing works achieve the goal of protecting image privacy while retaining overall visual effect, the generated protected image often contains a lot of noise as shown in Figure 1(a), leading to poor visual perception and poor sharing experience. Meanwhile, malicious users can easily aware that the image is protected and target it for attacks.

Motivated by the observation that the latent vector corresponding to the image can be decoupled into the detail and contour parts, in this paper, we propose HIPP, a reversible scheme based on latent vector decoupling to achieve image privacy protection for social images sharing. Unlike the existing schemes, our scheme applies a detail information extractor to decouple the vector into the detail and contour vectors in the latent space instead of directly utilizing encryption or other obfuscation techniques in the image space. The new vector is composed of the original contour vector and the replaced detail vector, which can be utilized to generate the protected image with high visual quality by generation models. Moreover, as protected images are indistinguishable from the real images, the malicious user cannot aware that the protected images displayed in SNP are fake images and analyze the fake privacy data. The main contributions of our scheme proposed in this paper are as follows:

- We propose a novel reversible image privacy protection scheme that generates naturally appealing protected images that are indistinguishable from real natural images, improving visual quality for better sharing experience and misleading malicious users into obtaining incorrect personal private information.
- We design a detail information extractor that can decouple the detail and contour information contained in the latent vector corresponding to the image. Meanwhile, generation models are applied to reconstructing the protected images with different details, which provides a new perspective for future research in the area of thumbnail-preserving image privacy protection.
- Experimental results demonstrate that the generated protected images appear natural without any visible noise. Moreover, our scheme achieves the lowest runtime and file expansion rate compared to existing methods, improving its practical value.

## 2 Related Work

### Image Privacy Protection for SNPs

As the earliest method for image privacy protection, traditional image encryption schemes [Singh and Singh, 2022] encrypt the whole image into a snowflake-like ciphertext by the delivered key, preserving no visual information related to the original image and limiting content sharing. Consequently, a partial image privacy protection approach, which only protects areas related to privacy within the image, has gained popularity. Initially, these areas are protected by simple techniques such as blurring (adding noise) [Neustaedter and Greenberg, 2003; Neustaedter *et al.*, 2006], pixelation (replacing pixels in the same sub block with their average value) [Lander *et al.*, 2001; Zhou and Pun, 2021], masking (covering all pixels by a specific value) [Park and Kim, 2020; Yu *et al.*, 2020], or object removal (filling with new content related to surrounding pixels) [Yi *et al.*, 2020; Wang *et al.*, 2021], which do not take multiple privacy conflicts into account. To address this, a hierarchical privacy protection scheme for privacy image sharing was proposed, allowing the reconstruction of the original version only when a sufficient number of individuals consent [Beugnon *et al.*, 2019]. In another approach, Liu et al. designed a system to achieve automated facial protection by learning sensitive relationships among users [Liu *et al.*, 2022]. Subsequently, the issue of privacy leakage caused by image forwarding still persisted, which was later addressed by a cross-platform protection scheme based on blockchain technology [Zhang *et al.*, 2022a; Zhang *et al.*, 2023a]. A crucial aspect of partial image protection mentioned above is the detection of privacy-sensitive areas within the image. However, no existing approach can guarantee that all sensitive areas can be found.

### Thumbnail-Preserving Image Privacy Protection

The objective of thumbnail-preserving image privacy protection is to make details within thumbnail blocks invisible while preserving the whole image thumbnail unchanged, which was first introduced by Wright et al. (2015). This scheme divides each thumbnail block into several sub blocks, and then rearranges pixels in every sub block, followed by rearranging sub blocks in the same thumbnail block which is sample but exposes the pixel lists of thumbnail blocks. Taijik et al. 2019 designed a sum-preserving encryption (SPE) algorithm to encrypt every two pixels while keeping their sum unchanged. Therefore, the sum of each block after encryption is completely identical to the original one. Although the scheme achieves nonce-respecting security, its time cost is quite high. Hence, a scheme named TPE-GAN was proposed to improve the efficiency by applying CycleGAN [Zhu *et al.*, 2017] to simulate randomized unary encoding [Chai *et al.*, 2022]. At

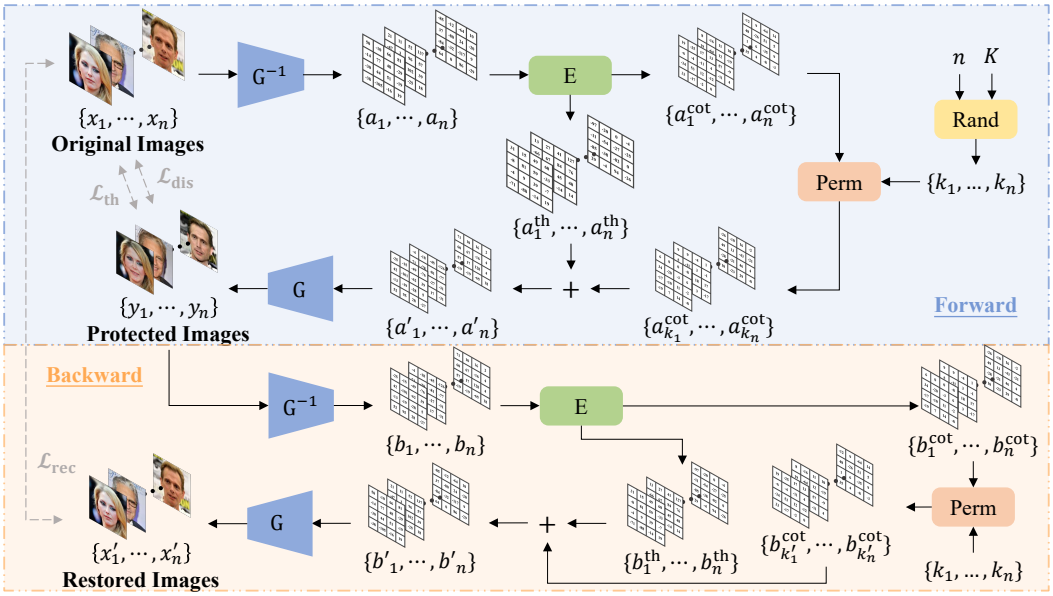


Figure 2: The overview of HIPP. Specifically, G and $G^{-1}$ form the latent-to-image mapper, E represents the detail information extractor. Rand and Perm are functions of random sequence generation and permutation, respectively. $n$ denotes the number of images are processed at once while $K$ is the key owned by the uploading user.

the same time, F-TPE came up with a multi-pixel SPE algorithm, achieving fast calculation on vector sets [Zhang *et al.*, 2022c].

Compared with completely thumbnail-preserving privacy protection schemes mentioned above, approximately thumbnail-preserving schemes achieve high efficiency at the expense of similarity between the protected and original thumbnails. Morohn et al. (2017) proposed two schemes named DRPE and LSB-TPE, which have possibility of decryption failures or incomplete recovery. Therefore, HF-TPE was proposed to ensure successful decryption, with the decrypted image being closer to the original one by constructing a sum-preserving data embedding algorithm [Zhang *et al.*, 2022b]. Subsequently, Ye et al. (2023) and Zhang et al. (2023b) proposed two completely reversible schemes to confirm that the restored image is identical to the original one by combining reversible data hiding with traditional encryption. To further reduce the running time, P$w$LM$\mu$ applies a chatic system with enhanced dynamics by $\mu$ [Chowdhury *et al.*, 2024]. Zhao et al. proposed a usability enhanced scheme, which utilizes SPE to encrypt the lowest seven bits while adjusting the protected thumbnail and storing extra information in the most significant bit of each pixel. While the existing works mentioned above perform well in preserving the original thumbnail, they all generate the protected images with noticeable noise.

## 3 Proposed Scheme

In HIPP, We aim to protect the privacy of image details from both malicious users and their analysis models while retaining the overall visual effect of the original image to ensure the protected image remains natural. Here, image details means any information within thumbnail blocks such as facial feature and building appearance. As images shared in SNP are often displayed in thumbnail form for coarse-grained content presentation, HIPP reconstructs image details for privacy protection, and preserves the thumbnail as unchanged as possible for image contour preservation.

### 3.1 Overview

Figure 2 illustrates the overall framework of the proposed HIPP, which consists of two primary modules: the latent-to-image space mapper (G and $G^{-1}$) and the detail information extractor E. Specifically, the $n$ original images are transformed into the latent vectors $\{a_1, \cdots, a_n\}$ by $G^{-1}$. And then, E divides the vector $a_i$ into two components: $a_i^{\text{th}}$ and $a_i^{\text{cot}}$, encapsulating the detail and contour information of the image $x_i$, respectively. A random sequence $\{k_1, \cdots, k_n\}$ is generated by $K$ and utilized to permute $a_i^{\text{cot}}$, where $k_i \in [1, n]$ and $k_i \neq k_j$ $(i, j \in [1, n], i \neq j)$. The permuted detail vector $a_{k_i}^{\text{cot}}$ is combined with $a_i^{\text{th}}$ to form the new latent vectors $a'_i$. Finally, the protected image $y_i$ is obtained by applying the generator G with $a'_i$ as input. The protected image $y_i$ has a thumbnail similar to $x_i$, but the details are different. The backward section in Figure 2 represents the inverse process of HIPP, where $k'_{k_i} = i$, meaning the order of $\{a'_1, \cdots, a'_n\}$ is exactly the same as that of $\{b_{k'_1}, \cdots, b_{k'_n}\}$. The restored image $x'_i$ is basically consistent with the original image $x_i$. Additionally, maintaining the correct order of the protected images is crucial for successful restoration. This order can be securely encoded into image filenames or embedded within the protected images themselves via data hiding methods.

### 3.2 Latent-to-Image Space Mapper

To extract detail information from images in the latent space, a space mapper, consisting of an image generator $G(\cdot) : \mathcal{Z} \rightarrow$

$\mathcal{I}$ and its inverse $\mathrm{G}^{-1}(\cdot) : \mathcal{I} \to \mathcal{Z}$, is necessary to establish a mapping between the latent and image spaces. The latent vector corresponding to the original image $x$ is obtained by

$$a = \mathrm{G}^{-1}(x), \tag{1}$$

the reconstructed image $\tilde{x}$ is acquired by

$$\tilde{x} = G(a) = G\left(G^{-1}(x)\right), \tag{2}$$

and the latent vector corresponding to $\tilde{x}$ is derived by

$$\tilde{a} = G^{-1}(\tilde{x}) = G^{-1}\left(G\left(G^{-1}(x)\right)\right). \tag{3}$$

Since the protected image $y$ is expected to be restored to a version as similar to the original image $x$ as possible, it is desirable to minimize the difference between $a$ and $\tilde{a}$ as well as between $x$ and $\tilde{x}$.

In HIPP, we choose Glow [Kingma and Dhariwal, 2018] and Generative Adversarial Network (GAN) for image generation, each with its own strengths. For Glow, the whole processing procedure is completely reversible, enabling $x$ and $\tilde{x}$, as well as $a$ and $\tilde{a}$, to be perfectly consistent without any loss. For GAN, we apply StyleGAN [Karras *et al.*, 2019] as $\mathrm{G}(\cdot)$ and in-domain GAN inversion as $\mathrm{G}^{-1}$ [Zhu *et al.*, 2020], which cannot achieve full reversibility but is capable of generating higher-quality protected images. The specific comparison results between Glow and GAN are presented in detail in the experimental section.

### 3.3 Detail Information Extractor

The detail information extractor E is crucial for both the generation of the protected image and the restoration of the original image, which aims to decouple image detail information from contour information within the latent vector $a$ by

$$a^{\mathrm{cot}}, a^{\mathrm{th}} = \mathrm{E}(a). \tag{4}$$

Here, $a^{\mathrm{cot}}$ is the detail vector and $a^{\mathrm{th}}$ denotes the contour vector, both having the same dimensions as $a$. To specifically demonstrate the role of E in HIPP, we assume that we have the original images $x_1$ and $x_2$ now, with their corresponding latent vectors $a_1$ and $a_2$, respectively. After inputting $a_1$ and $a_2$ into E, we obtain $a_1^{\mathrm{cot}}$, $a_1^{\mathrm{th}}$, $a_2^{\mathrm{cot}}$, and $a_2^{\mathrm{th}}$. Then, two new vectors are derived by

$$\begin{cases} a'_1 = a_2^{\mathrm{cot}} + a_1^{\mathrm{th}} \\ a'_2 = a_1^{\mathrm{cot}} + a_2^{\mathrm{th}} \end{cases}, \tag{5}$$

which are subsequently inputted into G to generate the protected images $y_1$ and $y_2$. $a'_1$ contains contour information from $a_1$ and detail information from $a_2$. Consequently, the generated protected image $y_1$ has a contour similar to $x_1$, but different details. In the backward process, we only need to perform a decoupling operation on the corresponding latent vector $b_1$ of $y_1$, which is identical to $a'_1$ under Glow or approximately equal to $a'_1$ under GAN:

$$b_1^{\mathrm{cot}}, b_2^{\mathrm{th}} = \mathrm{E}(b_1). \tag{6}$$

Here, $b_1^{\mathrm{cot}}$ and $b_1^{\mathrm{th}}$ are basically the same as $a_2^{\mathrm{cot}}$ and $a_1^{\mathrm{th}}$, respectively. Therefore, we can obtain $b'_1$ (which is essentially consistent with $a_1$) by

$$b'_1 = b_2^{\mathrm{cot}} + b_1^{\mathrm{th}} \tag{7}$$

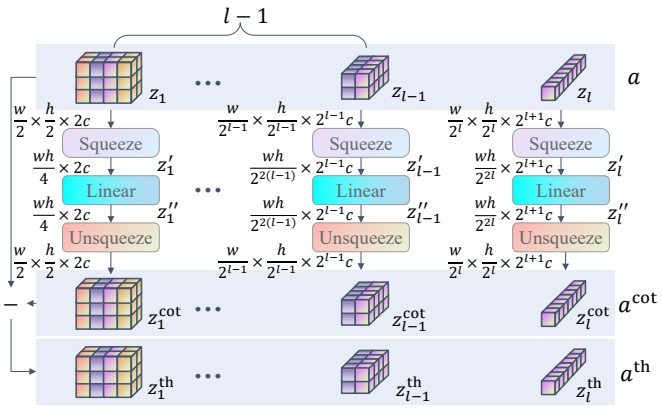


Figure 3: The architecture of the detail information extractor. $w$, $h$, and $c$ represent the width, height, and channel number of the original image $x$, respectively. $\{z_1, \cdots, z_l\}$ form the latent vector $a$ corresponding to $x$, where $l$ is the scale of flow in Glow.

and get the restored image $x'_1$ by

$$x'_1 = \mathrm{G}(b'_1). \tag{8}$$

As $a$, the latent vector mapped from $x$ by Glow, is composed of multiple sub-vectors $[z_1, \cdots, z_l]$ ($l$ denotes the scale of flow in Glow) of different dimensions, it is necessary to perform detail information extraction operations on each sub-vector separately as illustrated in Figure 3. First, $z_1$ with size of $w_i \times h_i \times \times c_i$ is reshaped into $z'_1$ of size $w_i h_i \times c_i$. $z''_i$ is generated by

$$z''_i = \mathrm{Linear}(z'_i), \tag{9}$$

where $\mathrm{Linear}(\cdot)$ represents a full connected layer with identical input and output dimensions. $z_i^{\mathrm{cot}}$ is obtained through reshaping $z''_i$, and then, $z_i^{\mathrm{th}}$ is derived through

$$z_i^{\mathrm{th}} = z_i - z_i^{\mathrm{cot}}. \tag{10}$$

For GAN, the vector $a$ has size of $2\log_2 \frac{h}{2} \times 512$ and reshaped into size of $1024 \log_2 \frac{h}{2}$ before $\mathrm{Linear}(\cdot)$ operation. Overall, E aims to decouple $a$ into $a^{\mathrm{cot}}$ and $a^{\mathrm{th}}$ without changing the dimensions.

### 3.4 Loss Function

The ultimate goal of the proposed scheme is to train the detail information extractor E to decouple the detail and contour information within $a$ as much as possible under frozen space mapper. The specific targets are: 1) the generated protected image $y$ is expected to have a contour similar to the original image $x$; 2) $y$ has details different from $x$; 3) the restored version of the protected image $x'$ is as close to $x$ as possible. Therefore, the optimization objectives of HIPP are

$$\begin{cases} \min\limits_{\mathrm{E}} \mathrm{Dif}\big(\mathrm{Thumb}(x, b), \mathrm{Thumb}(y, b)\big) \\ \max\limits_{\mathrm{E}} \mathrm{Dif}(x, y) \\ \min\limits_{\mathrm{E}} \mathrm{Dif}(x, x') \end{cases}. \tag{11}$$

Here, $\mathrm{Thumb}(x, b)$ sets the thumbnail block size to $b \times b$ and returns a thumbnail $t_x$ of $x$ with the same size. More pre-

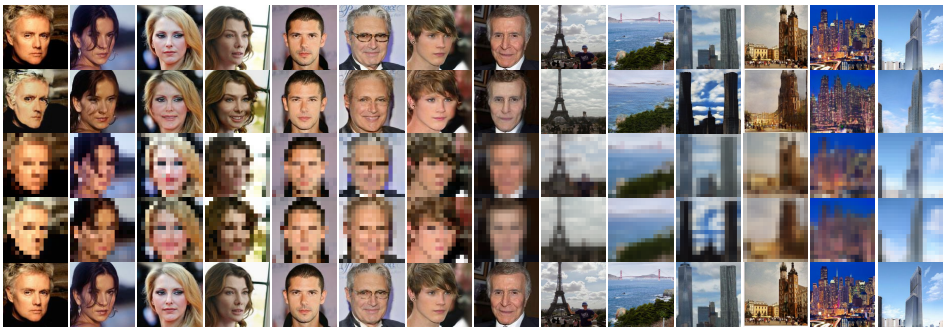

Figure 4: The visual performance of images processed by HIPP under different mappers. The five rows display the original images, the protected images, the original thumbnails, the protected thumbnails, and the restored images, respectively.

cisely, each pixel in the thumbnail is calculated by

$$t_x(i,j,p) = \frac{\sum_{s=\alpha}^{\alpha+b-1}\left(\sum_{q=\beta}^{\beta+b-1} x(s,q,p)\right)}{b \times b}, \tag{12}$$

where $i \in [1,w]$, $j \in [1,h]$, $p \in [1,c]$, $\alpha = \lceil \frac{i}{b} \rceil + 1$, and $\beta = \lceil \frac{j}{b} \rceil + 1$. And $\mathrm{Dif}(x,y)$ is utilized to evaluate the difference between $x$ and $y$. A smaller Dif value indicates greater similarity and less difference between the two inputted images. However, E cannot be directly trained with the above optimization objectives since it operates in the latent space $\mathcal{Z}$. The objectives must be transformed into loss functions related to the latent vectors. During training, a Gaussian-sampled vector $r$ with the same dimension as $a$ is used, and the loss functions shown in Figure 2 are introduced as follows.

*1) Detail and Contour Loss:* To achieve the first and second objectives, it is necessary to extract as much detail information from the vector as possible without including contour information. As $x$ and its corresponding thumbnail share the same contour information, we minimize

$$\mathcal{L}_{\text{th}} = \mathrm{MSE}(a^{\text{th}}, a^{\text{th}}_{\text{th}}) \tag{13}$$

to ensure that $a^{\text{th}}$ contains as little detail information as possible, where $a = \mathrm{G}^{-1}(x)$ and $a_{\text{th}} = \mathrm{G}^{-1}(\mathrm{Thumb}(x))$. Here, $\mathrm{MSE}(x,y)$ returns the Mean Squared Error (MSE) of the inputs, which measures the difference between them. The formula for MSE is

$$\mathrm{MSE}(x,y) = \frac{1}{whc}\sum_{p=1}^{c}\sum_{i=1}^{w}\sum_{j=1}^{h}(x(i,j,p) - y(i,j,p))^2. \tag{14}$$

Additionally, $\mathcal{L}_{\text{dis}}$, calculated by

$$\mathcal{L}_{\text{dis}} = \mathrm{MSE}(a', a), \tag{15}$$

is minimized to ensure that $a_{\text{th}}$ contains sufficient contour information. Here, $a' = r^{\text{cot}} + a^{\text{th}}$. Overall, $a^{\text{cot}}$ is expected to retain as much detail information as possible while reducing the likelihood of containing contour information by simultaneously minimizing losses $\mathcal{L}_{\text{th}}$ and $\mathcal{L}_{\text{dis}}$.

*2) Reconstruction Loss:* For the third objective, it is necessary to make the contour and detail vectors as independent of each other as possible so that E is able to extract the same detail vector from $a$ and $r' = a^{\text{cot}} + r^{\text{th}}$. Therefore, the loss

$$\mathcal{L}_{\text{rec}} = \mathrm{MES}(a'^{cot}, r^{cot}) + \mathrm{MSE}(r'^{cot}, a^{cot}) \tag{16}$$

is minimized to enhance the reconstruction ability, ensuring that $a'^{cot}$ and $r'^{cot}$ extracted during the backward process are as close as possible to $r^{cot}$ and $a^{cot}$ during the forward process, respectively. To this end,

$$\tilde{a} = a'^{\text{th}} + r'^{\text{cot}} \tag{17}$$

is basically the same as $a$, making $x$ as similar to $x'$ as possible.

Finally, the whole loss function is calculated by

$$\mathcal{L} = \mathcal{L}_{\text{th}} + \lambda_1\mathcal{L}_{\text{dis}} + \lambda_2\mathcal{L}_{\text{rec}}, \tag{18}$$

where $\lambda_1$ and $\lambda_2$ are hyperparameters used to adjust the relative weights of the loss terms $\mathcal{L}_{\text{dis}}$ and $\mathcal{L}_{\text{rec}}$ in the overall loss function $\mathcal{L}$. The optimization objective of HIPP is to minimize the whole loss function $\mathcal{L}$.

## 4 Experiments

### 4.1 Experimental Setup

**Datasets.** In the experiments, 50000 images are randomly sampled from CelebA [Liu *et al.*, 2015], Helen [Le *et al.*, 2012], and LSUN[Yu *et al.*, 2015] datasets and resized to $128 \times 128$ and $256 \times 256$ to form the training set. Additionally, we select 2000 images from the remaining images to form a testing set.

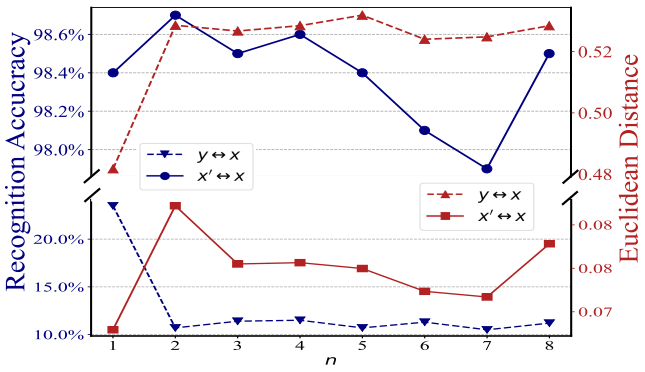


Figure 5: Euclidean distance and face recognition accuracy between different images. $x$, $y$, $x'$ represent the original, protected, and restored images, respectively.

| Block Size | $8 \times 8$ | $16 \times 16$ | $32 \times 32$ | $64 \times 64$ |
|---|---|---|---|---|
| TP-JPEG | **1.0** | **1.0** | **1.0** | **1.0** |
| ATPE | 0.6812 | 0.6437 | 0.5604 | 0.4147 |
| BIPU | **1.0** | **1.0** | **1.0** | **1.0** |
| TPE-GAN | 0.7796 | 0.5117 | 0.2942 | 0.1993 |
| F-TPE | **1.0** | **1.0** | **1.0** | **1.0** |
| HF-TPE | 0.9311 | 0.8647 | 0.8582 | 0.6954 |
| UE-JPEG | 0.7106 | 0.7106 | 0.7106 | 0.7106 |
| FVPP | <u>0.9635</u> | <u>0.9657</u> | 0.9727 | 0.9824 |
| P$w$LM$\mu$ | 0.9554 | 0.9348 | 0.9127 | 0.8653 |
| HIPP-Glow | 0.7542 | 0.8295 | 0.8932 | 0.9444 |
| HIPP-GAN | 0.8425 | 0.9318 | <u>0.9739</u> | <u>0.9868</u> |

Table 1: The SSIM value between thumbnails of the protected and original images under different schemes.We mark the top-2 results by bold and underlying.

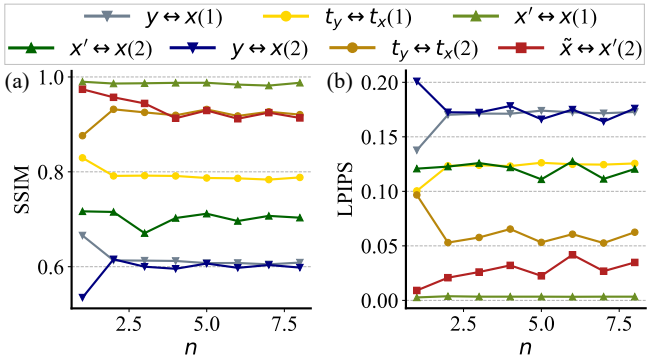


Figure 6: The values of SSIM and LPIPS between different images under different mappers. (a) SSIM. (b) LPIPS values. (1) means the results under Glow while (2) represents the results under GAN. $t_x$ and $t_y$ represent the thumbnails of the original image $x$ and the protected image $y$, respectively, while $\tilde{x} = \mathrm{G}\left(\mathrm{G}^{-1}(x)\right)$.

**Implementation Details.** During the training procedure of E, we directly apply the pretrained G and $\mathrm{G}^{-1}$ models, keeping their parameters frozen. Meanwhile, an Adam optimizer with $\beta_1 = 0, \beta_2 = 0.99, \epsilon = 10^{-8}$ is applied, the learning rate and iteration are set to $10^{-5}$ and 200000, respectively. As for hyperparameters $\lambda_1$ and $\lambda_2$, they are set to 10 and 200, respectively, when applying Glow as G. If utilizing StyleGAN as G and in-domain GAN inversion as $\mathrm{G}^{-1}$, the values of $\lambda_1$ and $\lambda_2$ are changed into 10 and 1000, respectively. In addition, when $n$ is set to 1, a random sampled latent vector $r$ is utilized to assist in transforming $a$ into $a'$ as described in the loss function section. When $n > 1$, the specific process is illustrated in Figure 2.

**Evaluation Metrics.** There are three indicators utilized in experiments to assess the performance of HIPP: Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM) [Wang *et al.*, 2004], and Learned Perceptual Image Patch Similarity (LPIPS) [C.-V.Yang *et al.*, 2014]. Specifically, PSNR quantifies the difference in intensity between the two images from a pixel-wise perspective. The higher the PSNR value, the smaller the difference between images. SSIM and LPIPS are both perceptual metrics that quantify the similarity between the two images. The SSIM value ranges from 0 to 1, with higher values indicating greater similarity. Conversely, the LPIPS value operates oppositely, where lower values signify higher similarity between the images.

**Baselines.** To comprehensively evaluate the performance of the proposed scheme, we select ten baselines for comparison: TP-JPEG [Wright *et al.*, 2015], ATPE [Marohn *et al.*, 2017], BIPU [Tajik *et al.*, 2019], TPE-GAN [Chai *et al.*, 2022], F-TPE [Zhang *et al.*, 2022c], HF-TPE [Zhang *et al.*, 2022b], FVPP [Zhang *et al.*, 2023b], UE-JPEG [Ye *et al.*, 2023], P$w$LM$\mu$ [Chowdhury *et al.*, 2024], and PR3 [Zhao *et al.*, 2024].

## 4.2 Experimental Results

### Evaluation of Detail Privacy Protection

The privacy objective of HIPP is reconstructing details within each thumbnail block to maximize the inconsistency between the protected and original images. In the specific evaluation experiments, we extract facial feature vectors from the two images and then calculate the Euclidean distance between these vectors by $d(p,q) = \sqrt{\sum_i (q_i - p_i)^2}$, where $p_i$ and $q_i$ denote the coordinates of points $p$ and $q$ on the $i$-th dimension, respectively. As shown by the red section in Figure 5, the distance between face feature vectors of the images remains at a high level (the Eucldean distance between two unrelated facial vectors is approximately 0.836) regardless of the value of $n$. Furthermore, we apply a tool named *face_recognition*[1] to judge whether faces in the input images belong to the same person. As shown by the blue section in Figure 5, the facial recognition accuracy of the protected images is quite low, which means that the proposed HIPP successfully reconstructs details in the original image. Meanwhile, as shown in Figure 6, the protected image $y$ is not similar to the original image $x$ from a perceptual perspective, indicating that unrelated details are successfully reconstructed by our scheme.

### Evaluation of Image Contour Preservation

Figure 4 provides a visual demonstration of HIPP. It is obvious that the protected images generated by HIPP shown in the second row remains a natural version without visible noise. Meanwhile, thumbnails of the original and protected images are quite similar as displayed in the third and fourth rows. In addition, we apply SSIM and LPIPS to evaluate the similarity between thumbnails of the protected and original images, as displayed in Figure 6 and Table 1. The SSIM value reaches its peak when the block size is $64 \times 64$ under HIPP-GAN, which is higher than that in other approximately thumbnail-preserving works. Although the protected thumbnail under HIPP cannot be made entirely identical to the original one, it is sufficiently similar for users to associate it with the original image during online browsing. Figure 6 indicates that the similarity value reaches its highest when $n = 1$ and shows no significant change as $n$ increases. Overall, the protected images generated by HIPP not only preserve the general contour of the original ones but also appear natural and are indistinguishable from realistic images.

[1]https://github.com/ageitgey/face_recognition

| Loss | HIPP-GAN | | | | | | HIPP-Glow | | |
|---|---|---|---|---|---|---|---|---|---|
| | $t_x \leftrightarrow t_y \uparrow$ | $x \leftrightarrow y \downarrow$ | $x \leftrightarrow x' \uparrow$ | $t_{\tilde{x}} \leftrightarrow t_y \uparrow$ | $\tilde{x} \leftrightarrow y \downarrow$ | $\tilde{x} \leftrightarrow x' \uparrow$ | $t_x \leftrightarrow t_y \uparrow$ | $x \leftrightarrow y \downarrow$ | $x \leftrightarrow x' \uparrow$ |
| $\mathcal{L}_{\text{th}}$ | 0.08841 | 0.02931 | 0.73266 | 0.11478 | 0.03238 | 1 | 0.24896 | 0.10123 | 0.99972 |
| $\mathcal{L}_{\text{dis}}$ | 0.84648 | 0.80269 | 0.71271 | 1 | 0.99997 | 0.99991 | 0.99996 | 0.99978 | 0.99978 |
| $\mathcal{L}_{\text{rec}}$ | 0.55233 | 0.29279 | 0.70299 | 0.65819 | 0.46142 | 0.99526 | 0.75279 | 0.73096 | 0.9998 |
| $\mathcal{L}_{\text{th}} + \mathcal{L}_{\text{dis}}$ | 0.57656 | 0.31004 | 0.09762 | 0.60235 | 0.40878 | 0.14392 | 0.4347 | 0.28993 | 0.44608 |
| $\mathcal{L}_{\text{th}} + \mathcal{L}_{\text{rec}}$ | 0.07365 | 0.02286 | 0.75240 | 0.06179 | 0.01664 | 0.99225 | 0.28529 | 0.14202 | 0.99982 |
| $\mathcal{L}_{\text{dis}} + \mathcal{L}_{\text{rec}}$ | 0.89701 | 0.86269 | 0.71269 | 0.99980 | 0.99915 | 0.99737 | 0.99995 | 0.99977 | 0.99977 |
| $\mathcal{L}_{\text{th}} + \mathcal{L}_{\text{dis}} + \mathcal{L}_{\text{rec}}$ | 0.73589 | 0.34157 | 0.76281 | 0.85804 | 0.5586 | 0.96253 | 0.82951 | 0.66583 | 0.98992 |

Table 2: The SSIM values of images under different training loss functions. $t_{\tilde{x}}$ represent the thumbnails of $\tilde{x}$.

| Scheme | TP-JPEG | ATPE | BIPU | F-TPE | HF-TPE | UE-JPEG | FVPP | PR3 | P$w$LM$\mu$ | HIPP-Glow | HIPP-GAN |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Size Expansion↓ | 2.32 | 2.10 | 1.73 | 3.30 | 2.11 | 1.68 | 3.99 | 2.27 | 2.20 | **1.10** | <u>1.16</u> |
| Run Time (s)↓ | 1.26 | 2.04 | 92.6 | 43.3 | 107.9 | 3.46 | 17.5 | 110.6 | 6.08 | **0.07** | <u>0.88</u> |

Table 3: The size expansion rate of the protected image and the running time of the whole processing procedure.We mark the top-2 results by bold and underlying.

**Evaluation of Reversibility**

Although HIPP cannot fully restore the protected images into their original versions, the restored images exhibit a very high similarity to the original ones as shown in Figure 4 and 6. As in-domain GAN inversion cannot perfectly map the latent vector to the image without any loss, we introduce a new comparison object $\tilde{x} = \mathrm{G}\left(\mathrm{G}^{-1}(x)\right)$ to isolate the impact of the latent-to-image mapper on the evaluation of E. Additionally, we evaluate the quality of the restored images by *face_recognition* with a threshold set to $0.3$ as shown in Figure 5. Even under stringent threshold, the probability of identifying the original and restored images as belonging to the same person remains very high. Moreover, as shown in Table 4, the restored image under HIPP-Glow is almost identical to the original version. For HIPP-GAN, the results are less satisfactory due to the inherent non-inevitability of GAN. However, employing a more outstanding GAN inversion as $\mathrm{G}^{-1}$ can improve the restoration quality.

| Scheme | TP-JPEG | ATPE | TPE-GAN | HF-TPE | HIPP-Glow | HIPP-GAN |
|---|---|---|---|---|---|---|
| PSNR↑ | 42.14 | <u>53.57</u> | 26.37 | **59.35** | 37.47 | 22.23 |
| SSIM↑ | 0.946 | 0.933 | 0.918 | <u>0.964</u> | **0.9899** | 0.7628 |

Table 4: The PSNR and SSIM value between the restored and original images under different schemes. We mark the top-2 results by bold and underlying.

**Ablation Study**

To analyze the indispensability of each component in $\mathcal{L}$, an ablation study is conducted by retraining E on different loss functions as shown in Table 2. Here, the final goal of training process is enlarging values of SSIM$(t_x, t_y)$, SSIM$(x, x')$, SSIM$(t_{\tilde{x}}, t_y)$, and SSIM$(\tilde{x}, x)$ while reducing values of SSIM$(x, y)$ and SSIM$(\tilde{x}, y)$. Specifically, $\mathcal{L}_{\text{th}}$ becomes $0$ when $a^{\text{cot}} = a$ and $a^{\text{th}} = 0$, $\mathcal{L}_{\text{dis}} = \text{MSE}(a', a) = \text{MSE}(a^{\text{cot}} + r^{\text{th}}, a^{\text{cot}} + a^{\text{th}})$ turns to $0$ when$a^{\text{cot}} = 0$ and $a^{\text{th}} = a$, and $\mathcal{L}_{\text{rec}}$ changes to $0$ when $\mathrm{E}(a) = a, 0$ or $\mathrm{E}(a) = 0, a$. Results show in Table 2 indicate that HIPP successfully achieves the training objectives if and only if the loss function is set to $\mathcal{L} = \mathcal{L}_{\text{th}} + \mathcal{L}_{\text{dis}} + \mathcal{L}_{\text{rec}}$.

**Efficiency Comparison**

To evaluate the running time in detail, we process $1000$ images at once and record the average running time per image as shown in Table 3. Compared to previous works, HIPP achieves a significant improvement in running time, saving processing time for users and enhancing its practicality. Meanwhile, the size expansion rate is calculated by comparing the file size of the protected image with that of the original image. Due to the absence of visible noise in the protected image, the images can be effectively compressed, resulting in a much smaller file size compared to other works. This not only improves the image uploading speed but also saves storage space.

## 5 Conclusion

In this paper, we propose a reversible image privacy protection scheme by preserving thumbnails, named HIPP, which can generate the protected image without any visible noise. The scheme maps the original image to the corresponding latent vector by the inversion of a generation model, decouples the vector into detail and contour vectors, and replaces the detail vector to reconstruct image details. Experimental results indicate that the generated protected image appears natural and has different details compared to the original version while preserving the whole contour unchanged. Furthermore, HIPP significantly reduces both runtime and file size of the protected image, thereby decreasing the time required for local processing and uploading, which enhances the practicability. In future work, we will explore ways to further decouple the image detail information from the contour information, aiming to achieve a higher level of privacy protection.

## Acknowledgements

This work was supported by National Natural Science Foundation of China (NSFC) under Grants No, 62372334, No. 62202340, the Fundamental Research Funds for the Central Universities under No. 2042025kf0054, the Natural Science Foundation of Hubei Province under No. 2025AFB455, the CCF-NSFOCUS 'Kunpeng' Research Fund under No. CCF-NSFOCUS 2023005, the Open Foundation of Henan Key Laboratory of Cyberspace Situation Awareness under No. HNTS2022004.

## References

[Barnes, 2006] S. B Barnes. A privacy paradox: Social networking in the united states. *First Monday*, 2006.

[Beugnon *et al.*, 2019] S. Beugnon, P. Puteaux, and W. Puech. Privacy protection for social media based on a hierarchical secret image sharing scheme. In *Proc. IEEE Int. Conf. Image Process. (ICIP)*, pages 679–683, 2019.

[C.-V.Yang *et al.*, 2014] C.-V.Yang, M. Chao, and M.-H. Yang. Single-image super-resolution: A benchmark. In *Proc. Eur. Conf. Comput. Vis.*, pages 372–386, 2014.

[Chai *et al.*, 2022] X. Chai, Y. Wang, X. Chen, Z. Gan, and Y. Zhang. TPE-GAN: Thumbnail preserving encryption based on GAN with key. *IEEE Signal Process. Lett.*, 29:972–976, 2022.

[Chowdhury *et al.*, 2024] K. Chowdhury, S. Deb, N. Kar, J. L. Sarkar, A. H. Alkhayyat, and L. Nkenyereye. Pwlm$\mu$-tpe: A visual-usability and privacy-enhanced thumbnail preserving encryption scheme of cloud computing for consumers. *IEEE Trans. Consum. Electron.*, pages 1–1, 2024.

[Deng *et al.*, 2023] X. Deng, C. Gao, and M. Xu. Pirnet: Privacy-preserving image restoration network via wavelet lifting. In *2023 IEEE/CVF Int. Conf. Computer Vis. (ICCV)*, pages 22311–22320, 2023.

[Dixon, 2024] X. Dixon. Instagram - statistics & facts. Technical report, Statista, 2024.

[Fan, 2019] L. Fan. A demonstration of image obfuscation with provable privacy. In *Proc. 2019 IEEE Int. Conf. Multimed. Expo Workshops (ICMEW)*, pages 608–608. IEEE, 2019.

[He *et al.*, 2024a] X. He, L. Li, F. Tong, and H. Peng. Multilevel privacy protection for social media based on 2-d compressive sensing. *IEEE Int. Things J.*, 11(4):6878–6892, 2024.

[He *et al.*, 2024b] X. He, M. Zhu, D. Chen, N. Wang, and X. Gao. Diff-privacy: Diffusion-based face privacy protection. *IEEE Trans. Circuits Syst. Video Technol.*, 34(12):13164–13176, 2024.

[Isaak and Hanna, 2018] J. Isaak and M. J Hanna. User data privacy: Facebook, cambridge analytica, and privacy protection. *Computer*, 51(8):56–59, 2018.

[Karras *et al.*, 2019] T. Karras, S. Laine, and T. Aila. A style-based generator architecture for generative adversarial networks. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pages 4401–4410, 2019.

[Kevin, 2024] H. Kevin. 41 up-to-date facebook facts and stats. Technical report, Wishpond, 2024.

[Kingma and Dhariwal, 2018] D. P Kingma and P. Dhariwal. Glow: Generative flow with invertible 1x1 convolutions. *Proc. Adv. Neural Inf. Process. Syst.*, 31, 2018.

[Krishna *et al.*, 2022] K. Krishna, R. Satyabrata, R. Umashankar, and M. Shashwat. Iehc: An efficient image encryption technique using hybrid chaotic map. *Chaos Solit. Fractals*, 158:111994, 2022.

[Lander *et al.*, 2001] K. Lander, V. Bruce, and H. Hill. Evaluating the effectiveness of pixelation and blurring on masking the identity of familiar faces. *Appl. Cognitive Psychol.*, 15(1):101–116, 2001.

[Le *et al.*, 2012] V. Le, J. Brandt, L. Bourdev, and T. S. Huang. Interactive facial feature localization. In *Proc. Eur. Conf. Comput. Vis.*, pages 679–692. Springer, 2012.

[Li *et al.*, 2023] D. Li, W. Wang, K. Zhao, and J. Dong. Riddle: Reversible and diversified de-identification with latent encryptor. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, page 8093–8102. IEEE, June 2023.

[Li *et al.*, 2024] B. Li, y. Wei, Y. Fu, Z. Wang, Y. Li, J. Zhang, R. Wang, and T. Zhang. Towards reliable verification of unauthorized data usage in personalized text-to-image diffusion models. In *Proc. IEEE Symp. Secur. Priv.*, 2024.

[Liu *et al.*, 2015] Z. Liu, P. Luo, X. Wang, and X. Tang. Deep learning face attributes in the wild. In *Proc. Int. Conf. Comput. Vis. (ICCV)*, December 2015.

[Liu *et al.*, 2022] J. Liu, L. Li, and N. Li. Learning and preserving relationship privacy in photo sharing. In *Proc. IEEE/ACM Int. Conf. Big Data Comput. Appl. Technol. (BDCAT)*, pages 170–173, 2022.

[Marohn *et al.*, 2017] B. Marohn, C. V. Wright, W. Feng, M. Rosulek, and R. B. Bobba. Approximate thumbnail preserving encryption. In *Proc. Multimed. Priv. Secur.*, pages 33–43. Association for Computing Machinery, 2017.

[Maxim *et al.*, 2020] M. Maxim, E. Ismail, and L.-T. Laura. Ciagan: Conditional identity anonymization generative adversarial networks. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, page 5446–5455. IEEE, June 2020.

[Mazzarolo *et al.*, 2021] G. Mazzarolo, J. C. F. Casas, A. D. Jurcut, and N.-A. Le-Khac. Protect against unintentional insider threats: The risk of an employee's cyber misconduct on a social media site. In *Cybercrime in context: The human factor in victimization, offending, and policing*, pages 79–101. Springer, 2021.

[Morales *et al.*, 2021] A. Morales, J. Fierrez, R. Vera-Rodriguez, and R. Tolosana. SensitiveNets: Learning agnostic representations with application to face images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 43(6):2158–2164, 2021.

[Morris *et al.*, 2023] J. Morris, S. Newman, K. Palaniappan, J. Fan, and D. Lin. "do you know you are tracked by photos

that you didn't take": Large-scale location-aware multi-party image privacy protection. *IEEE Trans. Dependable Secure Comput.*, 20(1):301–312, 2023.

[Narayanan and Shmatikov, 2009] A. Narayanan and V. Shmatikov. De-anonymizing social networks. In *Proc. IEEE Symp. Secur. Priv.*, pages 173–187. IEEE, 2009.

[Neustaedter and Greenberg, 2003] C. Neustaedter and S. Greenberg. The design of a context-aware home media space for balancing privacy and awareness. In *Proc. Int. Conf. Ubiquitous Comput.*, pages 297–314. Springer, 2003.

[Neustaedter *et al.*, 2006] C. Neustaedter, S. Greenberg, and M. Boyle. Blur filtration fails to preserve privacy for home-based video conferencing. *ACM Trans. Hum. Comput. Interact.*, 13(1):1–36, 2006.

[Park and Kim, 2020] S. Park and G. Kim. Block-based masking region relocation and detection method for image privacy masking. In *Proc. Int. Conf. Inf. Commun. Technol. Convergence*, pages 1586–1588. IEEE, 2020.

[Petrosyan, 2024] A. Petrosyan. Number of internet and social media users worldwide as of april 2024. Technical report, Statista, 2024.

[Ra *et al.*, 2013] M.-R. Ra, R. Govindan, and A. Ortega. P3: Toward Privacy-Preserving photo sharing. In *Proc. USENIX Symp. Netw. Syst. Des. Implementation*, pages 515–528, Lombard, IL, 2013. USENIX Association.

[Singh and Singh, 2022] K. N. Singh and A. F. Singh. Towards integrating image encryption with compression: A survey. *ACM Trans. Multimedia Comput. Commun. Appl. (TOMM)*, 18(3):1–21, 2022.

[Such and Criado, 2018] J. M Such and N. Criado. Multiparty privacy in social media. *Commun. ACM*, 61(8):74–81, 2018.

[Tajik *et al.*, 2019] K. Tajik, A. Gunasekaran, R. Dutta, B. Ellis, R. B. Bobba, M. Rosulek, C. V. Wright, and W. Feng. Balancing image privacy and usability with thumbnail-preserving encryption. In *Proc. Symp. Netw. Distrib. Syst. Secur. (NDSS)*, page 295, 2019.

[Wang *et al.*, 2004] Z. Wang, A. C Bovik, H. R Sheikh, and E. P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE trans. image process.*, 13(4):600–612, 2004.

[Wang *et al.*, 2021] Z.-M. Wang, M.-H. Li, and G.-S. Xia. Conditional generative convnets for exemplar-based texture synthesis. *IEEE Trans. Image Process.*, 30:2461–2475, 2021.

[Wright *et al.*, 2015] C. V. Wright, W. Feng, and F. Liu. Thumbnail-preserving encryption for JPEG. In *Proc. ACM Workshop Inf. Hiding Multimedia Secur.*, pages 141–146, 2015.

[Ye *et al.*, 2023] X. Ye, Y. Zhang, X. Xiao, S. Yi, and R. Lan. Usability enhanced thumbnail-preserving encryption based on data hiding for jpeg images. *IEEE Signal Process. Lett.IEEE Signal Process. Lett.*, 30:793–797, 2023.

[Yi *et al.*, 2020] Z. Yi, Q. Tang, S. Azizi, D. Jang, and Z. Xu. Contextual residual aggregation for ultra high-resolution image inpainting. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pages 7508–7517, 2020.

[Yu *et al.*, 2015] F. Yu, Y. Zhang, S. Song, A. Seff, and J. Xiao. Lsun: Construction of a large-scale image dataset using deep learning with humans in the loop. *arXiv preprint arXiv:1506.03365*, 2015.

[Yu *et al.*, 2020] J. Yu, M. Wu, C. Li, and S. Zhu. A street view image privacy detection and protection method based on mask-RCNN. In *Proc. IEEE Jt. Int. Inf. Technol. Artif. Intell. Conf.*, volume 9, pages 2184–2188. IEEE, 2020.

[Zeng *et al.*, 2015] Y. Zeng, Y. Sun, L. Xing, and V. Vokkarane. A study of online social network privacy via the tape framework. *IEEE J. of Sel. Topics Signal Process.*, 9(7):1270–1284, 2015.

[Zhang *et al.*, 2022a] M. Zhang, Z. Sun, H. Li, B. Niu, F. Li, Y. Xie, and C. Zheng. A blockchain-based privacy-preserving framework for cross-social network photo sharing. In *Proc. IEEE Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, pages 1–6, 2022.

[Zhang *et al.*, 2022b] Y. Zhang, R. Zhao, X. Xiao, R. Lan, Z. Liu, and X. Zhang. HF-TPE: High-fidelity thumbnail-preserving encryption. *IEEE Trans. Circuits Syst. Video Technol.*, 32(3):947–961, 2022.

[Zhang *et al.*, 2022c] Y. Zhang, W. Zhou, R. Zhao, X. Zhang, and X. Cao. F-TPE: Flexible thumbnail-preserving encryption based on multi-pixel sum-preserving encryption. *IEEE Trans. Multimedia*, pages 1–15, 2022.

[Zhang *et al.*, 2023a] M. Zhang, Z. Sun, H. Li, B. Niu, F. Li, Z. Zhang, Y. Xie, and C. Zheng. Go-sharing: A blockchain-based privacy-preserving framework for cross-social network photo sharing. *IEEE Trans. Dependable Secure Comput.*, 20(5):3572–3587, 2023.

[Zhang *et al.*, 2023b] Y. Zhang, X. Ye, X. Xiao, T. Xiang, H. Li, and X. Cao. A reversible framework for efficient and secure visual privacy protection. *IEEE Trans. Inf. Forensics Secur.*, 18:3334–3349, 2023.

[Zhao *et al.*, 2024] R. Zhao, Y. Zhang, W. Wen, X. Zhang, X. Cao, and Y. Xiang. Pr3: Reversible and usability-enhanced visual privacy protection via thumbnail preservation and data hiding. *IEEE Trans. Big Data*, pages 1–14, 2024.

[Zhou and Pun, 2021] Jizhe Zhou and Chi-Man Pun. Personal privacy protection via irrelevant faces tracking and pixelation in video live streaming. *IEEE Trans. Inf. Forensics Secur.*, 16:1088–1103, 2021.

[Zhu *et al.*, 2017] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proc. Int. Conf. Comput. Vis. (ICCV)*, pages 2242–2251, 2017.

[Zhu *et al.*, 2020] J. Zhu, Y. Shen, D. Zhao, and B. Zhou. In-domain gan inversion for real image editing. In *Proc. Eur. Conf. Comput. Vis.*, pages 592–608. Springer, 2020.