Differentially Private 2D Human Pose Estimation

Kaushik Bhargav Sivangi Paul Henderson Fani Deligianni
School of Computing Science, University of Glasgow
{kaushik.sivangi, paul.henderson, fani.deligianni}@glasgow.ac.uk

Abstract

Human pose estimation (HPE) underpins critical applications in healthcare, activity recognition, and human-computer interaction. However, the privacy implications of processing sensitive visual data present significant deployment barriers in critical domains. Differential Privacy (DP) provides formal guarantees but often results in steep performance costs. We introduce the first unified framework for differentially private 2D Human Pose Estimation (2D-HPE) that achieves strong privacy-utility trade-offs for structured visual prediction through complementary noise mitigation mechanisms. Our Feature-Projective DP integrates: (1) subspace projection that reduces noise variance by a factor k/p by restricting gradient updates to a k-principal subspace within the full p-dimensional parameter space, and (2) feature-level privacy, which selectively privatizes sensitive features while retaining public visual cues. Together these mechanisms yield a multiplicative utility gain under formal privacy constraints. Extensive experiments on MPII and HumanART datasets across privacy budgets ($\varepsilon \in \{0.2, 0.4, 0.6, 0.8\}$), clipping thresholds ($C \in \{0.01, 0.1, 1.0\}$) and training strategies demonstrate consistent improvements over vanilla DP-SGD. At $\varepsilon = 0.8$, our method achieves 82.61% PCKh@0.5, recovering 73% of the privacy induced performance gap. Cross-dataset evaluation on the HumanART confirms generalization (51.6 AP). Our study provides the first rigorous benchmark and a practical blueprint for privacy-preserving pose estimation in sensitive, real-world applications. Project page: https://bhairava2898.github.io/DP2DHPE/

1. Introduction

Human pose estimation (HPE) transforms raw visual data into structured keypoint representations of human posture and movements. This fundamental computer vision task enables numerous high impact applications in healthcare, activity recognition, human-computer interaction, sports analysis, and video games [5, 12, 39, 42, 45, 57]. However, as these systems are increasingly deployed into sensitive environments such as hospitals, homes, and workplaces, they introduce severe privacy risks at multiple levels [46].

At the data level, raw images contain identifiable biometric information exposed during collection and processing [67]. At the model level, trained networks can inadvertently memorize training data, enabling adversaries to extract sensitive information through model inversion, membership inference, and reconstruction attacks [20, 31, 65]. For instance, an adversary can exploit a model's weights [22, 53] or gradients [24] to reconstruct distinctive physical characteristics of patients or sensitive contextual information, such as the patient's home environment from the private training dataset [31]. This reconstruction could potentially identify individuals with specific medical conditions and reveal that they received treatment at a particular facility during the model's training period, thereby compromising both medical confidentiality and location privacy.

Previous privacy-preservation approaches in HPE rely primarily on data anonymization techniques, such as blurring, pixelation, and template-based shape modeling [3, 25, 49]. While these methods provide some level of privacy protection, they are often task-specific and can severely compromise the utility of the data for broader analysis. For instance, anonymization that removes facial features might preserve basic joint position information but destroy crucial clinical indicators needed for stress level assessment or abnormal motion pattern detection [6]. Furthermore, these methods do not offer formal privacy guarantees and remain vulnerable to more sophisticated attacks, limiting their applicability in highly sensitive contexts [11, 65]. Moreover, these approaches do not address the inherent vulnerability of neural networks to memorization attacks that can reconstruct training data [22, 53], limiting model sharing for research and clinical deployment. The inherent tension between improving model utility and ensuring robust privacy preservation represents a challenging research problem [2, 19, 69] that has not been adequately explored in the context of HPE.

Differential privacy (DP) provides a principled framework for mitigating these risks by offering provable guarantees against information leakage from both data and model

parameters [1, 17, 18, 23]. However, implementing DP through DP Stochastic Gradient Descent (DP-SGD) [1], typically results in substantial performance degradation, which is particularly problematic for fine-grained vision tasks like HPE where spatial precision is paramount [13, 15, 64].

In this work, we present the first systematic framework for differentially private learning in 2D Human Pose Estimation. We demonstrate that directly applying DP-SGD to 2D-HPE models leads to significant degradation in utility due to the fine-grained nature of keypoint prediction. We address the privacy-utility trade-off in 2D-HPE, through two complementary mechanisms. First, we employ projection-based DP-SGD that constrains noisy gradient updates to a learned k-dimensional subspace ($k \ll d$), reducing noise variance substantially. Second, we integrate Feature Differential Privacy (FDP), which relaxes differential privacy by decomposing gradient updates into public and private components, adding noise only to sensitive features. Finally, we propose a hybrid strategy that effectively combines projection and FDP, yielding multiplicative utility gains.

To summarise, our core contributions are as follows:

- **First systematic DP benchmark for pose estimation:** We establish comprehensive baselines for differentially private 2D-HPE across privacy budgets ($\varepsilon \in \{0.2, 0.4, 0.6, 0.8\}$), clipping thresholds ($C \in \{0.01, 1.0\}$), and training strategies on MPII and HumanART datasets.
- **Feature-Projective Private Learning:** We propose a joint mechanism that integrates two complementary noise reduction strategies, First, a public dataset is used to identify a low dimensional gradient subspace to filter noise. Second, from Feature Differential Privacy (FDP), we define the entire raw image as private and add noise only to this while simultaneously using its corresponding public feature to compute a noise free gradient that helps with improved utility.
- **Convergence Analysis of Feature-Projective DP:** We show theoretically that the combined effect of projection and FDP is multiplicative in terms of signal-to-noise ratio and convergence speed.

We conduct extensive experiments across diverse privacy budgets, clipping thresholds and training strategies on both MPII and HumanART datasets. The proposed feature-projective framework outperforms vanilla DP-SGD, demonstrating superior privacy-utility trade-offs.

2. Related Work

2.1. 2D Human Pose Estimation and Privacy

Markerless 2D-HPE identifies anatomical keypoints in images without physical markers, playing a fundamental role in human motion analysis for healthcare and activity recognition [10, 14]. While traditional heatmap methods based on CNN architectures [5, 33, 56, 59] and vision-based transformers [37, 38, 48, 60] achieve state-of-the-art performance, they suffer from quantization errors and usually result in large cumbersome networks that are prone to memorization. Regression approaches offer faster, end-to-end solutions but with reduced accuracy [36, 47, 54]. Coordinate classification approach addresses some of these limitations by treating pose estimation as classification over discretized coordinates [39], achieving strong performance with computational efficiency. Knowledge distillation techniques [8, 40, 62, 66] further enhance efficiency and inference speed by transferring knowledge from large models to compact architectures.

Protecting user privacy remains critical yet challenging for HPE, which relies on high-quality images. Recent work demonstrates that adversaries can reconstruct substantial portions of private training data solely by analyzing the parameters or gradients of a trained neural network [22, 53, 70]. Consequently, data sharing in sensitive domains provide limited information, preventing analysis of crucial clinical indicators that require body shape or facial information for stress assessment and abnormal motion detection. Most platforms implement rudimentary privacy protection through face blurring and pixelation [49], while more sophisticated methods include skin removal [25] and template-based shape modeling, and visual privacy layers that degrade private attributes [26]. However, these ad hoc methods lack formal privacy guarantees and suffer from the "onion effect" [11], where removing one protection layer exposes previously secured features, making them vulnerable to privacy attacks [65]. Recent GAN-based anonymization techniques [28–30] generate realistic anonymized figures, yet introduce artifacts due to pose detection errors and contextual mismatches. Adversarial learning approaches [43, 61] attempt to learn privatized features that obscure sensitive attributes while preserving task utility. However, these methods assume original data are unnecessary post-training, compromising interpretability, which is essential for clinical validation and diagnosis in healthcare applications. Moreover, even advanced anonymization cannot fully replace real data for training robust computer vision models [29], highlighting the need for formal privacy frameworks like differential privacy that provide quantifiable guarantees without sacrificing data authenticity.

2.2. Utility vs Privacy with DP optimization

DP-SGD [1] is one of the most common methods in developing privacy-preserved deep learning models because of the strong privacy guarantees compared to data-independent methods and its ability to scale to large datasets [17, 23].

However, DP settings induce a fundamental privacy-utility trade-off that compromise practical deployment of privacy preserved HPE [2]. DP-SGD performance depends on loss function smoothness [55], gradient dimensionality [19, 69] and clipping threshold selection [35]. It has also been shown that if the loss function lacks Lipschitz continuity, the performance of DP-SGD critically depends on carefully selecting an appropriate clipping threshold; otherwise, performance will not improve significantly, regardless of the amount of training data or iterations [19]. Recent research has sought to improve the utility-privacy trade-off by relaxing traditional differential privacy [21, 44, 51]. [51] enhanced the utility of DP by introducing Selective Differential Privacy, which protects only sensitive tokens within language models, while leaving non-sensitive elements unperturbed. [44] proposed Feature Differential Privacy, a generalized DP framework that explicitly categorizes features as either protected or public, enabling targeted noise application that enhances utility. Both methods demonstrate that targeted protection of sensitive attributes substantially mitigates the privacy-utility trade-off compared to classical DP.

3. Methodology

3.1. Preliminaries: Differentially Private Stochastic Gradient Descent (DP-SGD)

Privacy-preserving machine learning requires a rigorous mathematical framework to quantify privacy guarantees. DP provides such a framework by measuring in the context of databases how much the inclusion or exclusion of a single data point can influence the output of a randomized algorithm [17, 23]. This comparison allows us to bound the information leakage about any individual data point.

Consider two neighboring datasets-identical except for a single sample. The level of DP ensured by a randomized algorithm $\mathcal{M}$ is provided by the following definition.

Definition 1: (ε, δ)-Differential Privacy: A randomized algorithm $\mathcal{M}$ with domain $\mathcal{D}$ and range $\mathcal{R}$ is said to be (ε, δ)-DP if, for any subset $S \subseteq \mathcal{R}$ and for any neighboring datasets $d, d' \in \mathcal{D}$, the following condition holds:

$$\mathbb{P}[\mathcal{M}(d) \in S] \leq e^{\varepsilon} \cdot \mathbb{P}[\mathcal{M}(d') \in S] + \delta \tag{1}$$

In this definition, ε (privacy budget) controls the strength of the privacy guarantee. Smaller values of ε provide stronger privacy. The parameter δ represents the probability that the privacy guarantee fails and is typically set to be cryptographically small.

In the context of HPE, differential privacy ensures that the inclusion or exclusion of a single training image, which potentially contains identifiable biometric information, does not significantly affect the model's learned parameters or predictions. Therefore, DP trained models provide a formal measure of privacy protection [17], effectively mitigating risks from various attacks, including membership inference[52] or reconstruction attacks[70].

A common approach for ensuring differential privacy during neural network training is DP-SGD, which enforces an (ε, δ)-DP guarantee on gradient updates [1, 9, 16]. This mechanism involves clipping gradients to a fixed $L2$-norm threshold (C) and adding Gaussian noise calibrated based on desired privacy budget (ε, δ). This process ensures that no single training sample can disproportionately influence the model update [34].

3.2. Architecture

For our 2D-HPE models, we adopt TinyViT [58] as the backbone. This compact, four-stage efficient hierarchical vision transformer is well suited for resource-constrained vision tasks. Its smaller size is highly beneficial as the error bounds of DP-SGD are known to scale with number of parameters[7]. The model adopts a multi-stage architecture wherein the spatial resolution is progressively reduced and the feature representation expands. TinyViT follows a hybrid architectural design containing convolutional layers at the initial stages followed by self-attention mechanisms. Unlike standard ViT models, TinyViT employs a two-layer convolutional embedding. In the first stage of the network, it employs MBConv [27] blocks from MobileNetV2 to efficiently learn the low-level representation. The last three stages consist of transformer blocks hierarchically. Each stage consists of multi-head-self-attention (MHSA) layers, feed forward network (FFN) and 3 x 3 depthwise convolutions between the MHSA and FFN layers.

For keypoint localization, we augment the TinyViT backbone model with a coordinate classification output stage [39]. Given an input image $I \in \mathbb{R}^{C \times H \times W}$ and a ground truth keypoint $p_i = (x_i, y_i)$ for the i^{th} joint, the continuous coordinates are quantized into discrete bins via a splitting factor $k \geq 1$. Formally, the quantized coordinates are computed as:

$$p_i' = \left(\lfloor x_i \cdot k \rfloor \, , \, \lfloor y_i \cdot k \rfloor \right),$$

where $\lfloor \cdot \rfloor$ denotes the rounding operation. This binning reduces quantization error while preserving high localization precision. The complete architecture is depicted in Figure 1.

Within our network, the Convolutional head produces a 16-channel feature map, with each channel corresponding to a specific joint. These joint-specific features are Upsampled and flattened to form a compact representation used for classification over the discrete coordinate bins. To improve robustness, we employ Gaussian label smoothing on the classification targets. This smoothing accounts for spatial correlations by assigning soft labels that reflect the rel-

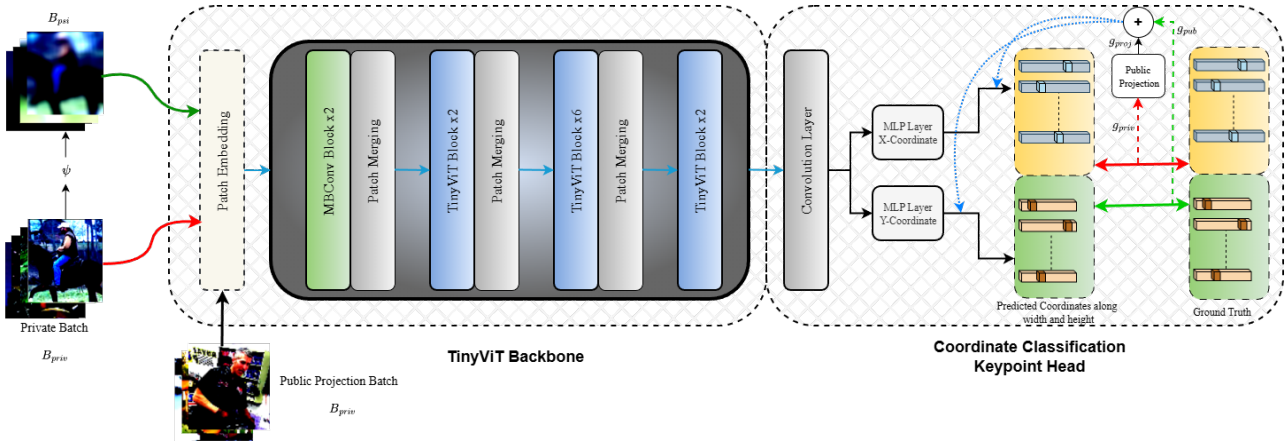

Figure 1. Overview of our private HPE pipeline coupling a TinyViT based backbone with Coordinate Classification Keypoint head. The Public feature batch B_{pub} is generated from the private batch B_{priv} using ψ both of which are given as input in a single iteration. Additionally, a public image set B_{pub}^{proj} independent of B_{priv} is used to calculate public gradients for projection at specified intervals. Red Arrow indicates the propagation of private gradients and Green Arrow indicates propagation of public feature gradients. Blue Dotted Arrow indicates the propagation of cumulative denoised gradient from Feature Projective DP.

evance of neighboring bins. Finally, the discrete classification outputs are decoded back into continuous coordinates to yield the final keypoint predictions.

3.3. Projection Based DP-SGD

Training dynamics in deep networks exhibit intrinsic low-dimensional structure, where meaningful gradient updates concentrate within a subspace significantly smaller than the full parameter space. We leverage this by identifying and projecting noisy gradients onto informative subspaces, filtering out less relevant directions while preserving signal quality under differential privacy constraints. In this way, we preserve the signal quality of gradient updates while adhering to DP constraints [68]. To estimate the intrinsic structure of the gradient space, we employ a small auxiliary public dataset S_{pub}, which is drawn from a similar distribution as that of private training set. This subset is used to estimate the principal subspace of the gradient covariance. Given the model parameters $w \in \mathbb{R}^p$, the second moment matrix of gradients over S_{pub} is calculated as:

$$M(w) = \frac{1}{m} \sum_{i=1}^{m} \nabla l(w, \tilde{z}_i) \nabla l(w, \tilde{z}_i)^T \tag{2}$$

where m denotes the number of public samples and $\tilde{z}_i$ represents an input sample from S_{pub}. The eigenvectors corresponding to the top k eigenvalues are stacked to form the projection matrix $\hat{V} \in \mathbb{R}^{p \times k}$ which forms the low-dimensional approximation of the full gradient space; this maps the p-dimensional gradients to a smaller k-dimensional subspace. This projection matrix is updated periodically to accommodate changes in gradient distributions over the training period.

In the DP-SGD setup, for each mini-batch sampled from the private dataset S_{priv}, per-sample gradients are computed and the sensitivity of each individual gradient is bounded by the clipping threshold C:

$$\tilde{g}_i = clip(\nabla l(w, z_i), C) \tag{3}$$

The clipped gradients are aggregated over the batch and Gaussian noise is added to ensure differential privacy

$$g = \frac{1}{B} \left(\sum_{i \in B} \tilde{g}_i + \mathcal{N}(0, \sigma^2 C^2 \mathbf{I}) \right), \tag{4}$$

where B is the size of the mini-batch and σ is the standard deviation. The full noisy gradient g is then projected onto the estimated low-dimensional subspace as

$$g_{proj} = (\hat{V}\hat{V}^T)g \tag{5}$$

This restricts the update direction to the subspace where the gradients exhibit the highest variance, thereby filtering out noise components residing in less informative directions. The model parameters are then updated using the projected gradient. Since the projection is applied as a post-processing step after noise addition, the overall DP guarantee remains intact.

3.4. Feature Projective DP-SGD for HPE

3.4.1. Feature Differential Privacy

To enhance the model utility in our 2D HPE task, we extend the standard DP-SGD framework using Feature Differential Privacy (FDP). FDP exploits the transformation of the training image into private and public variants, selectively applying differential privacy only to sensitive features while

freely utilizing non-sensitive(public) information [44]. Formally, let each sample be $x_i \in S_{data}$(a raw training image with keypoint labels), and let $\psi : S \to \mathcal{F}$ be a public feature map such that $\psi(x_i)$ is the public variant of the raw image x_i and let $f : [0,1] \to [0,1]$ be a trade-off function. A randomized mechanism $\mathcal{M}$ satisfies f-FDP with respect to ψ if, for any two datasets d, d' differing in exactly one image-label pair $x_i \neq x_i'$ but having identical public representations $(\psi(x_i) = \psi(x_i'))$ and for all subsets S for range of $\mathcal{M}$:

$$\mathbb{P}[\mathcal{M}(d) \in S] \leq 1 - f(\mathbb{P}[\mathcal{M}(d') \in S]) \tag{6}$$

Then, we say the mechanism is (ε, δ)-DP with respect to ψ iff it is f-FDP for $f(x) = 1-\delta-e^{\varepsilon}x$. Motivated by this definition, the FDP-SGD method, explicitly distinguishes between public and private (raw) images to improve pose estimation accuracy under the same privacy budget as standard DP. Specifically, for each image-keypoint pair, we define a public loss $l_{pub}(w, \psi(x))$ which captures the coarse pose estimation based on the definition of ψ. The private loss $l_{priv}(w, x)$ captures the sensitive, fine-grained details of the human that requires privacy protection. Then the overall loss can be given as:

$$l(w, x) = l_{priv}(w, x) + l_{pub}(w, \psi(x)) \tag{7}$$

3.4.2. Training with Feature-Projective DP

To maximize the privacy utility tradeoff, we introduce Feature-Projective DP, a hybrid approach that integrates the two approaches outlined in Sections 3.3, 3.4. The complete algorithmic details of this integrated approach are provided in Algorithm 1.

This approach synergizes two key ideas: First we adopt the FDP framework to decompose the total loss into a public component l_{pub} (computed on public features $\psi(x)$) and a private component l_{priv} (computed on raw image x). This ensures that DP noise is added only to the gradient of sensitive private component. Second, we apply the projection technique to filter noise, restricting the private gradient update to the most informative k-dimensional subspace. As shown in Algorithm 1, our Feature-projective DP method proceeds at each iteration t by first sampling two separate and independent batches from S_{data}. On the public batch B^t_{psi} we compute the gradient as:

$$g^t_{pub} = \frac{1}{|B^t_{psi}|} \sum_{x \in B^t_{psi}} \nabla l_{pub}(w_{t-1}, \psi(x)) \tag{8}$$

Similarly, on the private batch B^t_{priv}, we compute and clip the gradient of the private loss to the clipping norm C as $\tilde{g}$, then aggregate and add gaussian noise:

$$g^t_{priv} = \frac{1}{|B^t_{priv}|} \left(\sum_{x \in B^t_{priv}} \tilde{g} + \mathcal{N}(0, \sigma^2 C^2 I) \right) \tag{9}$$

We then denoise g^t_{priv} by applying the subspace projection from Eq.5 as a post-processing step given as:

$$g^t_{proj} = (\hat{V}_t \hat{V}_t^T) g^t_{priv} \tag{10}$$

The final gradient update g_t is the sum of the clean public component and the denoised private component. The model parameters are then updated as:

$$g_t = g^t_{pub} + g^t_{proj} \tag{11}$$

$$w_t = w_{t-1} - \eta_t g_t \tag{12}$$

where η_t denotes the learning rate.

3.4.3. Convergence Analysis of Feature-Projective DP

The convergence analysis of our method formally establishes the utility gain as observed from our empirical results and is a direct corollary of the separate analyses from [44, 68].

Let the empirical risk be $\hat{L}_n(w) = \frac{1}{n}\sum_{i=1}^{n} l(w, x_i)$ on a private dataset S_{priv} of size n.

Assumption 1. The loss $l(w, x)$ can be decomposed into public and private components as given in Eq. 7.

Assumption 2. The full loss $\hat{L}_n(w)$ is ρ-smooth, the full gradient $||\nabla l(w, x)||_2 \leq G$ is bounded where G defines the sensitivity for subspace reconstruction error and the private gradient is bounded by the threshold C as $||\nabla l_{priv}(w, x)||_2 \leq C$, where $C \leq G$.

Assumption 3. We have access to a separate public dataset S_{pub} of size m and $\hat{V}_t \in \mathbb{R}^{p \times k}$ is the k-dimensional projection matrix computed from top-k eigenspace (from Eq. 2) on S_{pub} at iteration w_{t-1}.

Assumption 4. Assuming the principal component of the gradient dominance condition is satisfied and under this, we denote the eigengap at iteration t as α_t and $\Lambda = \frac{1}{T}\sum_{t=1}^{T} 1/\alpha_t^2$ be average inverse squared eigengap and refer to $\gamma_2(\mathcal{W}, d_w)$ as the associated complexity measure (where $\mathcal{W}$ iterate set of the weights and d_w is distance between them), as defined in [69].

Under these assumptions, setting the total iterations $T = \mathcal{O}(n^2\varepsilon^2)$, the average expected gradient norm of feature-projective DP is bounded by:

$$\frac{1}{T}\sum_{t=1}^{T} \mathbb{E}||\nabla \hat{L}_n(w_t)||_2^2 \leq \underbrace{\tilde{\mathcal{O}}\left(\frac{k \cdot \rho \cdot C^2}{n\varepsilon}\right)}_{\text{Privacy Error}} + \underbrace{\mathcal{O}\left(\frac{\Lambda G^4 \rho^2 \gamma_2^2(\mathcal{W}, d_w) \ln p}{m}\right)}_{\text{Reconstruction Error}} \tag{13}$$

The convergence is bound by two terms: a reconstruction error inherited from use of public dataset S_{pub} and privacy error from the gaussian noise. By combining both the approaches, the privacy error scales with both the reduced dimension k and reduced gradient norm C which can be understood from the error bound changing from $\tilde{\mathcal{O}}(p \cdot G^2) \to \tilde{\mathcal{O}}(k \cdot C^2)$ which explains the feature-projective DP's higher utility for the same (ε, δ)-FDP guarantee.

Algorithm 1 Feature Projective DP-SGD

Require: Full dataset $\mathcal{S}_{\text{data}} = \{x_1, \ldots, x_n\}$, split into public subset $\mathcal{S}_{\text{pub}} \subset \mathcal{S}_{\text{data}}$ (size m) and private remainder $\mathcal{S}_{\text{priv}} = \mathcal{S}_{\text{data}} \setminus \mathcal{S}_{\text{pub}}$, public feature map ψ, combined loss $\ell(w, z)$ with public and private losses l_{pub}, l_{priv}, clip norm C, noise std. σ, subspace dim k, batch size B, iterations T, learning rate $\{\eta_t\}$.

Initialize model parameters $w_0 \in \mathbb{R}^p$.

for $t = 1, \ldots, T$ **do**

(1) Subspace identification on $\mathcal{S}_{\text{pub}}$:

Compute

$$M_t = \frac{1}{m} \sum_{z \in \mathcal{S}_{\text{pub}}} \nabla \ell(w_{t-1}, \tilde{z}) \, \nabla \ell(w_{t-1}, \tilde{z})^\top.$$

Compute the top-k eigenvectors of M_t.

Form the subspace basis $\hat{V}_t \in \mathbb{R}^{p \times k}$.

Compute the projector $\hat{V}_t \, \hat{V}_t^{\top} \in \mathbb{R}^{p \times p}$.

(2) Compute public and private feature gradient:

Sample public batch $B^t_{psi} \subset \psi(x) : x \in S_{priv}$

Compute

$$g^t_{pub} = \frac{1}{|B^t_{psi}|} \sum_{x \in B^t_{psi}} \nabla l_{pub}(w_{t-1}, \psi(x))$$

Sample private batch $B^t_{priv} \subset S_{priv}$

Compute the clipped gradient $\tilde{g}_t$, aggregate and add Gaussian noise

$$g^t_{priv} = \frac{1}{|B^t_{priv}|} \left(\sum_{x \in B^t_{priv}} \tilde{g}_t + \mathcal{N}(0, \sigma^2 C^2 I) \right)$$

Project: $g^t_{proj} = (\hat{V}_t \, \hat{V}_t^{\top}) \cdot g^t_{priv} \in \mathbb{R}^p$.

Merge Public and Private projected feature gradients

$$g_t = g^t_{pub} + g^t_{proj}$$

Update: $w_t = w_{t-1} - \eta_t \, g_t$.

end for

return w_T.

4. Experiments

4.1. Dataset and Implementation Details

In our experiments, we evaluated our framework on two widely used human pose datasets: MS COCO Keypoint Dataset [41] and MPII dataset [4]. Our methodology assumes that the COCO dataset serves as a public dataset used for pre-training the network weights, while MPII/ HumanART functions as a private dataset on which we apply the differential privacy techniques.

Specifically, our models are pretrained on the COCO *train2017* set, which consists of approximately 118k images with around 140k annotated human instances, each with 17 joint annotations. The *val2017* set consisting of around 5k images is used for validation. For evaluating the trade-off between utility and performance under various DP-SGD techniques we employ the MPII Human Pose Dataset consisting of 40k human instances, each labeled with 16 joint annotations. When transferring the model from COCO to MPII, we adjust for the keypoint discrepancy between datasets. We employ the Percentage of Correct Keypoints normalized by head (PCKh) [4] as an evaluation metric.

To further assess the generalization of our privacy-preserved pose estimation models under domain shift and visual diversity, we conduct additional experiments on the Human-Art dataset [32]. Human-Art is a recently introduced, large-scale human-centric benchmark designed to bridge natural and artificial visual domains. It contains 50,000 high-quality images with over 123,000 person instances across 20 diverse scenarios, spanning natural scenes (e.g., cosplay, drama, dance) and a wide spectrum of artistic styles (e.g., oil paintings, sculptures, digital art, watercolor, and murals). Compared to conventional datasets like MPII or COCO, Human-Art presents significantly greater challenges for pose estimation due to the presence of stylized or abstract human depictions, exaggerated or distorted body proportions, occlusions, artistic textures, and unconventional poses. We follow the standard MS COCO evaluation protocol and report the Average Precision (AP) as the primary metric.

Our experimental framework explores three distinct DP training scenarios: Fine-Tuning with frozen backbone, Full Fine-Tuning and, Training from scratch. For the first scenario, we specifically freeze the first three stages of the backbone and finetune the fourth stage and all instances of layer norm[13]. To generate the public feature map, we employ Gaussian blur as ψ which effectively suppresses facial and body structure details. Details on datasets, training and privacy related parameters are provided in the supplementary.

4.2. Results and Analysis on MPII dataset

For comprehensive comparison, all experimental results across training strategies, clipping thresholds and privacy budgets are visualized in Figure 2.

[1]Full tabular results are provided in the supplementary material.

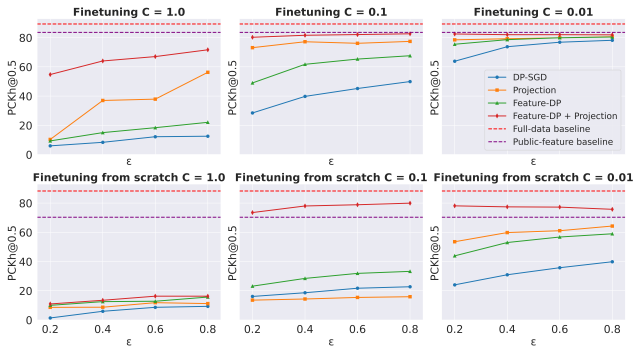

Figure 2. Comparison of PCKh@0.5 on MPII dataset across private and non-private methodologies under different training strategies with varied privacy budget (ε) and clipping thresholds (C).[1]

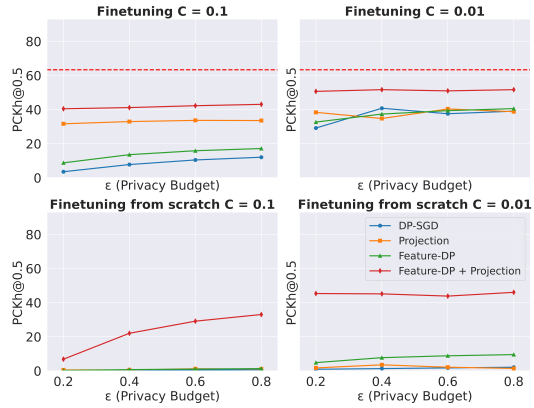

Figure 3. Comparison of AP on HumanART dataset across private and non-private methodologies under different training strategies with varied privacy budget (ε) and clipping thresholds (C).[1]

4.2.1. Non-Private Baseline Results

Table 1 presents baseline pose estimation performance of our model on the MPII dataset under three training strategies: (i) finetuning from a COCO-pretrained model, (ii) finetuning from scratch (initialization with COCO pretrained weights and all layers are trained), and (iii) training from scratch (random initialization). Additionally, we report results from using only public features (blurred images) under the same strategies to provide context for evaluating privacy-utility trade-offs. As expected, the finetuning strategy achieves the highest mean accuracy of 89.36% followed by finetuning from scratch (88.28%) and training from scratch (76.89%), which is to be expected. These non-private baselines establish upper bound performance references for evaluating differential privacy impact. When the model relies only on public features(gaussian blurred images), performance reduces significantly. While finetuning on public features maintain reasonable accuracy, the other training strategies yield substantially compromised results, confirming that fine-grained visual details in raw images are critical for accurate pose estimation.

Table 1. MPII Results: Non-Private Baselines for our HPE model on the MPII dataset

Training Strategy	Head	Shoulder	Elbow	Wrist	Hip	Knee	Ankle	Mean	Mean@0.1
Finetuning	97.07	95.86	89.59	83.61	89.29	85.31	81.48	89.36	31.33
Finetuning from scratch	96.45	95.84	88.07	82.18	88.78	83.01	79.45	88.28	28.11
Training from scratch	93.89	89.32	75.34	65.24	80.04	66.15	60.60	76.89	17.26
Finetuning on Public features	94.30	93.95	83.21	75.19	83.53	76.86	72.76	83.61	20.81
Finetuning from scratch on Public features	88.71	84.32	69.71	54.37	70.24	61.67	55.12	70.32	11.36
Training from scratch on Public features	15.31	20.01	17.19	12.59	25.04	16.90	10.91	17.99	0.78

4.2.2. DP-SGD Baseline Results

Figure 2 presents the PCKh@0.5 results on MPII dataset under the aforementioned training strategies using DP-SGD. Experiments were conducted across multiple settings with varying privacy parameters ($\varepsilon \in \{0.2, 0.4, 0.6, 0.8\}$) and clipping thresholds ($C \in \{0.01, 0.1, 1.0\}$). For standard finetuning with DP-SGD, lower clipping thresholds consistently yield better pose estimation results across different privacy levels. Specifically, at $C = 0.01$, the model achieves substantially higher accuracy of 63.85% mean PCKh@0.5 at the tightest privacy loss($\varepsilon = 0.2$) compared to $C = 0.1$ (28.46%) and $C = 1.0$ (5.94%). This is indeed because of the fact that the effective noise magnitude grows linearly with the C thus our results confirm this.

Notably, finetuning the COCO-pretrained TinyViT backbone significantly mitigates the DP induced performance degradation compared to training from scratch or finetuning from scratch [63]. This indicates that pretrained human pose based feature representations provide robust feature priors that enable DP-SGD to adapt effectively to private pose datasets, while maintaining resilience to noise corruption.

4.2.3. Performance Analysis of Subspace Projection

We maintain identical training strategies and privacy parameters to ensure direct comparison with both non-private and DP-SGD baseline methods. Our subspace projection approach demonstrates substantial performance improvements across multiple configurations. At the most restrictive clipping threshold ($C = 0.01$), projection yields significant gains from 63.85% to 78.48% at $\varepsilon = 0.2$ and from 78.17% to 80.63% at $\varepsilon = 0.8$. This enhancement occurs because, while Gaussian noise is injected uniformly across all gradient components, only a subset of directions carry meaningful pose-relevant information. By projecting onto the learned subspace, we effectively discard the noise in irrelevant directions, thereby improving signal-to-noise ratio and preserving essential pose estimation features. At $C = 0.1$, the projection approach consistently outperforms baseline DP-SGD. Finetuning increases accuracy from 73.13% to 77.41%, while training from scratch improves from 9.80% to 13.05%. However, for the fine-

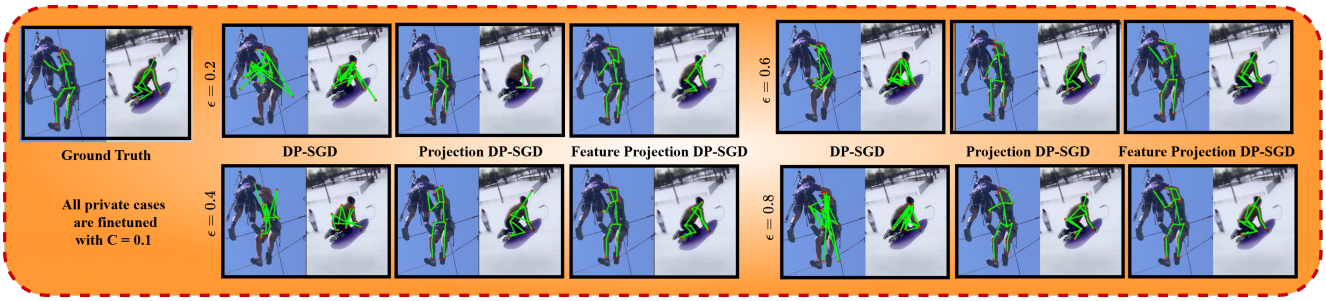

Figure 4. Depiction of qualitative results on DP-SGD, Projection DP-SGD and Feature Projection DP-SGD. We specifically show results on Finetuning with $C = 0.1$ at various privacy budgets.

tuning from scratch strategy, the curve plateaus slightly below regular DP-SGD. We attribute this phenomenon to the interaction between injected gaussian noise and subspace reconstruction error[68]. At the largest clipping threshold ($C = 1.0$), we observe non-monotonic patterns. Under this condition, the raw gradients become dominated by noise, leading to unstable parameter updates and local dips in accuracy.

4.2.4. Performance analysis of FDP and Feature-Projective DP

Feature DP consistently outperforms vanilla DP-SGD across all experimental configurations. Under finetuning with $C = 0.01$, FDP achieves substantial improvements, from 63.85% to 75.46% at $\varepsilon = 0.2$ (11.61% gain) and from 78.17% to 80.40% at $\varepsilon = 0.8$ (2.23% gain). This is consistently observed across all training strategies and clipping values. Integrating FDP with subspace projection results in the highest accuracy across all experimental settings. Even under the most challenging conditions with stringent clipping of $C = 1.0$, where standard DP-SGD achieves only 12.53%, Feature-projective DP attains 71.66%, representing a six fold relative gain.

The largest improvements occur when training from scratch. With $C = 0.1$ and $\varepsilon = 0.8$, vanilla DP-SGD achieves merely 6.85% accuracy, while FDP alone attains 11.22%. However, the combined Feature-Projective DP approach achieves 33.48%. This demonstrates that combining both techniques boosts utility drastically especially in large noise induced scenarios, where neither alone suffices to recover strong pose features from corrupted gradients. Figure 4 depicts few qualitative results across different privacy strategies along with ground truth.

4.3. Cross-Dataset Evaluation on HumanART

We further evaluate our feature-projective DP framework on the HumanART dataset, which contains stylized and artistic human figures with substantial visual domain shifts relative to natural images in MPII. As shown in Figure 3, our method maintains a strong privacy-utility balance across all privacy budgets, achieving 51.6 mAP at $\varepsilon = 0.8$ with finetuning strategy at $C = 0.01$. For clarity, we report only finetuning and finetuning from scratch at $C = \{0.01, 0.1\}$, as these are the only settings that yield stable and practically useful performance. Training from scratch or using $C = 1.0$ yields negligible accuracy. Results are visualised in Figure 3 while full tabular results are provided in supplementary. Notably, under non-private training, finetuning from scratch achieves higher accuracy (69.5 mAP) than finetuning by freezing the backbone (63.3 mAP), as expected from the greater capacity for task-specific adaptation. However, this trend reverses once DP is applied. We attribute this behavior to the addition of DP noise and clipping where updating a smaller subset of parameters concentrates the effective learning while reducing total injected noise. This observation aligns with prior findings[50] that DP noise disproportionately harms larger parameter regimes.

5. Conclusion

Our work presents the first differentially private (DP) approach to 2D human pose estimation (HPE), addressing critical privacy concerns while maintaining utility. Our results clearly establish that the synergistic combination of feature-level privacy and subspace projection dramatically enhances utility across all settings. Importantly, our proposed Feature-Projective DP 2D-HPE approach achieved up to 82.62% mean PCKh@0.5 on MPII and 51.6 mAP on HumanART at $\epsilon = 0.8$, significantly narrowing the gap to non-private performance under strong formal privacy guarantees. Crucially, the proposed approach requires no manual curation of private features, as it automatically protects the entire raw image, ensuring privacy preservation for both individuals and their spatial environments.

6. Acknowledgments

We acknowledge funding from EPSRC(EP\W01212X\1), the Royal Society (RGS\R2\212199) and Academy of Medical Sciences (NGR1\1678).

References

[1] Martin Abadi, Andy Chu, Ian Goodfellow, H Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, pages 308–318. ACM, 2016. 2, 3

[2] Wisam Abbasi, Paolo Mori, and Andrea Saracino. Trading-Off Privacy, Utility, and Explainability in Deep Learning-Based Image Data Analysis . *IEEE Transactions on Dependable and Secure Computing*, 22(01):388–405, 2025. 1, 3

[3] Shafiq Ahmad, Pietro Morerio, and Alessio Del Bue. Event anonymization: privacy-preserving person re-identification and pose estimation in event-based vision. *IEEE Access*, 2024. 1

[4] Mykhaylo Andriluka, Leonid Pishchulin, Peter Gehler, and Bernt Schiele. 2d human pose estimation: New benchmark and state of the art analysis. In *Proceedings of the IEEE Conference on computer Vision and Pattern Recognition*, pages 3686–3693, 2014. 6

[5] Bruno Artacho and Andreas Savakis. Unipose: Unified human pose estimation in single images and videos. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7035–7044, 2020. 1, 2

[6] Simone Barattin, Christos Tzelepis, Ioannis Patras, and Nicu Sebe. Attribute-preserving face dataset anonymization via latent code optimization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8001–8010, 2023. 1

[7] Raef Bassily, Adam Smith, and Abhradeep Thakurta. Private empirical risk minimization: Efficient algorithms and tight error bounds. In *2014 IEEE 55th annual symposium on foundations of computer science*, pages 464–473. IEEE, 2014. 3

[8] Fani Deligianni Bhargav Sivangi. Knowledge distillation with global filters for efficient human pose estimation. In *British Machine Vision Conference(BMVC)*, 2024. 2

[9] Franziska Boenisch, Antoni Kowalczuk, Jan Dubinski, Atiyeh Ashari Ghomi, Yi Sui, George Stein, Jiapeng Wu, Jesse C Cresswell, and Adam Dziedzic. Benchmarking robust self-supervised learning across diverse downstream tasks. *CoRR*, abs/2407.12588, 2024. 3

[10] Rohit Kumar Bondugula, Siba K Udgata, and Kaushik Bhargav Sivangi. A novel deep learning architecture and minirocket feature extraction method for human activity recognition using ecg, ppg and inertial sensor dataset. *Applied Intelligence*, 53(11):14400–14425, 2023. 2

[11] Nicholas Carlini, Matthew Jagielski, Chiyuan Zhang, Nicolas Papernot, Andreas Terzis, and Florian Tramer. The privacy onion effect: Memorization is relative. In *Advances in Neural Information Processing Systems*, pages 13263–13276. Curran Associates, Inc., 2022. 1, 2

[12] Aviral Chharia, Wenbo Gou, and Haoye Dong. Mv-ssm: Multi-view state space modeling for 3d human pose estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025. 1

[13] Soham De, Leonard Berrada, Jamie Hayes, Samuel L Smith, and Borja Balle. Unlocking high-accuracy differentially private image classification through scale. *arXiv preprint arXiv:2204.13650*, 2022. 2, 6

[14] Fani Deligianni, Yao Guo, and Guang-Zhong Yang. From emotions to mood disorders: A survey on gait analysis methodology. *IEEE journal of biomedical and health informatics*, 23(6):2302–2316, 2019. 2

[15] Jiawei Duan, Haibo Hu, Qingqing Ye, and Xinyue Sun. Analyzing and optimizing perturbation of dp-sgd geometrically. In *2025 IEEE 41st International Conference on Data Engineering (ICDE)*, pages 3439–3452, 2025. 2

[16] Lionel Dupuy, Sergio Yovine, Federico Pan, Nicolas Basset, and Thao Dang. Towards efficient active learning of pdfa. In *Proceedings of the 2022 International Conference on Learning Representations*. ICLR, 2022. 3

[17] Cynthia Dwork. The promise of differential privacy: a tutorial on algorithmic techniques. In *2011 IEEE 52nd Annual Symposium on Foundations of Computer Science, D (Oct. 2011)*, pages 1–2. Citeseer, 2021. 2, 3

[18] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography*, pages 265–284, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg. 2

[19] Huang Fang, Xiaoyun Li, Chenglin Fan, and Ping Li. Improved convergence of differential private SGD with gradient clipping. In *The Eleventh International Conference on Learning Representations*, 2023. 1, 3

[20] Jonas Geiping, Hartmut Bauermeister, Hannah Dröge, and Michael Moeller. Inverting gradients-how easy is it to break privacy in federated learning? *Advances in neural information processing systems*, 33:16937–16947, 2020. 1

[21] Aditya Golatkar, Alessandro Achille, Yu-Xiang Wang, Aaron Roth, Michael Kearns, and Stefano Soatto. Mixed differential privacy in computer vision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 3

[22] Niv Haim, Gal Vardi, Gilad Yehudai, michal Irani, and Ohad Shamir. Reconstructing training data from trained neural networks. In *Advances in Neural Information Processing Systems*, 2022. 1, 2

[23] Moritz Hardt and Kunal Talwar. On the geometry of differential privacy. In *Proceedings of the forty-second ACM symposium on Theory of computing*, pages 705–714, 2010. 2, 3

[24] Ali Hatamizadeh, Hongxu Yin, Pavlo Molchanov, Andriy Myronenko, Wenqi Li, Prerna Dogra, Andrew Feng, Mona G Flores, Jan Kautz, Daguang Xu, et al. Do gradient inversion attacks make federated learning unsafe? *IEEE Transactions on Medical Imaging*, 42(7):2044–2056, 2023. 1

[25] Nikolas Hesse, Christoph Bodensteiner, Michael Arens, Ulrich G. Hofmann, Raphael Weinberger, and A. Sebastian Schroeder. Computer vision for medical infant motion analysis: State of the art and rgb-d data set. In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, 2018. 1, 2

[26] Carlos Hinojosa, Juan Carlos Niebles, and Henry Arguello. Learning privacy-preserving optics for human pose estimation. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 2553–2562, 2021. 2

[27] Andrew Howard, Mark Sandler, Grace Chu, Liang-Chieh Chen, Bo Chen, Mingxing Tan, Weijun Wang, Yukun Zhu, Ruoming Pang, Vijay Vasudevan, Quoc V. Le, and Hartwig Adam. Searching for mobilenetv3. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019. 3

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

[29] Håkon Hukkelås and Frank Lindseth. Does image anonymization impact computer vision training? In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 140–150, 2023. 2

[30] Håkon Hukkelås, Morten Smebye, Rudolf Mester, and Frank Lindseth. Realistic full-body anonymization with surface-guided gans. In *Proceedings of the IEEE/CVF Winter conference on Applications of Computer Vision*, pages 1430–1440, 2023. 2

[31] Marija Jegorova, Chaitanya Kaul, Charlie Mayor, Alison Q. O'Neil, Alexander Weir, Roderick Murray-Smith, and Sotirios A. Tsaftaris. Survey: Leakage and privacy at inference time. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(7):9090–9108, 2023. 1

[32] Xuan Ju, Ailing Zeng, Jianan Wang, Qiang Xu, and Lei Zhang. Human-art: A versatile human-centric dataset bridging natural and artificial scenes. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 618–629, 2023. 6

[33] Aouaidjia Kamel, Bin Sheng, Ping Li, Jinman Kim, and David Dagan Feng. Hybrid refinement-correction heatmaps for human pose estimation. *IEEE Transactions on Multimedia*, 23:1330–1342, 2020. 2

[34] Weiwei Kong and Andres Munoz Medina. A unified fast gradient clipping framework for dp-sgd. *Advances in Neural Information Processing Systems*, 36:52401–52412, 2023. 3

[35] Jonathan Lebensold, Doina Precup, and Borja Balle. On the privacy of selection mechanisms with gaussian noise. In *International Conference on Artificial Intelligence and Statistics*, pages 1495–1503. PMLR, 2024. 3

[36] Jiefeng Li, Siyuan Bian, Ailing Zeng, Can Wang, Bo Pang, Wentao Liu, and Cewu Lu. Human pose regression with residual log-likelihood estimation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 11025–11034, 2021. 2

[37] Yizhuo Li, Miao Hao, Zonglin Di, Nitesh Bharadwaj Gundavarapu, and Xiaolong Wang. Test-time personalization with a transformer for human pose estimation. *Advances in Neural Information Processing Systems*, 34:2583–2597, 2021. 2

[38] Yanjie Li, Shoukui Zhang, Zhicheng Wang, Sen Yang, Wankou Yang, Shu-Tao Xia, and Erjin Zhou. Tokenpose: Learning keypoint tokens for human pose estimation. In *Proceedings of the IEEE/CVF International conference on computer vision*, pages 11313–11322, 2021. 2

[39] Yanjie Li, Sen Yang, Peidong Liu, Shoukui Zhang, Yunxiao Wang, Zhicheng Wang, Wankou Yang, and Shu-Tao Xia. Simcc: A simple coordinate classification perspective for human pose estimation. In *European Conference on Computer Vision*, pages 89–106. Springer, 2022. 1, 2, 3

[40] Zheng Li, Jingwen Ye, Mingli Song, Ying Huang, and Zhigeng Pan. Online knowledge distillation for efficient pose estimation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 11740–11750, 2021. 2

[41] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. 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. 6

[42] Peng Lu, Tao Jiang, Yining Li, Xiangtai Li, Kai Chen, and Wenming Yang. Rtmo: Towards high-performance one-stage real-time multi-person pose estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1491–1500, 2024. 1

[43] Kiwan Maeng, Chuan Guo, Sanjay Kariyappa, and G. Edward Suh. Bounding the invertibility of privacy-preserving instance encoding using fisher information. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023. 2

[44] Saeed Mahloujifar, Chuan Guo, G Edward Suh, and Kamalika Chaudhuri. Machine learning with privacy for protected attributes. In *2025 IEEE Symposium on Security and Privacy (SP)*, pages 2640–2657. IEEE, 2025. 3, 5

[45] Weian Mao, Yongtao Ge, Chunhua Shen, Zhi Tian, Xinlong Wang, Zhibin Wang, and Anton van den Hengel. Poseur: Direct human pose regression with transformers. In *European conference on computer vision*, pages 72–88. Springer, 2022. 1

[46] Nicole Martinez-Martin, Zelun Luo, Amit Kaushal, Ehsan Adeli, Albert Haque, Sara S Kelly, Sarah Wieten, Mildred K Cho, David Magnus, Li Fei-Fei, et al. Ethical issues in using ambient intelligence in health-care settings. *The lancet digital health*, 3(2):e115–e123, 2021. 1

[47] Xuecheng Nie, Jiashi Feng, Jianfeng Zhang, and Shuicheng Yan. Single-stage multi-person pose machines. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 6951–6960, 2019. 2

[48] Miroslav Purkrabek and Jiri Matas. Probpose: A probabilistic approach to 2d human pose estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025. 2

[49] Angel Ruiz-Zafra, Daniel Precioso, Blas Salvador, Simón P. Lubián-López, Javier Jiménez, Isabel Benavente-Fernández, Janet Pigueiras, David Gómez-Ullate, and Lionel C. Gontard. Neocam: An edge-cloud platform for non-invasive real-time monitoring in neonatal intensive care units. *IEEE Journal of Biomedical and Health Informatics*, 27(6):2614–2624, 2023. 1, 2

[50] Yinchen Shen, Zhiguo Wang, Ruoyu Sun, and Xiaojing Shen. Towards understanding the impact of model size on differential private classification. *arXiv preprint arXiv:2111.13895*, 2021. 8

[51] Weiyan Shi, Aiqi Cui, Evan Li, Ruoxi Jia, and Zhou Yu.

Selective differential privacy for language modeling. *arXiv preprint arXiv:2108.12944*, 2021. 3

[52] Reza Shokri, Marco Stronati, Congzheng Song, and Vitaly Shmatikov. Membership inference attacks against machine learning models. In *2017 IEEE symposium on security and privacy (SP)*, pages 3–18. IEEE, 2017. 3

[53] Hanling Tian, Yuhang Liu, Mingzhen He, Zhengbao He, Zhehao Huang, Ruikai Yang, and Xiaolin Huang. Simulating training dynamics to reconstruct training data from deep neural networks. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2025. 1, 2

[54] Alexander Toshev and Christian Szegedy. Deeppose: Human pose estimation via deep neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1653–1660, 2014. 2

[55] Bao Wang, Quanquan Gu, March Boedihardjo, Lingxiao Wang, Farzin Barekat, and Stanley J. Osher. DP-LSSGD: A stochastic optimization method to lift the utility in privacy-preserving ERM. In *Proceedings of The First Mathematical and Scientific Machine Learning Conference*, pages 328–351. PMLR, 2020. 3

[56] Chen Wang, Feng Zhang, Xiatian Zhu, and Shuzhi Sam Ge. Low-resolution human pose estimation. *Pattern Recognition*, 126:108579, 2022. 2

[57] Yihan Wang, Muyang Li, Han Cai, Wei-Ming Chen, and Song Han. Lite pose: Efficient architecture design for 2d human pose estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 13126–13136, 2022. 1

[58] Kan Wu, Jinnian Zhang, Houwen Peng, Mengchen Liu, Bin Xiao, Jianlong Fu, and Lu Yuan. Tinyvit: Fast pretraining distillation for small vision transformers. In *Computer Vision – ECCV 2022*, pages 68–85, Cham, 2022. Springer Nature Switzerland. 3

[59] Bin Xiao, Haiping Wu, and Yichen Wei. Simple baselines for human pose estimation and tracking. In *Proceedings of the European conference on computer vision (ECCV)*, pages 466–481, 2018. 2

[60] Yufei Xu, Jing Zhang, Qiming Zhang, and Dacheng Tao. Vitpose: Simple vision transformer baselines for human pose estimation. *Advances in Neural Information Processing Systems*, 35:38571–38584, 2022. 2

[61] Zheng Xu, Maxwell D. Collins, Yuxiao Wang, Liviu Panait, Sewoong Oh, Sean Augenstein, Ting Liu, Florian Schroff, and H. Brendan McMahan. Learning to generate image embeddings with user-level differential privacy. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023. 2

[62] Suhang Ye, Yingyi Zhang, Jie Hu, Liujuan Cao, Shengchuan Zhang, Lei Shen, Jun Wang, Shouhong Ding, and Rongrong Ji. Distilpose: Tokenized pose regression with heatmap distillation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2163–2172, 2023. 2

[63] Da Yu, Saurabh Naik, Arturs Backurs, Sivakanth Gopi, Huseyin A Inan, Gautam Kamath, Janardhan Kulkarni, Yin Tat Lee, Andre Manoel, Lukas Wutschitz, et al. Differentially private fine-tuning of language models. *arXiv preprint arXiv:2110.06500*, 2021. 7

[64] Lei Yu, Ling Liu, Calton Pu, Mehmet Emre Gursoy, and Stacey Truex. Differentially private model publishing for deep learning. In *2019 IEEE symposium on security and privacy (SP)*, pages 332–349. IEEE, 2019. 2

[65] Idris Zakariyya, Linda Tran, Kaushik Bhargav Sivangi, Paul Henderson, and Fani Deligianni. Differentially private integrated decision gradients (idg-dp) for radar-based human activity recognition. In *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2025. 1, 2

[66] Shihao Zhang, Baohua Qiang, Xianyi Yang, Xuekai Wei, Ruidong Chen, and Lirui Chen. Human pose estimation via an ultra-lightweight pose distillation network. *Electronics*, 12(12):2593, 2023. 2

[67] Ce Zheng, Wenhan Wu, Chen Chen, Taojiannan Yang, Sijie Zhu, Ju Shen, Nasser Kehtarnavaz, and Mubarak Shah. Deep learning-based human pose estimation: A survey. *ACM Computing Surveys*, 56(1):1–37, 2023. 1

[68] Yingxue Zhou, Zhiwei Steven Wu, and Arindam Banerjee. Bypassing the ambient dimension: Private sgd with gradient subspace identification. *arXiv preprint arXiv:2007.03813*, 2020. 4, 5, 8

[69] Yingxue Zhou, Steven Wu, and Arindam Banerjee. Bypassing the ambient dimension: Private {sgd} with gradient subspace identification. In *International Conference on Learning Representations*, 2021. 1, 3, 5

[70] Ligeng Zhu, Zhijian Liu, and Song Han. Deep leakage from gradients. *Advances in neural information processing systems*, 32, 2019. 2, 3