Recoverable Anonymization for Pose Estimation: A Privacy-Enhancing Approach

Wenjun Huang[1] Yang Ni[1] Arghavan Rezvani[1] SungHeon Jeong[1]
Hanning Chen[1] Yezi Liu[1] Fei Wen[2] Mohsen Imani[1]
[1] University of California, Irvine, USA [2] Texas A&M University, USA
{wenjunh3, m.imani}@uci.edu

Abstract

Human pose estimation (HPE) is crucial for various applications . However, deploying HPE algorithms in surveillance contexts raises significant privacy concerns due to the potential leakage of sensitive personal information (SPI) such as facial features, and ethnicity. Existing privacy-enhancing methods often compromise either privacy or performance, or they require costly additional modalities. We propose a novel privacy-enhancing system that generates privacy-enhanced portraits while maintaining high HPE performance. Our key innovations include the reversible recovery of SPI for authorized personnel and the preservation of contextual information. By jointly optimizing a privacy-enhancing module, a privacy recovery module, and a pose estimator, our system ensures robust privacy protection, efficient SPI recovery, and high-performance HPE. Experimental results demonstrate the system's robust performance in privacy enhancement, SPI recovery, and HPE.

1. Introduction

With the progression of computer vision, human pose estimation (HPE) has become a crucial and fundamental issue, attracting considerable scholarly attention. As a pivotal element of human-centric visual understanding, HPE establishes the groundwork for numerous advanced computer vision tasks, such as human action recognition [62], human parsing [53], motion prediction and retargeting [35, 40]. Consequently, it underpins a broad collection of applications, including human behavior analysis [58], violence detection [21], crowd riot scene identification [72], and autonomous driving [68].

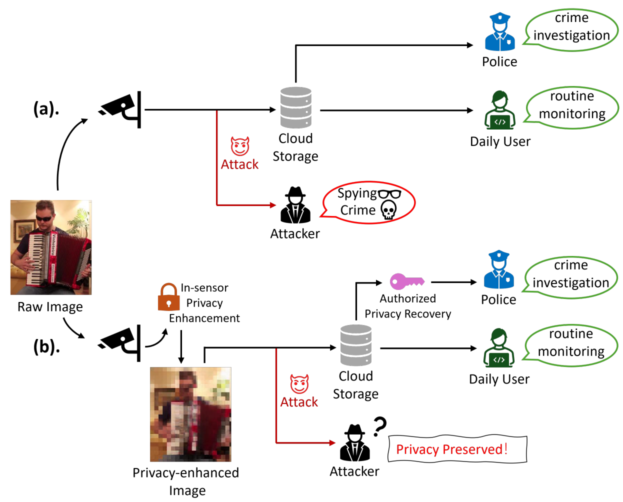

Figure 1. Motivation for our privacy-enhancing system. **(a).** Conventional surveillance systems are susceptible to leaks of SPI, which can be exploited for illicit surveillance and criminal activities. **(b).** Our system not only safeguards SPI against information misuse but also supports HPE. The privacy-enhanced images retain functionality for routine monitoring, while SPI remains recoverable by authorized personnel.

Due to the extensive computation involved in the applications above, users typically resort to cloud services for data processing and machine learning [25, 73, 74, 78]. However, when data is transmitted to cloud servers, sensitive personal information (SPI) such as facial features, gender, and ethnicity is inevitably shared. Privacy issues are particularly pronounced in surveillance contexts where HPE algorithms are widely deployed, as illustrated in Fig. 1(**a**). Ubiquitous surveillance systems collect and share vast amounts of data. While this data is valuable for legitimate users in various scenarios, such as routine monitoring, and crime investigations, it simultaneously raises significant privacy concerns for individuals and public safety. Without careful protection measures, SPI in raw data could be leaked

This work was supported in part by the DARPA Young Faculty Award, the National Science Foundation (NSF) under Grants #2127780, #2319198, #2321840, #2312517, and #2235472, the Semiconductor Research Corporation (SRC), the Office of Naval Research through the Young Investigator Program Award, and Grants #N00014-21-1-2225 and N00014-22-1-2067. Additionally, support was provided by the Air Force Office of Scientific Research under Award #FA9550-22-1-0253, along with generous gifts from Xilinx and Cisco.

and misused by malicious parties for harmful purposes. For instance, attackers might recognize individuals and surveil them for further criminal activities or even forge their identities [9]. Additionally, the leakage of SPI can introduce bias and compromise the fairness of analyses and judicial processes [17].

In response to data misuse, various legal regulations have been introduced [6, 15], and researchers are developing more advanced algorithms to consider personal privacy. For privacy enhancement in computer vision applications, a straightforward solution is to use very low-resolution data [39, 54]. Although these methods do not require specialized training to remove privacy features, they often fail to balance privacy enhancement and model performance effectively. Some approaches [3,8,57] employ additional modalities to enhance privacy. However, the need to install sensors for these extra modalities increases the cost of surveillance systems, impeding their widespread deployment. Another set of methods involves modifying images with handcrafted features such as blurring, adding noise, and pixelation [1, 10, 47]. Unfortunately, these techniques demand extensive domain knowledge, which may not be practical in real-world applications.

Recent privacy-enhancing systems adopt data-driven approaches that conceal SPI from various perspectives. For instance, Hukkelås *et al*. [30] propose a framework using a generative adversarial network (GAN) for full-body synthesis. Their approach generates new representations of individuals that effectively obscure SPI while preserving essential pose information. In another approach, Hinojosa *et al*. [24] introduces a hardware/software co-design framework. This framework optimizes both the point spread function of the camera lens and the neural network architecture, enabling the development of domain-specific computational cameras tailored for privacy-enhancing purposes. Furthermore, Dave *et al*. [16] present a training framework that autonomously removes SPI in a self-supervised manner, alleviating the need for extensive manual labeling efforts. Kansal *et al*. [34] propose a novel dual-stage framework that suppresses SPI from the discriminative features, and introduces a controllable privacy mechanism through differential privacy.

However, most of the previous work does not target HPE. Besides, all the aforementioned methods exhibit shortcomings in one or more of the following aspects:

(1). Recovery of Removed SPI: Privacy-enhanced images should allow authorized users to recover SPI when necessary. While SPI may not be essential for scientific research or routine monitoring, it remains critical for specific applications. To ensure data utility for various users, authorized personnel such as law enforcement officials should be able to recover original raw images from privacy-enhanced versions, particularly for investigative purposes.

(2). Preservation of Context: Effective privacy-enhancing systems should modify only the region of interest (e.g., humans) while preserving the background unchanged. Contextual information is crucial as the interpretation of actions can vary significantly depending on the surroundings [11,19,23,76]. For instance, distinguishing between someone jogging in a park and someone fleeing a store after theft requires intact contextual clues. Therefore, the context information should be preserved after privacy enhancement, to aid correct interpretation.

(3). Lightweight Deployment: Privacy-enhancing systems need to be lightweight for deployment near cameras. Transmitting data to cloud servers poses security risks such as interception and tampering during transmission [18, 55]. Deploying privacy-enhancing systems near cameras reduces these vulnerabilities by processing raw images locally before transmission [26,27,75]. Therefore, such systems must operate efficiently in real-time, considering limited computational resources and power constraints.

By addressing the limitations observed in previous work, we propose a privacy-enhancing system capable of generating privacy-enhanced portraits of individuals in images with minimal impact on HPE, as depicted in Fig. 1**(b)**. The privacy-enhancing module operates near the camera, processing raw images before transmission. This approach ensures that SPI in the privacy-enhanced images remains concealed from potential attackers, yet remains usable for HPE tasks and recoverable by authorized users through a privacy recovery module. Our approach begins by desensitizing raw images using conventional methods such as blurring, pixelation, or noise addition. These desensitized images serve as initial supervised inputs for the privacy-enhancing module, which then modifies original images to create privacy-enhanced versions in a trainable manner. To ensure the preservation of essential features for recovery and HPE, we optimize the privacy-enhancing process in conjunction with a privacy recovery model and a pose estimator. Through supervised and joint learning, our system achieves effective privacy protection, robust recovery capabilities, and maintains high performance in HPE tasks. The key contributions of this work are outlined as follows:

- To the best of our knowledge, we are the first to discuss reversibility, privacy recovery, and context preservation in privacy enhancement for HPE. We introduce a novel privacy-enhancing system designed to generate privacy-enhanced portraits of individuals in images, specifically adapted for downstream machine learning tasks such as HPE.
- We proposed an end-to-end joint learning policy for obfuscation, recovery, and pose estimation modules, with the ultimate aim of maintaining pose information and HPE performance after obfuscation and recovery.

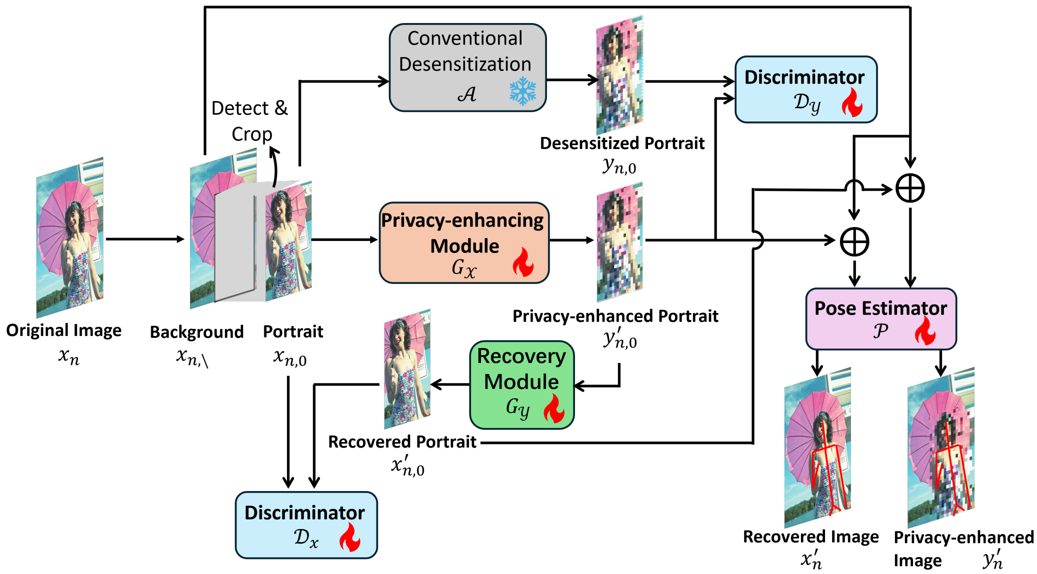

Figure 2. The complete pipeline of our proposed system. It contains a privacy-enhancing module $G_{\mathcal{X}}$ erasing private information, a module $G_{\mathcal{Y}}$ recovering the removed private information, two discriminators $D_{\mathcal{X}}, D_{\mathcal{Y}}$ for distinguishing the generated portraits, and a pose estimator $\mathcal{P}$ implementing pose estimation. 🔥 denotes the trainable modules, and ❄ denotes the frozen modules.

- We experimentally show that our system achieves robust performance in privacy protection, recovery of original images, and accurate human pose estimation. With joint training, on privacy-enhanced images, our model achieves around 10% higher average precision than the one that only finetunes the HPE model, while also equipped with strong obfuscation capability. On recovered images, our model further enhances the quality by around 3%, thanks to the accurate recovery and adaptive injection of HPE-related information.

2. Method

In this section, we elaborate on each component of our proposed system. As illustrated in Fig. 2, our system is composed of three modules: **(1). A privacy-enhancing module** (Sec. 2.1). We leverage an image-to-image style translation model using conditional generative adversarial networks (cGANs) [45] to generate privacy-enhanced portraits. The privacy-enhanced module is able to anonymize SPI in the images while preserving the features for the downstream tasks. The style translation is learned with the guide of a pose estimator such that necessary features are injected for downstream tasks. **(2). A privacy recovery module** (Sec. 2.2). In order to facilitate the reversibility given authorization, we use another pair of cGANs and jointly optimize them with the privacy-enhancing module to recover the SPI. **(3). A pose estimator** for human detection and pose estimation on both privacy-enhanced and recovered images (Sec. 2.3). All modules are tuned end-to-end to maintain pose estimation quality, where the pose estimator provides feedback for the first two modules.

2.1. Privacy Enhancing Module

Consider a set of images in the original domain $\{X_0, X_1, \cdots, X_n\} \in \mathcal{X}$. Each image X_n contains one or multiple people of portraits $\{x_{n,0}, x_{n,1}, \cdots, x_{n,i}\} \in \mathcal{X}$ with articulated pose annotations $\{p^x_{n,0}, p^x_{n,1}, \cdots, p^x_{n,i}\}$, where i denotes the portrait index in X_n. We leverage a pretrained lightweight object detector to detect all people and crops the regions to construct a data pool of $\{x_{0,0}, x_{0,1}, \cdots, x_{n,i}\} \in \mathcal{X}$ with poses $\{p^x_{0,0}, p^x_{0,1}, \cdots p^x_{n,i}\}$.

The goal of the privacy-enhancing module can be defined as follows: Given the pool of training articulated portraits $\{x_{0,0}, \cdots, x_{n,i}\}$ with poses $\{p^x_{0,0}, \cdots p^x_{n,i}\}$, we want to generate the paired privacy-enhanced portraits $\{y'_{0,0}, \cdots, y'_{n,i}\} \in \mathcal{Y}$ in the desensitized domain with poses $\{p^{y'}_{0,0}, \cdots p^{y'}_{n,i}\}$. $y'_{n,i}$ should maintain a high pose feature similarity with the paired portrait $x_{n,i}$ (i.e., $p^x_{n,i} \approx p^{y'}_{n,i}$) while removing the SPI in it. To achieve this in a learnable manner, we introduce a generator $G_{\mathcal{X}}$ and discriminator $D_{\mathcal{Y}}$.

The generator $G_{\mathcal{X}}$ generates the privacy-enhanced portrait $y'_{n,i} = G_{\mathcal{X}}(x_{n,i})$. To facilitate the generation, a discriminator $D_{\mathcal{Y}}$ is adopted to learn to distinguish the generated portraits $y'_{n,i}$ and the desensitization style guidance portraits $y_{n,i} = \mathcal{A}(x_{n,i})$, where $y_{n,i}$ is the privacy-enhanced portrait generated from a conventional desensitization method $\mathcal{A}$. Mathematically, $D_{\mathcal{Y}}$ distinguishes the portrait pair $(y_{n,i}, y'_{n,i})$ via a discriminator loss:

$$\begin{aligned}\mathcal{L}_{D_\mathcal{Y}} = &- \mathbb{E}_{(x,y)\sim p_{\text{data}}(x,y)}[\log D_\mathcal{Y}(y|x)] \\ &- \mathbb{E}_{x\sim p_{\text{data}}(x)}[\log(1 - D_\mathcal{Y}(G_\mathcal{X}(x)|x))]\end{aligned} \quad (1)$$

, where $D_\mathcal{Y}(a|b)$ is the discriminator's output probability that the a is real given the condition b.

On the other hand, $G_\mathcal{X}$ tries to trick $D_\mathcal{Y}$. Therefore, it is optimized via the following loss:

$$\mathcal{L}_{G_\mathcal{X}} = -\mathbb{E}_{x\sim p_{\text{data}}(x)}[\log D_\mathcal{Y}(G_\mathcal{X}(x)|x)] \quad (2)$$

. By constructing the adversarial relation, $G_\mathcal{X}$ and $D_\mathcal{Y}$ are trained jointly and boost the other's performance gradually.

However, the desired style that $G_\mathcal{X}$ should learn is not specified in the aforementioned adversarial training, impacts the training stability and potentially results in model collapse. Therefore, we introduce an extra loss term $\mathcal{L}_1$ that explicitly indicates the optimization direction:

$$\mathcal{L}_1 = \mathbb{E}_{(x,y)\sim p_{\text{data}}(x,y)}[\|y - G_\mathcal{X}(x)\|_1] \quad (3)$$

. While $\mathcal{L}_1$ guides the learning of the style, on the other hand, a too-low value hinders the injection of the necessary information for HPE. Therefore, inspired by Huber loss [28], we adopt a modified loss $\mathcal{L}_{\mathcal{X}\mathcal{Y}}$ to balance the style guidance and information injection:

$$\mathcal{L}_{\mathcal{X}\mathcal{Y}} = \begin{cases}\mathcal{L}_1 & \text{if } \mathcal{L}_1 \geq T, \\ 0 & \text{otherwise}\end{cases} \quad (4)$$

, where T is a predefined threshold. The total loss of the privacy-enhancing module is

$$\mathcal{L}_{enhance} = \mathcal{L}_{D_\mathcal{Y}} + \mathcal{L}_{G_\mathcal{X}} + \lambda_1 \mathcal{L}_{\mathcal{X}\mathcal{Y}} \quad (5)$$

, where λ_1 is a hyperparameter.

The remaining background denoted $X_{n,\setminus} = X_n \setminus \{x_{n,0}, \cdots, x_{n,i}\}$ is combined with the privacy-enhanced portraits $\{y'_{n,0}, \cdots, y'_{n,i}\}$ to form the privacy-enhanced image $Y'_n = X_{n,\setminus} \bigcup \{y'_{n,0}, \cdots, y'_{n,i}\}$.

2.2. Privacy Recovery Module

The privacy recovery module aims to recover the SPI hidden in the privacy-enhanced portraits $y' \in \mathcal{Y}$. The recovery problem can be defined as follows: Given the privacy-enhanced portraits $\{y'_{0,0}, \cdots, y'_{n,i}\} \in \mathcal{Y}$, the module recovers the SPI and generates the privacy-recovered portraits $\{x'_{0,0}, \cdots, x'_{n,i}\} \in \mathcal{X}$ as similar to the original portraits as possible.

The recovery module adopts a similar architecture as the privacy-enhancing module, consisting of a generator $G_\mathcal{Y}$ and a discriminator $D_\mathcal{X}$. However, one difference between the two modules is that the privacy recovery module takes the learnable generations $\{y'_{n,i}, \cdots, y'_{n,i}\}$ as input, but not the fixed inputs, such as $x_{n,i}$, and $y_{n,i}$. This is because the goal of the recovery module is specific to recover the SPI in the privacy-enhanced portraits, therefore, there is no use in force it learns the mapping from the traditional desensitized images $y_{n,i}$ to $x_{n,i}$. The generator $G_\mathcal{Y}$ generates the privacy-recovered portrait $x'_{n,i} = G_\mathcal{Y}(y'_{n,i})$, and the discriminator $D_\mathcal{X}$ distinguishes the portrait pair $(x_{n,i}, x'_{n,i})$. The $G_\mathcal{Y}$ is optimized via the loss

$$\mathcal{L}_{G_\mathcal{Y}} = -\mathbb{E}_{y'\sim p(y')}[\log D_\mathcal{Y}(G_\mathcal{Y}(y')|y')] \quad (6)$$

, and the $D_\mathcal{X}$ facilitates its performance by the loss

$$\begin{aligned}\mathcal{L}_{D_\mathcal{X}} = &- \mathbb{E}_{x\sim p_{\text{data}}(x), y'\sim p(y')}[\log D_\mathcal{X}(x|y')] \\ &- \mathbb{E}_{y'\sim p(y')}[\log(1 - D_\mathcal{X}(G_\mathcal{Y}(y')|y'))]\end{aligned} \quad (7)$$

. A consistency loss Eq. (8) is introduced in the recovery module to guide the whole privacy-enhancing and recovery process explicitly. It forces the recovered portraits to have a similar style to the original portraits.

$$\mathcal{L}_{\text{consistency}} = \mathbb{E}_{x\sim p_{\text{data}}(x)}[\|G_\mathcal{X}(G_\mathcal{Y}(x)) - x\|_1]) \quad (8)$$

. The total objective function of the privacy recovery module is

$$\mathcal{L}_{recovery} = \mathcal{L}_{G_\mathcal{Y}} + \mathcal{L}_{D_\mathcal{X}} + \lambda_2 \mathcal{L}_{\text{consistency}} \quad (9)$$

, where λ_2 is a hyperparameter that controls the style explicit guidance.

2.3. Pose Estimator

The pose estimator model $\mathcal{P}$ conducts pose estimation on Y'_n without seeing any SPI. Given a set of images $\{Y'_0, \cdots, Y'_n\}$, the model is optimized via multiple loss terms: a bounding box loss $\mathcal{L}_{bbox}$ that measures the overlap between the predicted bounding box $[y'_{n,i}]$ and the ground truth bounding box, a pose loss $\mathcal{L}_{pose}$ that measures the difference between the predicted keypoints and ground truth articulation keypoints, an object loss $\mathcal{L}_{obj}$ that classifies whether a keypoint is visible, and a classification loss $\mathcal{L}_{cls}$ that classifies the detected objects into predefined category (i.e., "human"). The loss function of a pose estimator is

$$\mathcal{L}_{PE_\mathcal{Y}} = \mathcal{L}_{bbox_\mathcal{Y}} + \mathcal{L}_{pose_\mathcal{Y}} + \mathcal{L}_{obj_\mathcal{Y}} + \mathcal{L}_{cls_\mathcal{Y}} \quad (10)$$

. Since the purpose of our system is to estimate human pose in both the privacy-enhanced images and the privacy-recovered images, $\mathcal{P}$ should be capable of implementing pose estimation on the images from both domains ($\mathcal{X}$ and $\mathcal{Y}$). Therefore, the pose estimator is trained on the pairs (y', x'). The total loss for the pose estimator is denoted as:

$$\mathcal{L}_{PE} = \mathcal{L}_{PE_\mathcal{X}} + \mathcal{L}_{PE_\mathcal{Y}} \quad (11)$$

. $\mathcal{L}_{PE_\mathcal{X}}$ is defined on recovered images and $\mathcal{L}_{PE_\mathcal{Y}}$ is for privacy-enhanced images.

Finally, we jointly optimize the privacy-enhancing, privacy-recovery, and pose estimation modules end-to-end with the following overall loss function:

$$\mathcal{L} = \mathcal{L}_{enhance} + \mathcal{L}_{recovery} + \lambda_3 \mathcal{L}_{PE} \quad (12)$$

, where λ_3 is a hyperparameter.

3. Experiments

3.1. Setup

Our system is developed using PyTorch [49] and is trained on an NVIDIA RTX A6000 GPU. The architecture employs a U-Net [52] model as the backbone for the generators and PatchGAN [32] for the discriminators. For pose estimation, we integrate YOLOv8 [33], although the model can be interchangeably replaced with alternative pose estimation algorithms to suit specific needs. Training of these modules employs distinct optimization strategies: the generators and discriminators utilize the Adam optimizer, whereas YOLOv8 employs the AdamW optimizer to potentially enhance training stability and performance. The initial learning rate is set at 0.000035, which undergoes exponential decay to facilitate convergence. Data augmentation techniques include random horizontal flipping and adjustments to hue, saturation, and brightness of the input images. We train our models with a batch size of 16.

The experiments are conducted on the widely used datasets: MPII Human Pose (MPII) [4], and Microsoft Common Objects in Context (COCO) [38]. The MPII dataset comprises approximately 25,000 images featuring over 40,000 individuals. Each pose within this dataset is manually annotated with up to 16 body joints. The COCO dataset encompasses over 200,000 labeled individuals, each annotated with 17 body joints, primarily focusing on people depicted at medium and large scales.

We assess our system utilizing established metrics for image quality and pose estimation. For the evaluation of privacy enhancement and recovery, we employ two commonly accepted metrics: the Peak Signal-to-Noise Ratio (PSNR) and the Structural Similarity Index Measure (SSIM) [66]. PSNR values range from 0 to ∞, with ∞ indicating perfect similarity, implying no discernible difference between the compared images. The SSIM varies from 0 to 1, where a value of 0 indicates no structural similarity between the images. Typically, image pairs are deemed to exhibit high similarity when the PSNR$\geq$ 30 and the SSIM $\geq$ 0.9 [31, 66]. For pose estimation, we utilize the Object Keypoint Similarity (OKS), analogous to the Intersection over Union (IoU) used in object detection. OKS is calculated based on the scale of the subject and the Euclidean distances between predicted keypoints and their corresponding ground truth points. To quantify the performance of our pose estimation, we employ the mean Average Precision (mAP) and mean Average Recall (mAR) at an OKS threshold of 0.5, denoted as mAP@0.5 and mAR@0.5.

The conventional desensitization methods used in our system comprise Gaussian blurring, where the kernel radius r is set to 8, and pixelation, with each pixel block having a side length of $r = 12$.

3.2. Results of Privacy Enhancement

Figure 3 presents a qualitative comparison of our privacy-enhancing module. In contrast to the original portraits, our privacy-enhanced portraits demonstrate superior visual privacy protection. The contours of the body, as well as the details of the face and clothing, are obscured, thereby preventing SPI through visual inspection. Compared to conventional desensitized portraits, our privacy-enhanced portraits achieve a competitive level of visual obfuscation while employing a distinct learnable approach.

Table 1 illustrates the quantitative comparison between privacy-enhanced portraits and raw images in terms of PSNR and SSIM. We also show the zero-shot HPE performance of a pretrained pose estimator pre-trained on both types of privacy-enhanced images. Compared to conventional desensitized portraits, our privacy-enhanced portraits attain similar PSNR and SSIM values when compared to the original portraits. This indicates that our method achieves comparable levels of privacy protection to the baseline. Effective privacy enhancement necessitates that the pose estimator, pretrained on original images, should fail to make accurate zero-shot inferences on the privacy-enhanced images. Our method significantly reduces the HPE performance, indicating that the pretrained pose estimator struggles to perform HPE accurately on the modified images. This substantial degradation in performance demonstrates the robust privacy enhancement capabilities of our approach. The privacy-enhancing module guided by pixelation demonstrates a lower image similarity to the original images and more significantly impacts the HPE performance of the pretrained pose estimator, compared to the module guided by blurring.

In addition to the visual obfuscation capability, we also expect the system to restore high efficacy in HPE by finetuning the pose estimator and adapting toward the privacy-enhanced images. However, with the conventional method, the carefully finetuned pose estimator model still observes a significant drop in performance. As shown in Tab. 2, the metric mAP@$0.5_{\{\text{joint, p}\}}$ was cut by around 15%, mainly due to the irrecoverable loss of information with the obfuscation. In contrast, when enabling the joint optimization of the three components within our system, there is a significant improvement in HPE performance, as evidenced by the data in the first two columns of Tab. 2. Both the mAP@0.5 and the mAR@0.5 of our method exceed those achieved

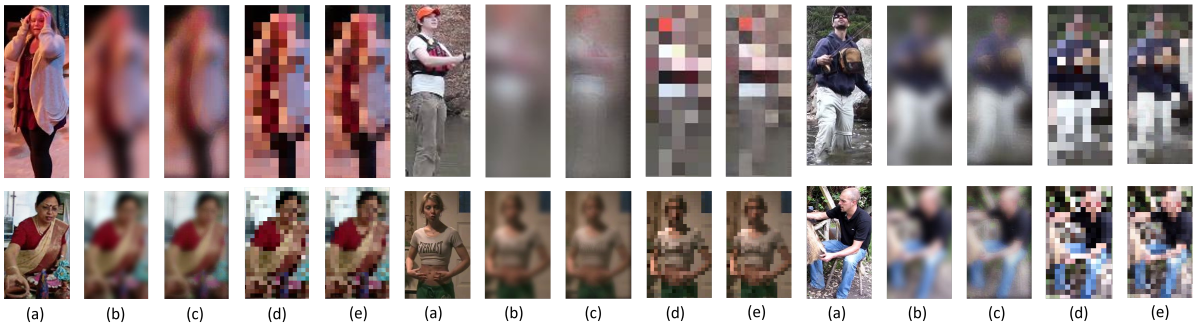

Figure 3. Qualitative comparison on privacy-enhanced portraits. (a) original portraits; (b)/(d) conventional desensitized portraits via blurring/pixelation; (c)/(e) privacy-enhanced portraits guided by blurring/pixelation. Enlarge for details.

Dataset	**MPII** (mAP@0.5$_{\{pre, o\}}$ = 83.9, mAR@0.5$_{\{pre, o\}}$ = 89.4)[a]				**COCO** (mAP@0.5$_{\{pre, o\}}$ = 86.2, mAR@0.5$_{\{pre, o\}}$ = 90.8)[a]			
Metrics	PSNR(o,p)↓[b]	SSIM(o,p)↓[b]	mAP@0.5$_{\{pre, p\}}$ ↓[c]	mAR@0.5$_{\{pre, p\}}$ ↓[c]	PSNR(o,p)↓[b]	SSIM(o,p)↓[b]	mAP@0.5$_{\{pre, p\}}$ ↓[c]	mAR@0.5$_{\{pre, p\}}$ ↓[c]
(1). Blurring								
Conventional	23.71	0.65	0.3	1.0	23.01	0.60	27.2	31.4
Ours	23.36	0.68	11.9	18.7	22.81	0.66	35.3	40.5
(2). Pixelation								
Conventional	19.97	0.53	0.1	0.6	19.34	0.49	0.1	0.3
Ours	20.89	0.56	0.2	0.5	20.15	0.54	0.2	0.3

[a] The subscript {pre, o} indicates a pose estimator pretrained on original images (pre), and tested on original images (o).
[b] A lower value indicates a lower similarity between the original image (o) and the privacy-enhanced image (p), showing a better privacy enhancement.
[c] A lower value indicates a better privacy enhancement. The subscript {pre, p} represents a pose estimator pretrained on original images (pre), and tested on privacy-enhanced images (p).

Table 1. Image Quality and Pose Estimation Performance of Privacy-enhanced Portraits.

with conventional desensitized portraits, with about 10% improvement in mAP0.5. Although these values are still marginally lower than those obtained by applying a pose estimator trained on original images to original images, this underscores that our system effectively incorporates valuable information into the privacy-enhanced portraits, thereby enhancing HPE performance.

3.3. Results of Privacy Recovery

A key strength of our system is that the anonymization process is reversible and we learn a uniform pose estimator for images before and after recovery. Figure 4 provides a qualitative comparison between the original portraits and the privacy-recovered portraits. The privacy-recovered portraits display visual quality that is on par with the original portraits. Distinguishing between the original and the privacy-recovered portraits through human visual inspection proves to be challenging, indicating effective restoration of SPI in the privacy-recovered images.

Table 2 presents the image quality metrics for the recovered images. The PSNR and SSIM values of the privacy-recovered portraits relative to the original portraits (i.e., PSNR(o, r) and SSIM(o,r) in Tab. 2) exceed 30 and 0.9, respectively, demonstrating that the privacy recovery module effectively restores the SPI. Surprisingly, the pose estimator, optimized jointly with other system components, outperforms a pose estimator trained solely on original images when applied to those images; it shows an average 3% improvement in mAP. This improvement is likely due to the privacy recovery module's dual function of not only restoring SPI from the portraits but also enhancing the HPE-related features during the recovery process, as guided by $\mathcal{L}_{PE_{\mathcal{X}}}$. Consequently, the privacy-recovered portraits retain the SPI while accentuating HPE-related features, thereby facilitating more accurate pose estimation. Additionally, the experimental results show that the system guided by blurring outperforms the other one (i.e., guided by pixelation) in terms of pose estimation on obfuscated and recovered images. Conversely, the system guided by pixelation more effectively restores the SPI from the privacy-enhanced images, achieving higher image quality (i.e., PSNR and SSIM metrics).

4. Discussion

4.1. Impact of Desensitization Guidance

Conventional desensitization guidance dictates the level of privacy enhancement in our module, influencing the style of the generated privacy-enhanced portraits. Severe desensitization, while increasing privacy, complicates the integration of HPE-related features, thereby hindering the joint training of the pose estimator and adversely affecting HPE

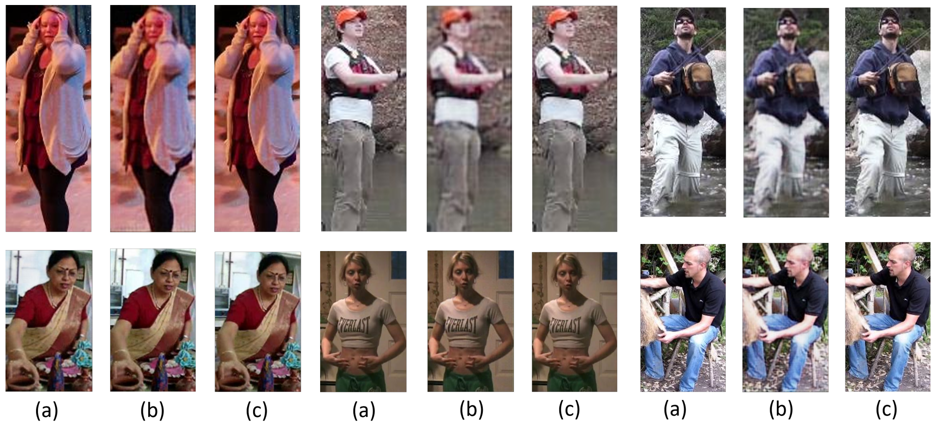

Figure 4. Qualitative results of the privacy-recovered portraits. (a) original portraits; (b)/(c) the portraits recovered from the privacy-enhanced portraits guided by blurring/pixelation. Enlarge for details.

Dataset	MPII (mAP@0.5$_{\{\text{pre, o}\}}$ = 83.9, mAR@0.5$_{\{\text{pre, o}\}}$ = 89.4)[a]					
Metrics	mAP@0.5$_{\{\text{joint, p}\}}$ ↑[b]	mAR@0.5$_{\{\text{joint, p}\}}$ ↑[b]	mAP@0.5$_{\{\text{joint, r}\}}$ ↑[b]	mAR@0.5$_{\{\text{joint, r}\}}$ ↑[b]	PSNR(o,r)↑[c]	SSIM(o,r)↑[c]
(1). Blurring						
Conventional	70.5	81.3	-	-	-	-
Ours	**81.5**	**88.8**	**87.4**	**92.4**	32.58	0.94
(2). Pixelation						
Conventional	65.2	77.9	-	-	-	-
Ours	74.9	84.9	87.1	91.3	**38.54**	**0.98**
Dataset	COCO (mAP@0.5$_{\{\text{pre, o}\}}$ = 86.2, mAR@0.5$_{\{\text{pre, o}\}}$ = 90.8)[a]					
Metrics	mAP@0.5$_{\{\text{joint, p}\}}$ ↑[b]	mAR@0.5$_{\{\text{joint, p}\}}$ ↑[b]	mAP@0.5$_{\{\text{joint, r}\}}$ ↑[b]	mAR@0.5$_{\{\text{joint, r}\}}$ ↑[b]	PSNR(o,r)↑[c]	SSIM(o,r)↑[c]
(1). Blurring						
Conventional	62.1	74.7	-	-	-	-
Ours	**75.3**	**84.9**	**89.0**	**92.5**	34.92	0.95
(2). Pixelation						
Conventional	59.4	65.6	-	-	-	-
Ours	70.3	81.1	88.6	92.0	**37.63**	**0.98**

[a] The subscript {pre, o} indicates a pose estimator pretrained on original images (pre), and tested on original images (o).
[b] A higher value indicates a better performance. The subscript {joint, o}, {joint, p}, {joint, r} represents a pose estimator joint trained with the privacy-enhancing and recovery modules, and tested on the original images (o), privacy-enhanced images (p), and privacy recovery images (r), respectively.
[c] A higher value indicates a higher similarity between the original image (o) and the privacy-recovered image (r), showing a better recovery.

Table 2. Image Quality and Pose Estimation Performance of Privacy-recovered Portraits.

Metrics	PSNR(o, p) ↓	SSIM(o, p) ↓	mAP@0.5$_{\{\text{joint, p}\}}$ ↑	mAR@0.5$_{\{\text{joint, p}\}}$ ↑
(1). Blurring[a]				
$r = 2$	29.22	0.84	82.5	89.3
$r = 4$	26.45	0.73	82.1	89.0
$r = 8$	23.36	0.68	81.5	88.8
$r = 12$	22.49	0.63	77.8	86.3
(2). Pixelation[b]				
$r = 4$	24.40	0.68	80.8	88.2
$r = 8$	21.36	0.55	76.2	85.3
$r = 12$	20.89	0.56	74.9	84.9
$r = 16$	20.09	0.47	60.7	75.2

[a] In blurring, r represents the radius of the blur kernel.
[b] In pixelation, r denotes the side length of each pixel block.

Table 3. Impact of Conventional Desensitization Guidance.

performance. Conversely, mild desensitization facilitates feature integration but may compromise privacy enhancement. Thus, the strategic selection of desensitization levels is crucial, as it significantly impacts overall system performance. Table 3 presents the performance of various conventional desensitization guidance methods, evaluating both portrait quality and HPE accuracy. As r increases, the capability for privacy enhancement improves, whereas the HPE performance deteriorates. Specifically, when r increases from 12 to 16 in pixelation, the similarity between the privacy-enhanced portraits and the original portraits remains relatively unchanged, yet there is a substantial decline in HPE performance. A similar pattern is observed in blurring when r changes from 8 to 12.

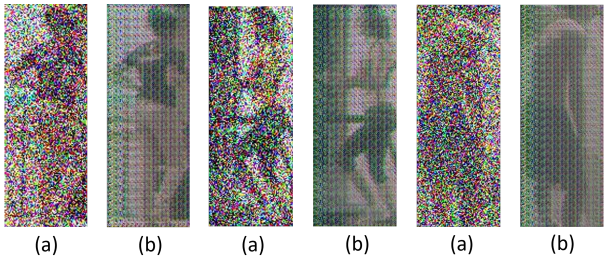

Figure 5. Qualitative results of the privacy-enhanced portraits guided by Gaussian noise. (a) conventional desensitized portraits; (b) privacy-enhanced portraits guided by Gaussian noise addition.

4.2. Impact of Adopting Noise Addition as Privacy Enhancement Guidance

Gaussian noise addition is another widely recognized conventional desensitization method. We explore its impact when utilized as guidance within our system. Figure 5 presents a qualitative comparison between conventional desensitized portraits and their corresponding privacy-enhanced counterparts. The privacy-enhanced portraits generated in the system exhibit numerous artifacts, diverging from the guaidance and compromising privacy preservation. We hypothesize that this deviation arises because Gaussian noise addition introduces a random pattern, which is challenging to learn through Eq. (3).

4.3. Backbone & Model Lightweightness

To facilitate deployment in surveillance environments, our privacy-enhancing module must be sufficiently lightweight to operate on edge devices without sacrificing its privacy-enhancement capabilities. We evaluate the impact of different backbones of the privacy-enhancing module on overall performance. Table 4 displays the results in terms of privacy-enhancement, HPE performance, and inference speed. Although both U-Net and ResNet backbones effectively capture the patterns of conventional desensitization, the privacy-enhanced portraits generated with a ResNet backbone exhibit poorer HPE performance. This suggests a failure in integrating HPE-related features effectively, potentially due to the absence of skip connections that are present in U-Net for transferring low-level information across the network. Further evaluations conducted on the NVIDIA Jetson AGX Orin [46] reveal that U-Net configurations 7 and 8 achieve desirable inference speeds, maintaining real-time processing capabilities (i.e., 30 FPS), which surpass those of the ResNet backbones. Given these findings, U-Net emerges as the more suitable backbone for our privacy-enhancing module, considering both performance metrics and latency requirements.

Metrics	PSNR(o, p)↓	SSIM(o, p)↓	PSNR(o, r)↑	SSIM(o, r)↑	mAP@0.5$_{\{joint, p\}}$ ↑	mAR@0.5$_{\{joint, p\}}$ ↑	FPS↑[a]
(1). Blurring							
U-Net 7	23.36	0.68	32.58	0.94	81.5	88.8	63.71
U-Net 8	23.34	0.68	32.55	0.94	81.4	88.8	59.84
ResNet 6	23.86	0.69	32.46	0.93	68.7	79.1	39.95
ResNet 9	23.91	0.69	32.41	0.93	67.5	78.8	34.82
(2). Pixelation							
U-Net 7	20.89	0.56	38.54	0.98	74.9	84.9	61.22
U-Net 8	20.81	0.55	38.63	0.98	75.2	85.0	57.19
ResNet 6	21.15	0.57	38.51	0.98	60.5	73.9	37.46
ResNet 9	21.18	0.57	38.45	0.97	60.9	74.1	33.96

[a] Frame per second (FPS) is measured at inference speed on NVIDIA Jetson AGX Orin.

Table 4. Impact of Backbone Architecture and Inference Speed on Edge Device.

5. Related Work

5.1. Pose Estimation

Multiple approaches exist for addressing HPE, with recent advancements in deep learning demonstrating superior performance compared to earlier methods [50, 61, 70, 71]. Notable recent deep-learning-based algorithms include [5, 22, 36, 37, 43]. These methods are typically discussed separately concerning single-person and multi-person scenarios. In single-person pose estimation, the objective is to localize joint positions in images containing only one person [12, 43, 59, 60]. In contrast, multi-person pose estimation methods can be categorized into top-down and bottom-up approaches. Top-down methods [7,13,37,56,64,69] first employ person detectors to identify individual persons in an image, then apply single-person pose estimation to each detected person. In contrast, bottom-up methods [42, 63, 65] first detect all body keypoints in an image and subsequently group them into distinct person instances.

5.2. Privacy Enhancing Methods

Naive image privacy-enhancing techniques such as masking, blurring, or pixelation are commonly employed in practice [1, 10]. However, these methods tend to remove semantic information and significantly degrade the quality of privacy-enhanced images, rendering the data unusable for many applications. Some efforts have explored addressing the issue through additional modalities [2, 48], but these approaches are often impractical and lack scalability. [14, 77] involve encrypting feature vectors of visual data to ensure privacy. However, encrypting large volumes of visual data is complex and resource-intensive. Recent studies have leveraged deep generative models to anonymize data while preserving its utility for downstream applications. They either inpaint missing regions [29, 44] or transform original regions [20,41,51,67]. However, much of prior work has focused primarily on face anonymization, leaving other identifiers such as clothing and body type untouched, which can compromise privacy. While some efforts have targeted full-body anonymization [30, 44], these approaches often lack recoverability, limiting their applicability.

6. Conclusion

We propose a privacy-enhancing system for HPE that addresses the critical need for protecting SPI while maintaining the performance of HPE tasks. The privacy-enhancing module, privacy recovery module, and pose estimator work in unison to anonymize SPI, allow for its recovery by authorized personnel, and ensure the preservation of contextual information essential for accurate behavior interpretation. Our experimental results demonstrate that the system achieves robust performance in privacy protection, accurate recovery of original images, and high-precision HPE.

References

[1] Prachi Agrawal et al. Person de-identification in videos. *IEEE Transactions on Circuits and Systems for Video Technology*, 21(3):299–310, 2011. 2, 8

[2] Shafiq Ahmad et al. Person re-identification without identification via event anonymization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 11132–11141, 2023. 8

[3] Shafiq Ahmad et al. Event anonymization: Privacy-preserving person re-identification and pose estimation in event-based vision. *IEEE Access*, 2024. 2

[4] Mykhaylo Andriluka et al. 2d human pose estimation: New benchmark and state of the art analysis. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2014. 5

[5] Bruno Artacho et al. Omnipose: A multi-scale framework for multi-person pose estimation. *arXiv preprint arXiv:2103.10180*, 2021. 8

[6] Eduard Barnoviciu et al. Gdpr compliance in video surveillance and video processing application. In *2019 International Conference on Speech Technology and Human-Computer Dialogue (SpeD)*, pages 1–6. IEEE, 2019. 2

[7] Yuanhao Cai et al. Learning delicate local representations for multi-person pose estimation. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16*, pages 455–472. Springer, 2020. 8

[8] Ting Cao et al. In-bed human pose estimation from unseen and privacy-preserving image domains. In *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)*, pages 1–5. IEEE, 2022. 2

[9] Anupama Chadha et al. Deepfake: an overview. In *Proceedings of second international conference on computing, communications, and cyber-security: IC4S 2020*, pages 557–566. Springer, 2021. 2

[10] Datong Chen et al. Tools for protecting the privacy of specific individuals in video. *EURASIP Journal on Advances in Signal Processing*, 2007:1–9, 2007. 2, 8

[11] Hanning Chen et al. Taskclip: Extend large vision-language model for task oriented object detection. *arXiv preprint arXiv:2403.08108*, 2024. 2

[12] Xianjie Chen et al. Articulated pose estimation by a graphical model with image dependent pairwise relations. *Advances in neural information processing systems*, 27, 2014. 8

[13] Yilun Chen et al. Cascaded pyramid network for multi-person pose estimation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7103–7112, 2018. 8

[14] Hang Cheng et al. Person re-identification over encrypted outsourced surveillance videos. *IEEE Transactions on Dependable and Secure Computing*, 18(3):1456–1473, 2019. 8

[15] Warren B Chik. The singapore personal data protection act and an assessment of future trends in data privacy reform. *Computer Law & Security Review*, 29(5):554–575, 2013. 2

[16] Ishan Rajendrakumar Dave et al. Spact: Self-supervised privacy preservation for action recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20164–20173, 2022. 2

[17] Julia Dressel et al. The dangers of risk prediction in the criminal justice system. 2021. 2

[18] Sahar Ebadinezhad. A systematic literature review on information security leakage: Evaluating security threat. In *Proceedings of Third International Conference on Sustainable Expert Systems: ICSES 2022*, pages 993–1007. Springer, 2023. 2

[19] Susan T Tufts Fiske. Social cognition: From brains to culture. 2020. 2

[20] Oran Gafni et al. Live face de-identification in video. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9378–9387, 2019. 8

[21] Guillermo Garcia-Cobo et al. Human skeletons and change detection for efficient violence detection in surveillance videos. *Computer Vision and Image Understanding*, 233:103739, 2023. 1

[22] Zigang Geng et al. Human pose as compositional tokens. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 660–671, 2023. 8

[23] Robert Gifford. Environmental psychology: Principles and practice. 2007. 2

[24] Carlos Hinojosa et al. Learning privacy-preserving optics for human pose estimation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 2573–2582, 2021. 2

[25] Wenjun Huang et al. Exploration of using a pressure sensitive mat for respiration rate and heart rate estimation. In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 298–301. IEEE, 2021. 1

[26] Wenjun Huang et al. Ecosense: Energy-efficient intelligent sensing for in-shore ship detection through edge-cloud collaboration. *arXiv preprint arXiv:2403.14027*, 2024. 2

[27] Wenjun Huang et al. Intelligent sensing framework: Near-sensor machine learning for efficient data transmission. *IEEE Sensors Journal*, 2024. 2

[28] Peter J Huber. Robust estimation of a location parameter. In *Breakthroughs in statistics: Methodology and distribution*, pages 492–518. Springer, 1992. 4

[29] Håkon Hukkelås et al. Deepprivacy: A generative adversarial network for face anonymization. In *International symposium on visual computing*, pages 565–578. Springer, 2019. 8

[30] Håkon Hukkelås et al. Deepprivacy2: Towards realistic full-body anonymization. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 1329–1338, 2023. 2, 8

[31] Quan Huynh-Thu et al. Scope of validity of psnr in image/video quality assessment. *Electronics letters*, 44(13):800–801, 2008. 5

[32] Phillip Isola et al. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017. 5

[33] Glenn Jocher et al. Ultralytics YOLO, Jan. 2023. 5

[34] Kajal Kansal et al. Privacy-enhancing person re-identification framework-a dual-stage approach. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 8543–8552, 2024. 2

[35] Moritz Kappel et al. High-fidelity neural human motion transfer from monocular video. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1541–1550, 2021. 1

[36] Ke Li et al. Pose recognition with cascade transformers. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1944–1953, 2021. 8

[37] Yanjie Li et al. Tokenpose: Learning keypoint tokens for human pose estimation. In *Proceedings of the IEEE/CVF International conference on computer vision*, pages 11313–11322, 2021. 8

[38] Tsung-Yi Lin et al. Microsoft coco: Common objects in context. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*, pages 740–755. Springer, 2014. 5

[39] Jixin Liu et al. Indoor privacy-preserving action recognition via partially coupled convolutional neural network. In *2020 International Conference on Artificial Intelligence and Computer Engineering (ICAICE)*, pages 292–295. IEEE, 2020. 2

[40] Zhenguang Liu et al. Motion prediction using trajectory cues. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 13299–13308, 2021. 1

[41] Jhon Lopez et al. Privacy-preserving optics for enhancing protection in face de-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12120–12129, 2024. 8

[42] Zhengxiong Luo et al. Rethinking the heatmap regression for bottom-up human pose estimation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 13264–13273, 2021. 8

[43] Weian Mao et al. Poseur: Direct human pose regression with transformers. In *European conference on computer vision*, pages 72–88. Springer, 2022. 8

[44] Maxim Maximov et al. Ciagan: Conditional identity anonymization generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5447–5456, 2020. 8

[45] Mehdi Mirza et al. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014. 3

[46] NVIDIA Corporation. Nvidia jetson agx orin developer kit user guide. Available from NVIDIA, 2023. 8

[47] José Ramón Padilla-López et al. Visual privacy protection methods: A survey. *Expert Systems with Applications*, 42(9):4177–4195, 2015. 2

[48] Marina Paolanti et al. Person re-identification with rgb-d camera in top-view configuration through multiple nearest neighbor classifiers and neighborhood component features selection. *Sensors*, 18(10):3471, 2018. 8

[49] Adam Paszke et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019. 5

[50] Leonid Pishchulin et al. Poselet conditioned pictorial structures. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 588–595, 2013. 8

[51] Zhongzheng Ren et al. Learning to anonymize faces for privacy preserving action detection. In *Proceedings of the european conference on computer vision (ECCV)*, pages 620–636, 2018. 8

[52] Olaf Ronneberger et al. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pages 234–241. Springer, 2015. 5

[53] Tao Ruan et al. Devil in the details: Towards accurate single and multiple human parsing. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 4814–4821, 2019. 1

[54] Michael Ryoo et al. Privacy-preserving human activity recognition from extreme low resolution. In *Proceedings of the AAAI conference on artificial intelligence*, volume 31, 2017. 2

[55] Nurul I Sarkar et al. A secure long-range transceiver for monitoring and storing iot data in the cloud: design and performance study. *Sensors*, 22(21):8380, 2022. 2

[56] Dahu Shi et al. End-to-end multi-person pose estimation with transformers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11069–11078, 2022. 8

[57] Vinkle Srivastav et al. Human pose estimation on privacy-preserving low-resolution depth images. In *International conference on medical image computing and computer-assisted intervention*, pages 583–591. Springer, 2019. 2

[58] Torben Teepe et al. Towards a deeper understanding of skeleton-based gait recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1569–1577, 2022. 1

[59] Jonathan Tompson et al. Efficient object localization using convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 648–656, 2015. 8

[60] Alexander Toshev et al. Deeppose: Human pose estimation via deep neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1653–1660, 2014. 8

[61] Fang Wang et al. Beyond physical connections: Tree models in human pose estimation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 596–603, 2013. 8

[62] Heng Wang et al. Action recognition with improved trajectories. In *Proceedings of the IEEE international conference on computer vision*, pages 3551–3558, 2013. 1

[63] Haixin Wang et al. Regularizing vector embedding in bottom-up human pose estimation. In *European Conference on Computer Vision*, pages 107–122. Springer, 2022. 8

[64] Jian Wang et al. Graph-pcnn: Two stage human pose estimation with graph pose refinement. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XI 16*, pages 492–508. Springer, 2020. 8

[65] Yihan Wang et al. Lite pose: Efficient architecture design for 2d human pose estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13126–13136, 2022. 8

[66] Zhou Wang et al. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 5

[67] Taihong Xiao et al. Adversarial learning of privacy-preserving and task-oriented representations. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 12434–12441, 2020. 8

[68] Feiyi Xu et al. Action recognition framework in traffic scene for autonomous driving system. *IEEE Transactions on Intelligent Transportation Systems*, 23(11):22301–22311, 2021. 1

[69] Sen Yang et al. Transpose: Keypoint localization via transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 11802–11812, 2021. 8

[70] Yi Yang et al. Articulated pose estimation with flexible mixtures-of-parts. In *CVPR 2011*, pages 1385–1392. IEEE, 2011. 8

[71] Bangpeng Yao et al. Modeling mutual context of object and human pose in human-object interaction activities. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 17–24. IEEE, 2010. 8

[72] Balasubramanian Yogameena et al. Computer vision based crowd disaster avoidance system: A survey. *International journal of disaster risk reduction*, 22:95–129, 2017. 1

[73] Chang Yu et al. Advanced user credit risk prediction model using lightgbm, xgboost and tabnet with smoteenn. *arXiv preprint arXiv:2408.03497*, 2024. 1

[74] Chang Yu et al. Credit card fraud detection using advanced transformer model. *arXiv preprint arXiv:2406.03733*, 2024. 1

[75] Sanggeon Yun et al. Hypersense: Hyperdimensional intelligent sensing for energy-efficient sparse data processing. *Advanced Intelligent Systems*, page 2400228, 2024. 2

[76] Sanggeon Yun et al. Missiongnn: Hierarchical multimodal gnn-based weakly supervised video anomaly recognition with mission-specific knowledge graph generation. *arXiv preprint arXiv:2406.18815*, 2024. 2

[77] Bowen Zhao et al. Freed: An efficient privacy-preserving solution for person re-identification. In *2022 IEEE Conference on Dependable and Secure Computing (DSC)*, pages 1–8. IEEE, 2022. 8

[78] Qi Zheng et al. Advanced payment security system:xgboost, lightgbm and smote integrated. *arXiv preprint arXiv:2406.04658*, 2024. 1