Pre-capture Privacy via Adaptive Single-Pixel Imaging

Yoko Sogabe Shiori Sugimoto Ayumi Matsumoto Masaki Kitahara
NTT Corporation, Japan
{yoko.sogabe, shiori.sugimoto, ayumi.matsumoto, masaki.kitahara}@ntt.com

Abstract

As cameras become ubiquitous in our living environment, invasion of privacy is becoming a significant concern. A common approach to privacy preservation is to remove personally identifiable information from a captured image, but there is a risk of the original image being leaked. In this paper, we propose a pre-capture privacy-aware imaging method that captures images from which the details of a pre-specified anonymized target have been eliminated. The proposed method applies a single-pixel imaging framework in which we introduce a feedback mechanism called an aperture pattern generator (APG). The introduced APG adaptively outputs the next aperture pattern to avoid sampling the anonymized target by using already acquired data as a clue. Furthermore, the anonymized target can be set to any object without changing hardware. Except for the removed detailed features of the anonymized target, the captured images are of comparable quality to those captured by a general camera and can be used for various computer vision applications. We target faces and license plates and experimentally show that the proposed method can capture clear images in which detailed features of the anonymized target are eliminated, achieving both privacy and utility.

1. Introduction

As a result of technological innovations in networking, semiconductors, computer vision, and more, cameras have become ubiquitous in our living environment. The use of such cameras with computer vision is expected to have various practical applications. However, the widespread use of cameras raises concerns about privacy and may be subject to social backlash and legal restrictions. Thus, to promote the utilization of computer vision, it is necessary to overcome privacy and utility trade-offs.

A common approach to privacy preservation is to remove personal data from the captured image data after capturing. However, there is a risk that the data before removal may be leaked. Such an approach in which personal data is removed after capturing is called post-capture privacy. In contrast, pre-capture privacy is an approach based on computational

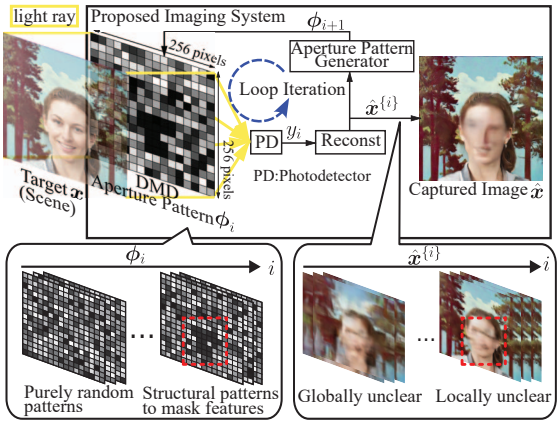

Figure 1. Overview of proposed imaging system. The system captures a single image gradually by repeatedly acquiring the incident light through the aperture pattern multiple times based on a single-pixel imaging framework. The entire reconstructed captured image gradually becomes clear. The next aperture pattern is generated to avoid sampling the anonymized target by using the current unclear reconstructed captured image. This feedback mechanism, which is proposed in this paper, results in the optical elimination of the anonymized target.

imaging in which personal data is not captured (either optically or at the sensor level), which ultimately enhances the level of security. Prior studies on pre-capture [27, 43] used thermal cameras to estimate the location of the face to avoid sampling it. These imaging systems were designed by focusing on face anonymization, and it is difficult to apply them to anonymize anything other than faces. With cameras in public places, however, there is a wide variety of objects that should not be captured, i.e., *anonymized targets*. Examples include faces, textual information (license plates), fingerprints, and irises.

In this paper, we propose a pre-capture privacy-aware imaging method that captures images in which the details of the anonymized target are optically eliminated. An aperture pattern generator (APG) is introduced in a single-pixel imaging framework. The APG implicitly estimates the location of the anonymized target from the unclear image, which is reconstructed from the already acquired data, and outputs the next aperture pattern to avoid sampling that location.

Assuming that the anonymized targets are either faces or license plates, our quantitative experiments by simulations show that both privacy and utility can be achieved. In addition, a prototype imaging system is assembled to verify its application in the real world.

Our contributions are as follows:

Pre-capture privacy. We introduce imaging that optically excludes detailed features of the anonymized target by adaptively controlling the aperture, achieving anonymization at the point of capture.

Utility. The captured anonymized images are of comparable quality to those captured by a general camera, except for the local degradation of the anonymized target, and can be used for computer vision tasks.

Variability of anonymized targets. The anonymized target can be any object other than a face. In this case, the aperture pattern generation network only needs to be retrained using the existing pre-trained recognition model for the anonymized target, without changing the hardware.

2. Related Work

Traditionally, the approach to privacy preservation has been post-capture privacy, but recent advances in computational imaging have made pre-capture privacy possible. In relation to our work, we outline methods for pre-capture privacy and computational imaging techniques that are closely related to privacy preservation.

Pre-capture privacy for face. The most difficult part of pre-capture privacy is determining the location of the anonymized target before capturing. In [43] and [27], a thermal camera is used to estimate the location of faces. The thermal camera detects a face silhouette by assuming the temperature of faces. Another camera, which can control the shutter pixel-by-pixel, captures an anonymized image by turning off the shutter at the detected silhouette. The captured images are natural, except for the faces, which are masked. Therefore, they can be used in any computer vision application. However, such imaging systems have been designed by focusing on face anonymization, and it is difficult to apply them to anonymizing anything other than faces.

Pre-capture privacy for specific applications. Some studies have achieved pre-capture privacy by capturing an image that can only be used for specific applications but the image is globally degraded to the point that personal data is unrecognizable. In [26], moderately degraded images are captured with a defocus lens attached to a camera to blur captured images. This preserves privacy while using the camera for a specific application such as full-body motion tracking. In [35], an imaging system combining optical convolution, and pooling and quantization in sensor circuits is developed. By jointly optimizing HW parameters with a face detector and recognizer, the system enables face detection while limiting identification. In [14], a lens's point spread function and human pose estimation network are jointly trained in an end-to-end fashion. This makes it possible to degrade private attributes while maintaining important features for human pose estimation. In [37], an end-to-end trained phase mask is inserted into the aperture plane to capture an image that is strongly blurred to protect privacy while enabling depth estimation. In [4], the coded aperture on a lensless camera and classifier network are jointly trained in an end-to-end fashion. This makes it difficult for a malicious user to reconstruct the image while still being suitable for the trained classifier. Meanwhile, the captured anonymized image should be usable for not only one task but various tasks such as people flow analysis, character recognition, and object detection.

Computational imaging in relation to privacy. FlatCam [3], the coded aperture camera [22, 40], and single-pixel imaging [10] are based on compressed sensing (CS) theory [6]. CS-based imaging destroys spatial information in the sensor image (raw image) and visually eliminates privacy in the sensor image. However, because the original image, which includes personal data, can potentially be recovered from the sensor image by CS reconstruction methods, it is not classified as pre-capture. In [25], which is a modified version of FlatCam, facial information is eliminated in software by detecting the face through CS reconstruction. This approach is essentially classified as post-capture privacy.

To overcome the trade-off between privacy and utility, it is necessary to be able to set arbitrary anonymized targets and to be able to capture images without global degradation for use in any application. In contrast to prior studies, the proposed method satisfies all of these requirements.

3. Adaptive Single-Pixel Imaging for Privacy

Figure 1 shows an overview of the proposed pre-capture privacy-aware imaging method. The proposed method is based on a single-pixel imaging (SPI) framework and introduces a feedback mechanism, called an aperture pattern generator (APG), using a deep learning model. SPI gradually captures a single image by repeatedly acquiring incident light through the aperture pattern, where the entire reconstructed image gradually becomes clear. The APG generates the next aperture pattern to avoid sampling the anonymized target from the current unclear reconstructed image. This feedback mechanism make it possible to eliminate the anonymized target optically. Sec. 3.1 describes and formulates the principle of conventional SPI, and Sec. 3.2 describes how anonymization is achieved through the feedback mechanism by APG.

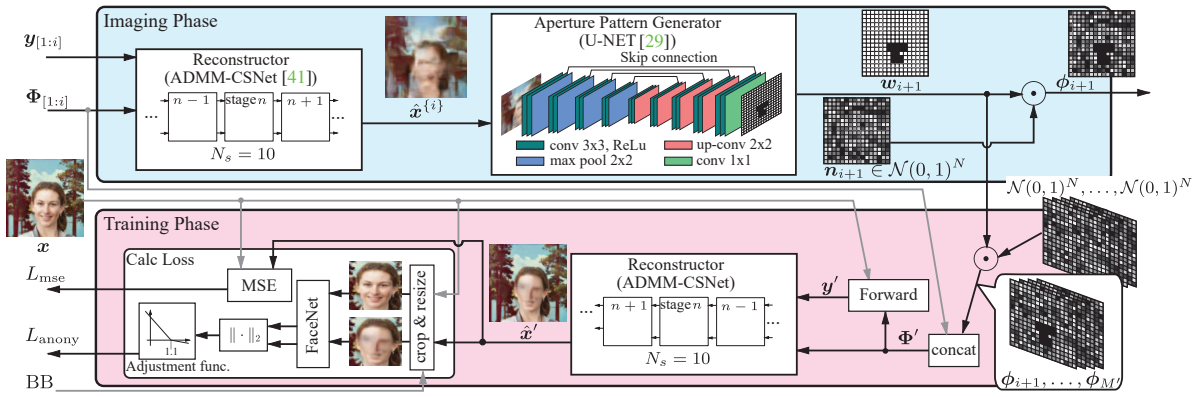

Figure 2. Proposed network architecture. The aperture pattern generator (APG) implicitly estimates the location of the anonymized target from $\hat{\boldsymbol{x}}^{\{i\}}$, which is reconstructed from the already acquired data, and outputs the next aperture pattern $\boldsymbol{\phi}_{i+1}$ to avoid sampling that location. The APG and reconstructor are jointly trained. The two reconstructors share the network weights.

3.1. Single-Pixel Imaging

We explain the mathematical principles, followed by the optical implementation. SPI is an imaging method based on CS [6]. While we use SPI, our focus is not on the low sampling rates typical of CS, but rather on achieving privacy with higher sampling rates to preserve image quality for downstream tasks. The target image $\boldsymbol{x} \in \mathbb{R}^N$ (an image with a total of N pixels) is not acquired directly but is reconstructed from $\boldsymbol{y}$ and $\boldsymbol{\Phi}$. First, $\boldsymbol{x}$ is optically modulated to a measurement $\boldsymbol{y} \in \mathbb{R}^M$ with fewer $M(< N)$ elements using a measurement matrix $\boldsymbol{\Phi} \in \mathbb{R}^{M \times N}$, and then $\boldsymbol{y}$ is acquired.

$$\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{x} \tag{1}$$

Then $\boldsymbol{x}$ is reconstructed from $\boldsymbol{y}$ and $\boldsymbol{\Phi}$.

$$\hat{\boldsymbol{x}} = \mathrm{Recon}(\boldsymbol{y}, \boldsymbol{\Phi}) \tag{2}$$

The reconstruction is typically solved by an iterative algorithm [36]. A deep unrolled network, which is an algorithm that combines the advantages of deep learning techniques and traditional iterative reconstruction algorithms, has also been developed [41, 42]. We choose an unrolled network, ADMM-CSNet [41], as the CS reconstructor because of its high computational speed and accuracy. In this paper, y_i denotes the i-th elements of $\boldsymbol{y}$, $\boldsymbol{y}_{[1,i]}$ denotes the sub-vector from the 1st to i-th elements of $\boldsymbol{y}$, $\boldsymbol{\phi}_i \in \mathbb{R}^N$ denotes the i-th row vector of $\boldsymbol{\Phi}$, and $\boldsymbol{\Phi}_{[1,i]}$ denotes sub-matrix from the 1st row to the i-th row of $\boldsymbol{\Phi}$. M/N is referred to as the sampling rate.

SPI involves a photodetector (PD) and digital micromirror device (DMD) as shown in Figure 1. The light ray from the target is modulated by the aperture pattern $\boldsymbol{\phi}_i$ displayed on the DMD, and the modulated light is then acquired in the PD ($y_i = \boldsymbol{\phi}_i \cdot \boldsymbol{x}$). The above process is repeated M times to obtain $\boldsymbol{y}$.

Additionally, SPI can also be reconstructed using $\boldsymbol{y}_{[1,i]}$ at the $i(< M)$-th acquisition. $\hat{\boldsymbol{x}}^{\{i\}}(= \mathrm{Recon}(\boldsymbol{y}_{[1,i]}, \boldsymbol{\Phi}_{[1,i]}))$ represents the reconstructed image at the i-th acquisition. When i is small, $\hat{\boldsymbol{x}}^{\{i\}}$ is inaccurate, and the accuracy is expected to increase as i increases.

3.1.1 Block-based CS

In the block-based CS [13], the target image is partitioned into small non-overlapping blocks which are acquired independently but reconstructed jointly. This can reduce the computational cost of reconstruction.

Suppose that we capture an $L \times L$ image ($N = L \times L$ pixels in total) by dividing it into $B \times B$-pixel blocks ($n = B \times B$ total pixels in a block). $N_b = N/n$ is the number of blocks. As N_b measurements are acquired every i, $\boldsymbol{y} \in \mathbb{R}^M$ is acquired for $M'(= M/N_b)$ iterations, and Φ is re-defined as an $M' \times N$ matrix. Additionally, $\boldsymbol{y}_i$ denotes the measurements of N_b blocks at the i-th acquisition and can be written as

$$\boldsymbol{y}_i \in \mathbb{R}^{N_b} = \begin{bmatrix} y_{i,1} \\ y_{i,2} \\ \vdots \\ y_{i,Nb} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\phi}_{i,1} \cdot \boldsymbol{x}_1 \\ \boldsymbol{\phi}_{i,2} \cdot \boldsymbol{x}_2 \\ \vdots \\ \boldsymbol{\phi}_{i,Nb} \cdot \boldsymbol{x}_{Nb} \end{bmatrix} = \mathrm{Forward}(\boldsymbol{\phi}_i, \boldsymbol{x}), \tag{3}$$

where $y_{i,j}$ denotes the measurement of the j-th block at i-th acquisition, $\boldsymbol{\phi}_{i,j} \in \mathbb{R}^n$ denotes the j-th block of $\boldsymbol{\phi}_i$, and $\boldsymbol{x}_j \in \mathbb{R}^n$ denotes the j-th block of $\boldsymbol{x}$. The measurements from the 1st to i-th acquisition are written as

$$\boldsymbol{y}_{[1:i]} \in \mathbb{R}^{iN_b} = [\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_i]^T \tag{4}$$

3.2. Adaptive Aperture Generation for Privacy

Introducing a feedback mechanism via an APG into SPI enables anonymization. In SPI, aperture patterns generated from random normal distributions are typically used, but in the proposed method, aperture patterns are derived through a feedback mechanism via the APG. As shown in Figure 2, the aperture pattern at $i+1$ ($\boldsymbol{\phi}_{i+1}$) is adaptively generated from the unclear provisional reconstructed image at the i-th acquisition ($\hat{\boldsymbol{x}}^{\{i\}}$) to avoid sampling the anonymized target. When i is sufficiently small, $\hat{\boldsymbol{x}}^{\{i\}}$ is expected to be an unclear image, and face silhouettes can be detected even through individuals cannot be identified. As a very simple example, it is possible to avoid acquiring detailed parts of

Algorithm 1 Training procedure (**lines 15–23 are skipped for imaging procedure**)

Require: $\boldsymbol{x}$, BB(Bounding box of the anonymized target)

1: $\boldsymbol{w} \leftarrow \mathbf{1}^N$
2: $\boldsymbol{\phi}_1 \leftarrow \mathcal{N}(0,1)^N$
3: $\boldsymbol{\Phi}_{[1,1]} \leftarrow [\boldsymbol{\phi}_1]^T$
4: **for** $i \leftarrow 1, \ldots, M'$ **do**
5: $\quad \boldsymbol{y}_i \leftarrow \text{Forward}(\boldsymbol{\phi}_i, \boldsymbol{x})^\dagger$
6: $\quad \boldsymbol{y}_{[1,i]} \leftarrow [\boldsymbol{y}_{[1,i-1]}, \boldsymbol{y}_i]$
7: $\quad$ **if** $i \in \{\lfloor K^n \rfloor | n \in \mathbb{N}\}$ **then**
8: $\quad\quad \hat{\boldsymbol{x}}^{\{i\}} \leftarrow \text{Recon}(\Theta_R, \boldsymbol{y}_{[1,i]}, \boldsymbol{\Phi}_{[1,i]})$
9: $\quad\quad \boldsymbol{w} \leftarrow \text{APG}(\Theta_G, \hat{\boldsymbol{x}}^{\{i\}})$
10: $\quad$ **end if**
11: $\quad \boldsymbol{w}_{i+1} \leftarrow \boldsymbol{w}$
12: $\quad \boldsymbol{n}_{i+1} \leftarrow \mathcal{N}(0,1)^N$
13: $\quad \boldsymbol{\phi}_{i+1} \leftarrow \boldsymbol{w}_{i+1} \odot \boldsymbol{n}_{i+1}$
14: $\quad \boldsymbol{\Phi}_{[1,i+1]} \leftarrow [\boldsymbol{\Phi}_{[1,i]}, [\boldsymbol{\phi}_{i+1}]^T]$
15: $\quad$ **if** $i \in \{\lfloor K^n \rfloor | n \in \mathbb{N}\}$ **then**
16: $\quad\quad \boldsymbol{\phi}_{i+2}, \ldots, \boldsymbol{\phi}_{M'} \leftarrow \boldsymbol{w}_{i+1} \odot (\mathcal{N}(0,1)^N, \ldots, \mathcal{N}(0,1)^N)$
17: $\quad\quad \boldsymbol{\Phi}' \leftarrow [\boldsymbol{\Phi}_{[1,i+1]}, [\boldsymbol{\phi}_{i+2}, \ldots, \boldsymbol{\phi}_{M'}]^T]$
18: $\quad\quad \boldsymbol{y}'_{i+1}, \ldots, \boldsymbol{y}'_{M'} \leftarrow \text{Forward}(\boldsymbol{\phi}_{i+1}, \boldsymbol{x}), \ldots, \text{Forward}(\boldsymbol{\phi}_{M'}, \boldsymbol{x})$
19: $\quad\quad \boldsymbol{y}' \leftarrow [\boldsymbol{y}_{[1,i]}, \boldsymbol{y}'_{i+1}, \ldots, \boldsymbol{y}'_{M'}]$
20: $\quad\quad \hat{\boldsymbol{x}}' \leftarrow \text{Recon}(\Theta_R, \boldsymbol{y}', \Phi')$
21: $\quad\quad$ Calculate $\mathcal{L}_G$ using $\boldsymbol{x}, \hat{\boldsymbol{x}}'$, BB, and update Θ_G
22: $\quad\quad$ Calculate $\mathcal{L}_R$ using $\boldsymbol{x}, \hat{\boldsymbol{x}}', \hat{\boldsymbol{x}}^{\{i\}}$, and update Θ_R
23: $\quad$ **end if**
24: **end for**
25: **return** $\hat{\boldsymbol{x}} \leftarrow \text{Recon}(\Theta_R, \boldsymbol{y}, \boldsymbol{\Phi})$

Θ_G and Θ_R are the network weights of the APG and reconstructor, respectively. †This operation is optical acquisition using the DMD and the PD in the real imaging process.

the face by setting $\boldsymbol{\phi}_{i+1}$ to zero for the location of each facial part after the i-th acquisition. Repeating the above process up to the M'-th acquisition should produce a reconstructed image $\hat{\boldsymbol{x}}$ in which only facial features are masked.

Alg. 1 shows the pseudo-code. The proposed system captures a single anonymized image by repeatedly performing the process of the optical acquisition using $\boldsymbol{\phi}_i$ (line 5) and the adaptive generation of $\boldsymbol{\phi}_{i+1}$ by the APG (lines 8–13). The APG generates $\boldsymbol{w}$, which is the sampling weight at each pixel. The APG consists of a U-NET deep learning model [29] (#steps=5, #channels=64) with outputs clipped within the range $[0,1]$ and takes $\hat{\boldsymbol{x}}^{\{i\}} \in \mathbb{R}^N$ as input and outputs $\boldsymbol{w} \in [0,1]^N$ as shown in line 9 of Alg. 1. The next aperture pattern $\boldsymbol{\phi}_{i+1}$ is defined as follows:

$$\boldsymbol{\phi}_{i+1} = \boldsymbol{w}_{i+1} \odot \boldsymbol{n}_{i+1}, \tag{5}$$

where $\boldsymbol{n}_{i+1}$ is a random normal distribution vector ($\mathcal{N}(0,1)^N$), and $\odot$ denotes an element-wise product. Because the compressed sensing theory states that a clear image can be obtained by using $\boldsymbol{\Phi}$ of random bases, $\boldsymbol{n}_{i+1}$ is used as the original basis and then is partly attenuated by $\boldsymbol{w}_{i+1}$ to suppress the acquisition of information at each pixel.

In addition, to accelerate the training and imaging process, the adaptive feedback (line 8–9 of Alg. 1) operates only at exponential intervals (line 7), and the previous $\boldsymbol{w}$ is reused (line 11). Because a random vector $\boldsymbol{n}_{i+1}$ is generated at each i (line 12), a different $\boldsymbol{\phi}_{i+1}$ is obtained. K can be changed in the training and imaging phase. A too large K causes loss of anonymity.

3.2.1 Loss Function

The APG is trained with the following loss function to output $\boldsymbol{w}$ such that only the anonymized target is not sampled.

$$\mathcal{L}_G = \alpha L_{\text{mse}} + (1-\alpha) L_{\text{anony}}, \tag{6}$$

where α is a balancing parameter. L_{mse} and L_{anony} are used to evaluate image quality and the degree of anonymity, respectively. The loss function is evaluated using the target image $\boldsymbol{x}$, which is known at the training phase, and a reconstructed image which depends on $\boldsymbol{w}_{i+1}$. Note that instead of $\hat{\boldsymbol{x}}^{\{i+1\}}$, which is the reconstructed image at the $(i+1)$-th acquisition, we use $\hat{\boldsymbol{x}}'$, which is the reconstructed image when $\boldsymbol{w}_{i+1}$ is reused until the M'-th acquisition. Since hundreds of acquisitions are required for a single image, the impact of an aperture pattern ($\boldsymbol{w}_{i+1}$) is small. To amplify the minute effects of a single $\boldsymbol{w}_{i+1}$ and facilitate learning, we use the reconstructed image assuming that $\boldsymbol{w}_{i+1}$ is reused until the end (M'), namely $\hat{\boldsymbol{x}}'$, as shown in lines 16–20 of Alg. 1.

L_{mse} and L_{anony} are calculated from $\boldsymbol{x}$ and $\hat{\boldsymbol{x}}'$ as shown in the 'Training Phase' in Figure 2. L_{mse} is the mean squared error between $\boldsymbol{x}$ and $\hat{\boldsymbol{x}}'$. L_{anony} must be small when anonymity is high. L_{anony} depends on the anonymized target (face and license plate), the details of which are defined as follows:

Face Anonymization. FaceNet [31] is used as a facial feature extractor. In FaceNet, the distance of feature vectors from two face images is less than 1.1 when the two faces are the same individual. Following this rule, L_{anony} is calculated as follows: first, $\boldsymbol{x}$,$\hat{\boldsymbol{x}}'$, and BB (bounding box of the face) are given. A face image pair is created by cropping $\boldsymbol{x}$ and $\hat{\boldsymbol{x}}'$ using BB, and the cropped image pair is resized to 160×160 to match the input of FaceNet. Then we calculate the distance of the feature vectors from the output of FaceNet (average if there is more than one face) and enter the distance value into an adjustment function. The adjustment function is a modified Leaky ReLU function, i.e.,

$$f(x) = \begin{cases} -(x-1.1) & \text{if } x < 1.1 \\ -0.01 \times (x-1.1) & otherwise \end{cases} \tag{7}$$

License Plate Anonymization. The basic procedure is the same as that for faces; please refer to the supplementary material for details.

	Face	License plate
Optimizer	Adam	
l_r	1.0×10^{-4} (halved at every 5 epochs)	
#epochs	40	
#mini_batches	4	
Data augmentation	Random crop, Random resize $[0.5 : 2]$	
	Random rotate $[-10° : 10°]$	
Training time	about three weeks	
α	0.999	0.1
Dataset	BSDS500 [2], DIV2K [1]	
	CelebA [21]	Cars [18]

Table 1. Training Parameters

3.2.2 Robustness to Reconstruction Attacks

We need to consider what kind of image attackers will obtain when all acquisition values, namely $\mathbf{\Phi}$ and $\boldsymbol{y}$, are leaked. The anonymity level during training is assessed using $\hat{\boldsymbol{x}}'$ reconstructed by our training's reconstructor. However, as attackers may use various reconstruction methods, anonymity should ideally be robust against any reconstruction method. To achieve this, the reconstructor is specifically trained for the $\mathbf{\Phi}$ property produced by the APG, which should enable it to surpass the performance of the attackers' reconstructors. For this purpose, the reconstructor is alternately trained with the APG using a specific equation, which is line 22 of Alg. 1.

$$\mathcal{L}_{\mathrm{R}} = \frac{1}{N}\|\boldsymbol{x} - \hat{\boldsymbol{x}}^{\{i\}}\|^2 + \frac{1}{N}\|\boldsymbol{x} - \hat{\boldsymbol{x}}'\|^2 \tag{8}$$

This perspective is also discussed in Sec 4.3.

4. Simulated Experiment

We verify the effectiveness of the proposed method through a simulation experiment in which $\boldsymbol{y}_i = \mathrm{Forward}(\boldsymbol{\phi}_i, \boldsymbol{x})$ is operated on a computer with the image in the dataset as $\boldsymbol{x}$. We assume two anonymized targets, a face (Sec. 4.1) and a license plate (LP) (Sec. 4.2). Although it is difficult to compare the proposed method to other pre-capture privacy-preserving methods, we conduct a quantitative comparison with the simplest method using defocus blurring with a lens. To simulate defocus blurring, a 31×31 Gaussian blur with $\sigma = 16$ is applied to the input image. σ is adjusted so that the anonymity is almost the same as that of the proposed method. We compare 'Original', 'Defocus', and 'Ours', which correspond to general cameras, cameras with the defocus lens attached, and the proposed method, respectively. The target is 256×256 RGB images ($N = 65536$), and the block size B is 32. Higher sampling rates generally improve image accuracy but make anonymity harder to maintain. To assess anonymity under challenging conditions, we use a high sampling rate ($M/N = 0.5$), resulting in $M' = 512$. We set $K = 4$, determining the 5 feedback iterations (N_f). The programs are written in Python (TensorFlow v2.9.1) and run in Ubuntu 20.04 with an Intel Xeon Platinum 8275CL (memory: 1152GB) and a NVIDIA Tesla A100 (40GB). Other training parameters are shown in Table 1. All images in the dataset are separated into training and testing images at a ratio of 9:1. The anonymized targets (face or LP) are detected from the training images by the pre-trained detector, and their bounding boxes (BB) are stored in advance. A total of 64K images are prepared for training.

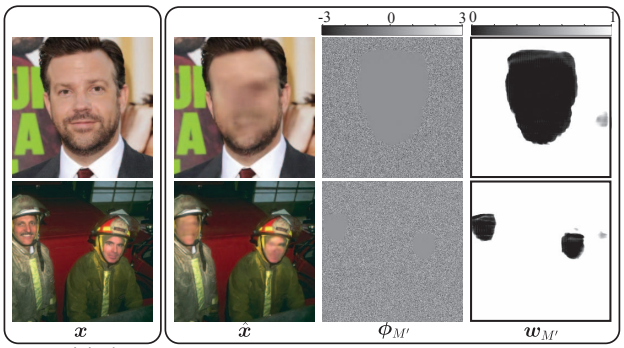

Figure 3. Images captured from simulations of face anonymization

4.1. Face Anonymization

4.1.1 Training

The pre-trained Retinaface[1] [33] is used for the face detector, the pre-trained FaceNet[2] [31] for the facial feature extractor, and ADMM-CSNet [41] for the CS reconstruction. The detector and feature extractor are used to prepare the training data and compute the loss function but are not used in the imaging phase. ADMM-CSNet is modified to be applicable to block-based CS and pre-trained using $\mathbf{\Phi}$ of a random normal distribution matrix with the sampling rate in the range of 0.0 to 0.5 in advance, and the pre-trained weights are used as initial values. The training images contain a roughly even mix of faces and non-faces. The ratio of the number of faces in each image is adjusted to 5:4:1 for zero, one, and two or more faces. Since the face images are unaligned, faces appear in various positions.

4.1.2 Results

Figure 3 shows the output images. As shown by $\hat{\boldsymbol{x}}$, the clothing, letters, and background are accurate, while detailed information on the face is concealed, making it difficult to identify the person. Additionally, it remains effective even when multiple faces are presented. $\boldsymbol{w}_{M'}, \boldsymbol{\phi}_{M'}$ indicates that the face area is set to zero values to avoid acquiring features. Figure 4 shows the progression of $\hat{\boldsymbol{x}}^{\{i\}}$ and $\boldsymbol{w}_{i+1}$ for better understanding of the role of the introduced APG. The APG can estimate the location of faces from $\hat{\boldsymbol{x}}^{\{i\}}$ and generate $\boldsymbol{w}_{i+1}$ to avoid sampling at the location. As for

[1] https://github.com/peteryuX/retinaface-tf2
[2] https://github.com/davidsandberg/facenet

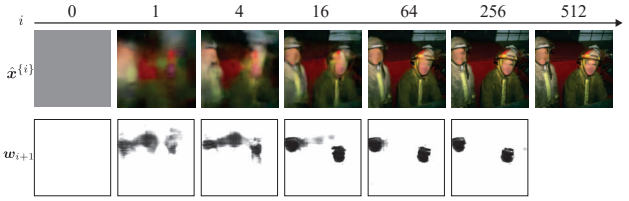

Figure 4. $\hat{\boldsymbol{x}}^{\{i\}}$ and $\boldsymbol{w}_{i+1}$ in progress. $\hat{\boldsymbol{x}}^{\{i\}}$ and $\boldsymbol{w}_{i+1}$ are calculated at '$i \in \{\lfloor K^n \rfloor | n \in \mathbb{N}\}$' (where $K = 4$). When $i = 0$, $\hat{\boldsymbol{x}}^{\{i\}}$ and $\boldsymbol{w}_{i+1}$ are the initial values (not calculated). When $i = 1$, $\hat{\boldsymbol{x}}^{\{i\}}$ is reconstructed, and then the APG generates $\boldsymbol{w}_{i+1}$ to avoid sampling the faces, although the region may be slightly inaccurate. The same applies hereinafter at $i = 4, 16, 64, 256$. $\boldsymbol{w}_{i+1}$ gradually becomes more accurate. Finally, when $i = 512$, $\hat{\boldsymbol{x}}(= \hat{\boldsymbol{x}}^{\{M'\}})$ is reconstructed and outputted as the captured image.

Method	Anonymity		Image quality	
	LFW(↓)	AgeDB-30(↓)	PSNR(↑)	PASCAL VOC2007(↑)
Original	0.999	0.987	-	0.6912
Defocus	0.659	0.569	21.06	0.2535
Ours	0.675	0.558	31.64	0.6078

Table 2. Results of anonymity and image quality in face anonymization. 'LFW' and 'AgeDB-30' indicate AUC value in 1:1 face verification test. 'PASCAL VOC2007' indicates mAP on object detection.

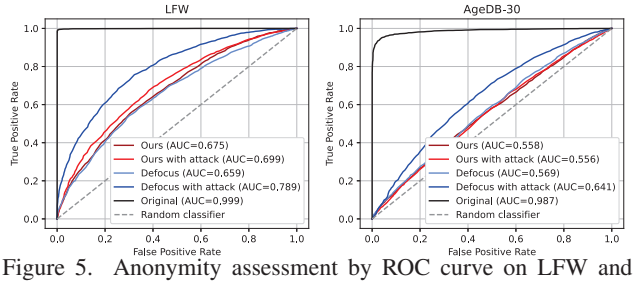

Figure 5. Anonymity assessment by ROC curve on LFW and AgeDB-30.

computational time, the generation time of $\boldsymbol{\Phi}$ per image is about 0.35 seconds.

The quantitative evaluation is conducted to assess anonymity and image quality, the result of which are shown in Table 2. For the anonymity assessment, we perform a face recognition test and evaluate the area under curve (AUC) of the receiver operating characteristic (ROC) curve. We test the LFW [15] and AgeDB-30 [24] dataset using the pre-trained ArcFace[3] model [8]. A set of 'Defocus' and 'Ours' images are obtained through Gaussian blur and simulation of our method, respectively, from the original images in the dataset. The images are resized once to 256×256, each operation is applied, and then they are resized to back to the original size. The column values of 'LFW' and 'AgeDB-30' indicate the AUC value. If the faces is recognized completely randomly, the AUC will be

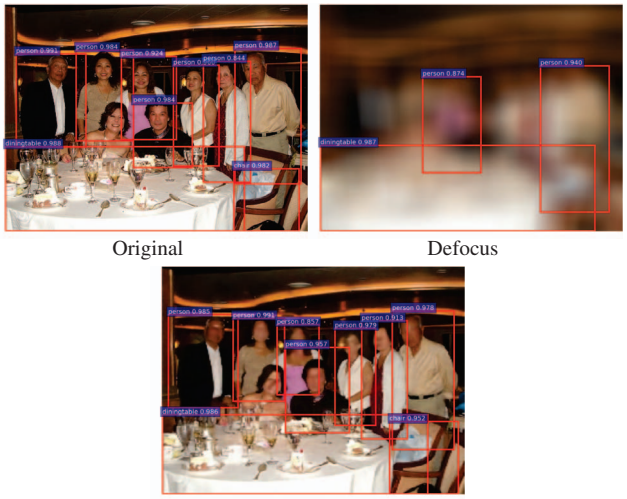

Figure 6. Results of object detection. In 'Ours', all objects are detected although the positions of bounding boxes are slightly different from that of 'Original'. 'Defocus' detects only a few objects.

0.5. Figure 5 shows the ROC curve. In AgeDB-30, 'Ours' and 'Defocus' are close to the random classifier. The results of the CFP-FP [32] and FGLFW [9] dataset can be found in our supplementary material. The results show that the proposed method and the defocus lens method achieve a high degree of anonymity.

The image quality is evaluated by using an image quality metric and score of object recognition. For the image quality metric, PSNR is calculated by masking the face area. Next, object recognition scores are evaluated to assess whether objects other than faces are accurately captured. The dataset PASCAL VOC 2007 [12] (20 object classes) and the detector model Faster-RCNN[4] [28] are used. The training and testing images of PASCAL VOC 2007 are converted to 'Original', 'Defocus', and 'Ours', respectively, in advance. This conversion procedure is the same as that of the face recognition test. The Faster-RCNN models is the trained on the training images. The PASCAL VOC2007 test is performed, and the mean average precision (mAP) is reported in the 'PASCAL VOC2007' column. The values show that the proposed method clearly captures objects other than faces. In contrast, 'Defocus' cannot be used for object recognition due to overall image degradation. Figure 6 shows an example of object detection.

Image restoration Attacks. We evaluate the anonymity in the case of an image restoration attack. Assuming that an attacker can access a set of original ($\boldsymbol{x}$) and reconstructed images ($\hat{\boldsymbol{x}}$), the attacker could train a network to recover the faces. For this purpose, we utilize Panini-Net [39], which is the most advanced GAN-based model for face image restoration and can handle various types of image degradations. The training images are converted by using "Ours" and "Defocus" respectively, and the model is trained

[3] https://github.com/peteryuX/arcface-tf2

[4] https://github.com/smallcorgi/Faster-RCNN_TF

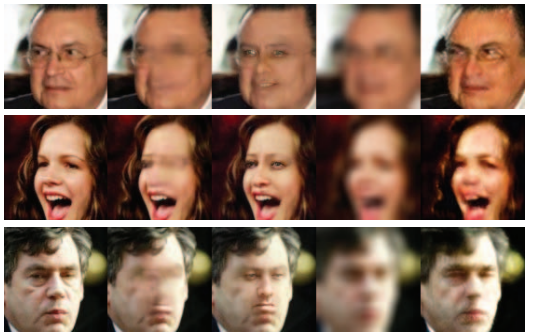

Figure 7. Image Restoration Attack

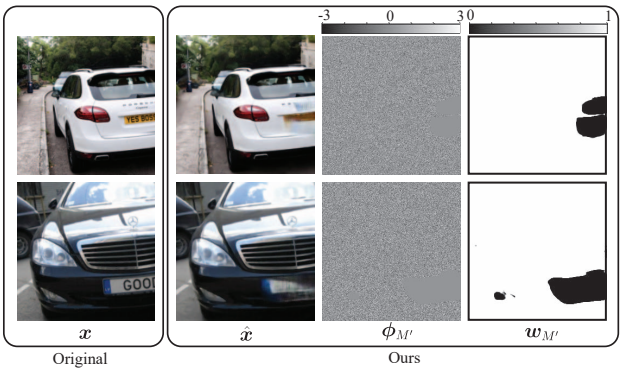

Figure 8. Images captured from simulations of LP anonymization

by each of the converted training images. Figure 7 shows examples of restored images. With 'Ours', Panini-Net frequently restores the face of a noticeably different person, whereas it is restored quite accurately with 'Defocus'. The quantitative results of the face recognition test under the image restoration attack are presented as 'Ours with attack' and 'Defocus with attack' in Figure 5. The image restoration attack is very effective against 'Defocus' but has little to no effect on 'Ours'. The results show that the proposed method is robust against image restoration attacks whereas the defocus lens is not.

In summary, the results of our quantitative evaluation demonstrate that only the proposed method achieves anonymity while providing clear imaging for other objects.

4.2. License Plate Anonymization

The second anonymized target is set to be vehicle license plates (LPs). Since the basic experimental procedure follows that of the face version (Sec. 4.1), we focus on the differences in this section. The training processes are largely the same; for details, see the supplementary material.

4.2.1 Results

Figure 8 shows the output images. As shown by $\hat{x}$, detailed information on the LPs is concealed. $\phi_{M'}$ and $w_{M'}$ indicate that the LP area is set to zero to avoid acquiring features.

Quantitative evaluation is conducted in term of anonymity and image quality. Table 3 shows the results of the quantitative evaluation. For the anonymity assessment, we follow the ALPR-Unconstrained[5] test condition [34], where an LP is considered correct if all characters are correctly recognized. ALPR-Unconstrained is used for LP detection and recognition. As shown in Table 3, no LP could be correctly identified in 'Ours' and 'Defocus'. However, the image quality of the proposed method is higher than that of 'Defocus' and is comparable to that of the original. As in the case of faces, only the proposed method achieves both anonymity and utility.

Method	Anonymity	Image quality	
	ALPR(↓)	PSNR(↑)	PASCAL VOC2007(↑)
Original	0.715	-	0.6912
Defocus	0.0	21.06	0.2535
Ours	0.0	31.81	0.6117

Table 3. Results of anonymity and image quality in license plate anonymization. 'ALPR' indicates scores of LP recognition test. 'PASCAL VOC2007' indicates mAP on object detection.

ADMM-CSNet (jointly trained) | ADMM-CSNet (not jointly trained) | ADMM(TV) | PnP(BM3D)

Figure 9. Comparison of reconstruction methods

4.3. Reconstruction Attacks

In this section, assuming all of the acquired data ($\mathbf{\Phi}$ and $\boldsymbol{y}$) has been leaked, we compare three reconstruction methods to verify that the details of the face cannot be recovered. In Sec. 4.1 and 4.2, we evaluated the degree of anonymity using the reconstructor (ADMM-CSNet) which is trained jointly with the aperture pattern generator. As described in Sec. 3.2.2, because the reconstructor is trained to recover data as accurately as possible, including the face, the evaluation should be reliable. However, an attacker who obtains $\mathbf{\Phi}$ and $\boldsymbol{y}$ may reconstruct the target image by any CS reconstruction method. In this section, we evaluate three different reconstruction methods: ADMM-CSNet which is not jointly trained, the alternating directions method of multipliers (ADMM) [5] with total variation (TV) [30] regularization, and plug-and-play (PnP) [38] with BM3D [7], as shown in Figure 9. ADMM-CSNet, which is jointly trained, recovers the face most accurately, indicating that that the anonymity evaluation in Sec 4.1 and 4.2 is reliable.

4.4. Ablation Study

Comparison to Simple Detection and Masking. As the APG is similar to a semantic segmentation network with only two instances (face and background), we compare the APG with a simple segmentation model to verify its

[5] https://github.com/sergiomsilva/alpr-unconstrained

Method	Anonymity
APG	0.675
Segmentation	0.832

Table 4. Anonymity by simple semantic segmentation compared to APG. 'Anonymity' indicates the AUC value of face recognition on the LFW dataset using the pre-trained Arcface model. The segmentation model compromises anonymity by failing to detect faces in low-quality reconstructed images when i is small.

K	Anonymity	Times[sec] (#feedbacks)
1.5	0.675	0.83 (16)
2	0.676	0.54 (9)
4	0.675	0.35 (5)
8	0.683	0.26 (3)
16	0.701	0.26 (3)

Table 5. Anonymity (AUC for face recognition on LFW using Arcface) and GPU computing times for CS reconstruction and APG, by varying K.

effectiveness. We evaluate the APG trained by Alg. 1 and the face segmentation model trained using CelebAMask-HQ [20] as shown Table 4. The network structure of both models is exactly the same. The segmentation model compromises anonymity due to its inability to detect faces in low-quality reconstructed images when i is small. Additional details are in the supplementary material.

Determination of K**.** If K is too large, the amount of feedback can be reduced but anonymity would be lost. Table 5 shows how changing K affects anonymity. When K exceeds 8, the anonymity begins to decline, and speed does not improve significantly. However, reducing K to less than 4 does not result in any anonymity improvement at all. As a result, we prioritized anonymity and set $K = 4$.

5. Prototype

As shown in Figure 10, we assembled a rough prototype of the proposed system based on the single-pixel imaging implementation in [11]. To simplify implementation, our prototype is degraded in two aspects compared to the simulated version: the aperture pattern is binary and the captured images are monochrome.

The subject is a paper printout of one of the images in CelebA, the sampling rate M/N is 0.75, and $K = 4$. Since only one PD is used, the PD sequentially acquires the light from each block by switching off the blocks other than the target block. Furthermore, due to inadequate control and synchronization, the system operates slowly. As a result, it takes about thirty seconds to capture one image. Note that the bottleneck is not the processing time of the introduced APG because the total GPU computing time is less than 0.5 seconds. Figure 10 also shows the captured images in non-adaptive conventional SPI and in our system, which

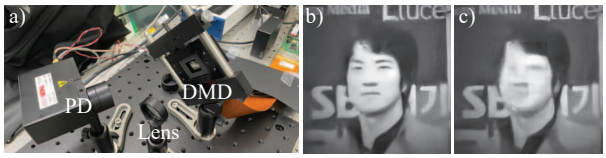

Figure 10. (a) Prototype of the proposed system and images captured by (b) non-adaptive conventional SPI and (c) our system. We use a Thorlabs PMM02 as the PD. The analog to digital converter is a National Instruments USB-6223. The DMD is a Vialux V7001-VIS for intensity modulation. The objective lens is a Thorlabs LB1901.

demonstrate that the proposed method was effective in the actual experiment.

6. Discussion and Conclusion

We have presented a pre-capture privacy-aware imaging method based on single-pixel imaging that adaptively generates aperture patterns using a deep learning model. The introduced aperture pattern generator outputs the next aperture pattern by exploiting the data already acquired so as to exclude features of the anonymized target. Through simulation experiments on face and license plate anonymization, we show that our proposed method can anonymize images while maintaining image quality.

However, the following should be considered with regards to the proposed method:

Real-time imaging. Real-time imaging is difficult because thousands of acquisition values must be sequentially performed for a single image. The fundamental bottleneck of SPI is the operating frequency of the DMD. Recent studies [16, 17] have achieved more than 100 fps by mechanically moving a DMD or modulating light with LEDs instead. By combining these implementations with the proposed method, real-time imaging should be feasible.

Anonymity for reconstruction using temporal adjacency. All experiments in this paper are evaluated assuming that a single image is recovered from a single $\boldsymbol{\Phi}$ and $\boldsymbol{y}$. However, when the proposed method is applied to video, a reconstruction attack may exploit even multiply pairs of $\boldsymbol{\Phi}$ and $\boldsymbol{y}$ derived from the previous and next frames. We have not evaluated the anonymity for such a situation.

Reconstruction-free inference. [19, 23] suggest that inference without reconstruction may outperform inference after reconstruction. To more comprehensively evaluate anonymity, it is necessary to assess the proposed method from this perspective.

For future work, we plan to improve the hardware implementation for real-time imaging. In addition, we plan to conduct further evaluations to expand the scope of anonymized targets and examine the case where multiple types of anonymized targets are specified simultaneously.

References

[1] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, July 2017. 5

[2] Pablo Arbelaez, Michael Maire, Charless Fowlkes, and Jitendra Malik. Contour detection and hierarchical image segmentation. *IEEE Transactions on pattern analysis and machine intelligence*, 33(5):898–916, May 2011. 5

[3] M Salman Asif, Ali Ayremlou, Aswin Sankaranarayanan, Ashok Veeraraghavan, and Richard G Baraniuk. Flatcam: Thin, lensless cameras using coded aperture and computation. *IEEE Transactions on Computational Imaging*, 3(3):384–397, 2016. 2

[4] Eric Bezzam, Martin Vetterli, and Matthieu Simeoni. Learning rich optical embeddings for privacy-preserving lensless image classification. *arXiv preprint arXiv:2206.01429*, 2022. 2

[5] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, Jonathan Eckstein, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine learning*, 3(1):1–122, 2011. 7

[6] Emmanuel J Candès and Michael B Wakin. An introduction to compressive sampling. *IEEE signal processing magazine*, 25(2):21–30, 2008. 2, 3

[7] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on image processing*, 16(8):2080–2095, 2007. 7

[8] 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 (CVPR)*, June 2019. 6

[9] Weihong Deng, Jiani Hu, Nanhai Zhang, Binghui Chen, and Jun Guo. Fine-grained face verification: Fglfw database, baselines, and human-dcmn partnership. *Pattern Recognition*, 66:63–73, 2017. 6

[10] Marco F. Duarte, Mark A. Davenport, Dharmpal Takhar, Jason N. Laska, Ting Sun, Kevin F. Kelly, and Richard G. Baraniuk. Single-pixel imaging via compressive sampling. *IEEE Signal Processing Magazine*, 25(2):83–91, 2008. 2

[11] Matthew P Edgar, Graham M Gibson, Richard W Bowman, Baoqing Sun, Neal Radwell, Kevin J Mitchell, Stephen S Welsh, and Miles J Padgett. Simultaneous real-time visible and infrared video with single-pixel detectors. *Scientific reports*, 5(1):10669, 2015. 8

[12] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The PASCAL Visual Object Classes Challenge 2007 (VOC2007) Results. http://www.pascal-network.org/challenges/VOC/voc2007/workshop/index.html. 6

[13] Lu Gan. Block compressed sensing of natural images. In *2007 15th International conference on digital signal processing*, pages 403–406. IEEE, 2007. 3

[14] Carlos Hinojosa, Juan Carlos Niebles, and Henry Arguello. Learning privacy-preserving optics for human pose estimation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 2573–2582, October 2021. 2

[15] Gary B. Huang, Manu Ramesh, Tamara Berg, and Erik Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst, October 2007. 6

[16] Hongxu Huang, Lijing Li, Yuxuan Ma, and Mingjie Sun. 25,000 fps computational ghost imaging with ultrafast structured illumination. *Electronic Materials*, 3(1):93–100, 2022. 8

[17] Patrick Kilcullen, Tsuneyuki Ozaki, and Jinyang Liang. Compressed ultrahigh-speed single-pixel imaging by swept aggregate patterns. *Nature Communications*, 13(1):7879, 2022. 8

[18] Jonathan Krause, Michael Stark, Jia Deng, and Li Fei-Fei. 3D object representations for fine-grained categorization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, pages 554–561, 2013. 5

[19] Kuldeep Kulkarni and Pavan Turaga. Reconstruction-free action inference from compressive imagers. *IEEE transactions on pattern analysis and machine intelligence*, 38(4):772–784, 2015. 8

[20] Cheng-Han Lee, Ziwei Liu, Lingyun Wu, and Ping Luo. Maskgan: Towards diverse and interactive facial image manipulation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 8

[21] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep learning face attributes in the wild. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, December 2015. 5

[22] Patrick Llull, Xuejun Liao, Xin Yuan, Jianbo Yang, David Kittle, Lawrence Carin, Guillermo Sapiro, and David J. Brady. Coded aperture compressive temporal imaging. *Opt. Express*, 21(9):10526–10545, May 2013. 2

[23] Suhas Lohit, Kuldeep Kulkarni, Pavan Turaga, Jian Wang, and Aswin C Sankaranarayanan. Reconstruction-free inference on compressive measurements. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 16–24, 2015. 8

[24] Stylianos Moschoglou, Athanasios Papaioannou, Christos Sagonas, Jiankang Deng, Irene Kotsia, and Stefanos Zafeiriou. AgeDB: The first manually collected, in-the-wild age database. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1997–2005, 2017. 6

[25] Thuong Nguyen Canh and Hajime Nagahara. Deep compressive sensing for visual privacy protection in flatcam imaging. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, pages 3978–3986, 2019. 2

[26] Francesco Pittaluga and Sanjeev J. Koppal. Privacy preserving optics for miniature vision sensors. *Proceedings of*

the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 07-12-June:314–324, 2015. 2

[27] Francesco Pittaluga, Aleksandar Zivkovic, and Sanjeev J Koppal. Sensor-level privacy for thermal cameras. In *2016 IEEE International Conference on Computational Photography (ICCP)*, pages 1–12. IEEE, 2016. 1, 2

[28] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28, 2015. 6

[29] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention - MICCAI*, volume 9351 of *Lecture Notes in Computer Science*, pages 234–241. Springer, 2015. 3, 4

[30] Leonid I Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena*, 60(1-4):259–268, 1992. 7

[31] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 815–823, 2015. 4, 5

[32] Soumyadip Sengupta, Jun-Cheng Chen, Carlos Castillo, Vishal M Patel, Rama Chellappa, and David W Jacobs. Frontal to profile face verification in the wild. In *IEEE Conference on Applications of Computer Vision*, February 2016. 6

[33] Sefik Ilkin Serengil and Alper Ozpinar. Hyperextended lightface: A facial attribute analysis framework. In *2021 International Conference on Engineering and Emerging Technologies (ICEET)*, pages 1–4. IEEE, 2021. 5

[34] Sergio Montazzolli Silva and Claudio Rosito Jung. License plate detection and recognition in unconstrained scenarios. In *Proceedings of the European conference on computer vision (ECCV)*, pages 580–596, 2018. 7

[35] Jasper Tan, Salman S Khan, Vivek Boominathan, Jeffrey Byrne, Richard Baraniuk, Kaushik Mitra, and Ashok Veeraraghavan. CANOPIC: Pre-digital privacy-enhancing encodings for computer vision. In *2020 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6. IEEE, 2020. 2

[36] Jie Tang, Brian E Nett, and Guang-Hong Chen. Performance comparison between total variation TV-based compressed sensing and statistical iterative reconstruction algorithms. *Physics in Medicine & Biology*, 54(19):5781, 2009. 3

[37] Zaid Tasneem, Giovanni Milione, Yi-Hsuan Tsai, Xiang Yu, Ashok Veeraraghavan, Manmohan Chandraker, and Francesco Pittaluga. Learning phase mask for privacy-preserving passive depth estimation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 504–521. Springer, 2022. 2

[38] Singanallur V Venkatakrishnan, Charles A Bouman, and Brendt Wohlberg. Plug-and-play priors for model based reconstruction. In *2013 IEEE Global Conference on Signal and Information Processing*, pages 945–948. IEEE, 2013. 7

[39] Yinhuai Wang, Yujie Hu, and Jian Zhang. Panini-net: Gan prior based degradation-aware feature interpolation for face restoration. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 2022. 6

[40] Zihao W Wang, Vibhav Vineet, Francesco Pittaluga, Sudipta N Sinha, Oliver Cossairt, and Sing Bing Kang. Privacy-preserving action recognition using coded aperture videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2019. 2

[41] Yan Yang, Jian Sun, Huibin Li, and Zongben Xu. Admm-csnet: A deep learning approach for image compressive sensing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(3):521–538, 2020. 3, 5

[42] Jian Zhang and Bernard Ghanem. ISTA-Net: Interpretable optimization-inspired deep network for image compressive sensing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1828–1837, 2018. 3

[43] Yupeng Zhang, Yuheng Lu, Hajime Nagahara, and Rin-Ichiro Taniguchi. Anonymous camera for privacy protection. *Proceedings of International Conference on Pattern Recognition (ICPR)*, pages 4170–4175, 2014. 1, 2