Differentially Private 2D Human Pose Estimation

Kaushik Bhargav Sivangi, Paul Henderson, Fani Deligianni

University of Glasgow
Kaushik.Sivangi@glasgow.ac.uk, Paul.Henderson@glasgow.ac.uk, Fani.Deligianni@glasgow.ac.uk

Abstract

Human pose estimation (HPE) has become essential in numerous applications including healthcare, activity recognition, and human-computer interaction. However, the privacy implications of processing sensitive visual data present significant deployment barriers in critical domains. While traditional anonymization techniques offer limited protection and often compromise data utility for broader motion analysis, Differential Privacy (DP) provides formal privacy guarantees but typically degrades model performance when applied naively. In this work, we present the first comprehensive framework for differentially private 2D human pose estimation (2D-HPE) by applying Differentially Private Stochastic Gradient Descent (DP-SGD) to this task. To effectively balance privacy with performance, we adopt Projected DP-SGD (PDP-SGD), which projects the noisy gradients to a low-dimensional subspace. Next, we incorporate Feature Differential Privacy(FDP) to selectively privatize only sensitive features while retaining public visual cues. Finally, we propose a hybrid feature-projective DP framework that combines both approaches to balance privacy and accuracy for HPE. We evaluate our approach on the MPII dataset across varying privacy budgets, training strategies, and clipping norms. Our combined feature-projective method consistently outperforms vanilla DP-SGD and individual baselines, achieving up to 82.61% mean PCKh@0.5 at $\epsilon = 0.8$, substantially closing the gap to the non-private performance. This work lays foundation for privacy-preserving human pose estimation in real-world, sensitive applications.

Introduction

Human pose estimation (HPE) represents a fundamental task that transforms raw visual content consisting of humans into structured representations of human positioning and movements. This sophisticated conversion of visual data into anatomically meaningful keypoint configurations enables numerous high impact applications in healthcare, activity recognition, human-computer interaction, sports analysis, and video games (Artacho and Savakis 2020; Mao et al. 2022; Wang et al. 2022b; Lu et al. 2024; Li et al. 2022).

As these human motion analysis systems are increasingly deployed into sensitive environments such as hospitals and homes, protecting user privacy has become a critical concern (Martinez-Martin et al. 2021). HPE systems that capture and process raw images pose significant privacy risks in two key ways: first, the raw images themselves contain sensitive personal information that could be exposed during collection and processing (Zheng et al. 2023); second, even trained models can inadvertently reveal information about individuals in training data when subjected to sophisticated privacy attacks such as model inversion or membership inference techniques (Geiping et al. 2020; Jegorova et al. 2023; Zakariyya et al. 2025). For instance, an adversary can exploit a model's weights (Haim et al. 2022) or gradients (Hatamizadeh et al. 2023) to reconstruct distinctive physical characteristics of patients or sensitive contextual information, such as the patient's home environment from the private training dataset (Jegorova et al. 2023). 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 approaches to privacy preservation in HPE have focused primarily on data anonymization techniques, such as blurring, pixelation, and template-based shape modeling (Ahmad, Morerio, and Del Bue 2024; Ruiz-Zafra et al. 2023; Hesse et al. 2018). 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 (Barattin et al. 2023). Furthermore, these methods typically lack theoretical privacy guarantees and remain vulnerable to more sophisticated attacks, limiting their applicability in highly sensitive contexts (Zakariyya et al. 2025). Moreover, these approaches do not address the inherent vulnerability of neural networks to memorization attacks that can reconstruct training data (Haim et al. 2022). This limitation significantly limits the potential to share models trained in sensitive multimedia datasets in the wider research community and clinical applications. The inherent tension between improving model utility and ensuring robust privacy preservation represents a challenging research problem (Abbasi, Mori, and Saracino 2025; Fang et al. 2023; Zhou, Wu, and Banerjee 2021) that has not been adequately explored in the con-

text of HPE. Differential privacy (DP) offers a promising framework to address these concerns by providing mathematical guarantees against data leakage (Hardt and Talwar 2010; Dwork 2021; Abadi et al. 2016; Dwork et al. 2006). However, implementing DP in deep learning models, particularly through Differentially Private Stochastic Gradient Descent (DP-SGD)(Abadi et al. 2016), typically results in substantial performance degradation that can render models impractical for real-world applications (Yu et al. 2019; De et al. 2022) . This degradation is problematic in vision tasks such as HPE, where the utility of the model relies on maintaining fine-grained spatial precision.

In this work, we present the first comprehensive study of differential privacy for the 2D-HPE task. 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. To address the privacy-utility trade-off in 2D-HPE, we explore multiple complementary strategies. First, we adopt a projection based DP-SGD approach that leverages the low-dimensional structure of the gradient space. By projecting the noisy gradients onto a principal subspace from a small public set, we mitigate privacy-induced noise without compromising the DP guarantees. Second, we incorporate Feature Differential Privacy (FDP), a privacy framework that relaxes differential privacy on public features, while rigorously protects sensitive features, thereby reducing the amount of calibrated noise required. Finally, we propose a hybrid strategy that integrates both projection and FDP effectively, combining their strengths.

To summarise, our core contributions are as follows:

- We present the first systematic exploration of differentially private learning for 2D human pose estimation, establishing baseline performance under various privacy regimes.
- We propose a projection-based DP-SGD mechanism adopted for pose estimation, which improves performance by restricting gradient updates to low-dimensional subspaces that capture the most informative directions.
- We adapt and extend the Feature Differential Privacy (FDP) framework to the HPE domain, allowing selective private updates only for sensitive features. This approach automatically protects both human subjects and their surrounding spatial context without requiring manual feature curation. Additionally, we integrate this with the projection method to yield a unified feature-projective DP approach demonstrating consistent best privacy-utility trade-off across various settings.
- We perform a thorough evaluation across multiple DP regimes, training strategies, and gradient clipping thresholds on the MPII dataset, demonstrating the effectiveness of the proposed method in improving the privacy-utility trade-off.

Related Work

2D Human Pose Estimation and Privacy

Markerless 2D-HPE identifies anatomical keypoints in images without physical markers, thus playing a fundamental role in human motion analysis for several applications in healthcare and activity recognition (Deligianni, Guo, and Yang 2019; Bondugula, Udgata, and Sivangi 2023). Although traditional heatmap methods originally based on CNN architectures (Artacho and Savakis 2020; Kamel et al. 2020; Wang et al. 2022a; Xiao, Wu, and Wei 2018) and later leveraging vision-based transformers (Li et al. 2021b,c; Xu et al. 2022) achieved state-of-the-art performance, they suffer from quantization errors and usually result in large cumbersome networks. Regression approaches offer faster, end-to-end solutions but with reduced accuracy (Toshev and Szegedy 2014; Nie et al. 2019; Li et al. 2021a). Coordinate classification approach addresses these limitations by treating pose estimation as a classification problem with discretized coordinates (Li et al. 2022). This method achieved state-of-the-art performance while maintaining computational efficiency. Knowledge distillation techniques (Ye et al. 2023; Li et al. 2021d; Zhang et al. 2023; Bhargav Sivangi 2024) have been also applied to enhance the efficiency of HPE with significantly less model parameters and inference speed by transfering knowledge from large heatmap based models to smaller architectures.

Protecting the privacy of users is of paramount importance in all real-world applications and a challenging task for human pose estimation, which depends on high-quality images. Recent research works(Haim et al. 2022; Zhu, Liu, and Han 2019) have demonstrated that adversaries can successfully reconstruct substantial portions of private training data solely by analyzing the parameters and gradients of a trained neural network. For these reasons, in critical applications, data sharing initiatives provide limited information, which prevents the analysis of crucial clinical indicators that require body shape or facial information to detect levels of stress and identify abnormal motion patterns. In practice, most platforms implement rudimentary privacy protection based on face de-identification with blurring and pixelation (Ruiz-Zafra et al. 2023). Some approaches to protecting people's privacy have evolved from basic techniques to more sophisticated methods, although significant challenges persist. Early work by (Hesse et al. 2018) employed skin removal and template-based shape modeling, which compromised clinical utility despite privacy benefits. (Hinojosa, Niebles, and Arguello 2021) proposed an end-to-end framework that integrates an optimized optical encoder with a CNN decoder, incorporating a visual privacy protection layer to degrade private attributes while preserving essential features. These methods lack theoretical privacy guarantees and remain vulnerable to adversarial attacks (Zakariyya et al. 2025). Furthermore,traditional methods like blurring or masking significantly degrade data quality, reducing their effectiveness in downstream tasks like instance segmentation and pose estimation. Recent image anonymization techniques (Hukkelås and Lindseth 2023a; Hukkelås et al. 2023) have been developed to address privacy concerns, employing realistic GAN-based approaches such as to generate anonymized human figures and faces without severe visual distortion. Although realistic anonymization improves data utility compared to traditional obfuscation, it still introduces artifacts or unnatural appearances due to challenges like limited pose detection accuracy and contextual mismatches.

Consequently, even advanced realistic anonymization techniques cannot yet fully replace real, non-anonymized data for training robust computer vision models (Hukkelås and Lindseth 2023b).

Utility vs Privacy with DP optimization

DP-SGD 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 (Hardt and Talwar 2010; Dwork 2021). In DP-SGD the process involves clipping the gradients first to bound the maximum norm of individual gradients and subsequently adding Gaussian noise. Clipping the gradients ensures that the gradients' sensitivity is bounded so that any single data point would have very limited influence (Lebensold, Precup, and Balle 2024). In this way, DP is maintained by preventing a single data point disproportionally influencing the deep learning model. However, there is a trade-off between privacy and utility in DP settings that can compromise the ability to develop privacy-preserved HPE algorithms that would be useful in practice (Abbasi, Mori, and Saracino 2025). The performance of DP-SGD has been shown to depend on the smoothness of the loss function (Wang et al. 2020) and the dimension of the gradient (Fang et al. 2023; Zhou, Wu, and Banerjee 2021). 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 (Fang et al. 2023). Recent research has sought to improve the utility-privacy trade-off by relaxing traditional differential privacy. (Shi et al. 2021) enhanced the utility of DP by introducing Selective Differential Privacy, which protects only sensitive tokens within language models, while leaving non-sensitive elements unperturbed. (Mahloujifar et al. 2025) 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.

Methodology

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 (Dwork 2021; Hardt and Talwar 2010). 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:

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

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 (Dwork 2021), effectively mitigating risks from various attacks, including membership inference(Shokri et al. 2017) or reconstruction attacks(Zhu, Liu, and Han 2019).

DP-SGD A common approach for ensuring differential privacy during neural network training is DP-SGD, which enforces an (ϵ, δ)-DP guarantee on gradient updates (Abadi et al. 2016; Dupuy et al. 2022; Boenisch et al. 2024). 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 (Kong and Munoz Medina 2023).

Lightweight Architecture for 2D-HPE

For our 2D-HPE models, we adopt TinyViT (Wu et al. 2022) as the backbone. It is a small sized four-stage efficient hierarchical vision transformer which is well suited for resource-constrained vision tasks. The model adopts a multi-stage architecture where in 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 (Howard et al. 2019) blocks from MobileNetV2 to efficiently learn the low-level representation. The last three stages consists 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 (Li et al. 2022). 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:

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

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.

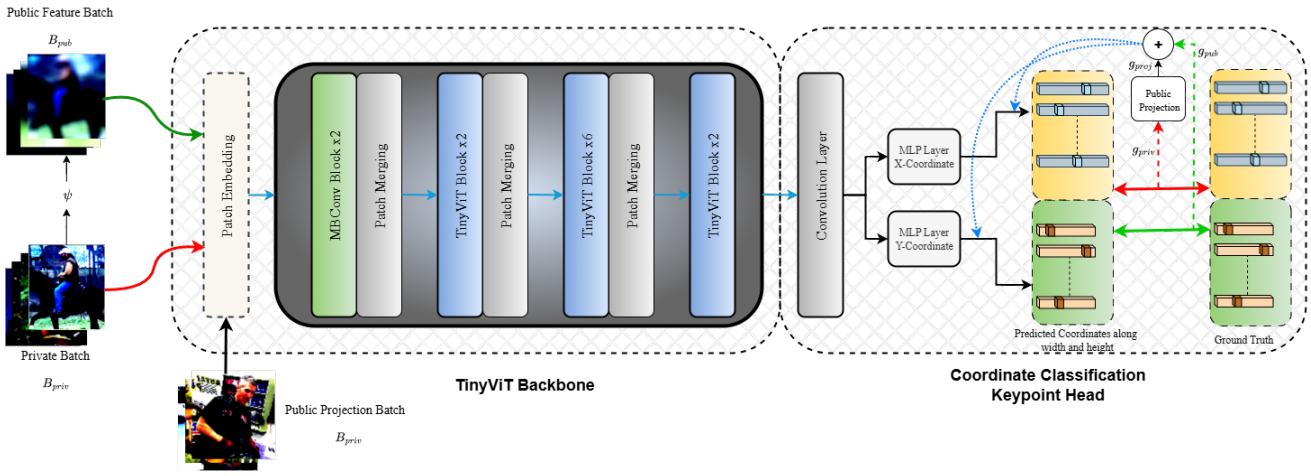


Figure 1: Overview of our private HPE pipeline coupling a TinyViT based backbone with Coordinate Classification Keypoint head. The Public feature batch B_{pub} is generated from the private batch B_{priv} using ψ both of which are given as input in a single iteration. Additionally, a public image set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

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

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 (Zhou, Wu, and Banerjee 2020). 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 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

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

where B is the size of the mini-batch and σ is the standard deviation. The full noisy gradient g is then projected on to 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.

Feature Projective DP-SGD for HPE

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 (Mahloujifar et al. 2025). 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 $f(x) = 1-\delta-e^{\epsilon}x$. 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:

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

Training: At every iteration, FDP-SGD updates parameters using two separate batches drawn independently from S_{data}. On the public batch B_{pub} we compute the gradient as:

$$g_{pub} = \frac{1}{|B_{pub}|} \sum_{x \in B_{pub}} \nabla l_{pub}(w, \psi(x)) \tag{8}$$

Similarly, on the private batch B_{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:

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

The parameters are updated at iteration t by combining both the public and private gradients as:

$$g = g_{priv} + g_{pub} \tag{10}$$

$$w_t = w_{t-1} - \eta_t g_t \tag{11}$$

where η_t denotes the learning rate. By applying noise only to the private gradient component, FDP-SGD significantly improves the accuracy of human pose estimation compared to vanilla DP-SGD under identical (ϵ, δ) privacy constraints.

To further enhance utility, we integrate FDP-SGD with our previously described subspace projection technique, creating Feature-Projective DP. This combined approach focuses parameter updates on the most informative gradient directions while maintaining the privacy benefits of feature decomposition. The complete algorithmic details of this integrated approach are provided in Algorithm 1.

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, iterationss 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 $V_t \in \mathbb{R}^{p \times k}$.
  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 $B_{pub}^t \subset \psi(x) : x \in S_{priv}$
  Compute

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

  Sample private batch $B_{priv}^t \subset S_{priv}$
  Compute the clipped gradient $\tilde{g}_t$, aggregate and add Gaussian noise

$$g_{priv}^t = \frac{1}{|B_{priv}^t|} \sum_{x \in B_{priv}^t} \tilde{g}_t + \mathcal{N}(0, \sigma^2 C^2 I)$$

  Project: $g_{proj}^t = (\hat{V}_t \hat{V}_t^{\top}) \cdot g_{priv}^t \in \mathbb{R}^p$.
  Merge Public and Private projected feature gradients

$$g_t = g_{pub}^t + g_{proj}^t$$

  Update: $w_t = w_{t-1} - \eta_t g_t$.
end for
return w_T.

Experiments

Dataset and Implementation Details

In our experiments, we evaluated our framework on two widely used human pose datasets: MS COCO Keypoint Dataset (Lin et al. 2014) and MPII dataset (Andriluka et al. 2014). Our methodology assumes that the COCO dataset serves as a public dataset used for pre-training the network weights, while MPII functions as a private dataset on which we apply the differential privacy techniques. The MS COCO Keypoint Dataset is employed for pretraining purposes.

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 tech-

niques 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) (Andriluka et al. 2014) as an evaluation metric. To further assess the generalization ability of our pose estimation models under domain shift and visual diversity, we conduct additional experiments on the Human-Art dataset (Ju et al. 2023). Human-Art is a recently introduced, large-scale human-centric dataset specifically designed to bridge the gap between natural and artificial visual domains. It contains 50,000 high-quality images with over 123,000 person instances drawn from 20 diverse scenarios, covering both natural scenes (e.g., cosplay, drama, dance) and a wide spectrum of artistic representations (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. This is primarily due to the presence of stylized or abstract depictions of human figures, exaggerated or distorted body proportions, occlusions, artistic textures, and unconventional pose. The results on HumanART are presented in supplementary.

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(De et al. 2022). To generate the public feature map, we employ Gaussian blur which effectively suppresses facial and body structure details. Details on datasets, training and privacy related parameters are provided in the supplementary.

Experimental Results and Analysis

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

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

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 ($\epsilon \in \{0.2, 0.4, 0.6, 0.8\}$) 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($\epsilon = 0.2$) 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 (Yu et al. 2021). 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.

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 $\epsilon = 0.2$ and from 78.17% to 80.63% at $\epsilon = 0.8$. This enhancement occurs because, while Gauissian 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 finetuning 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(Zhou, Wu, and Banerjee 2020). 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

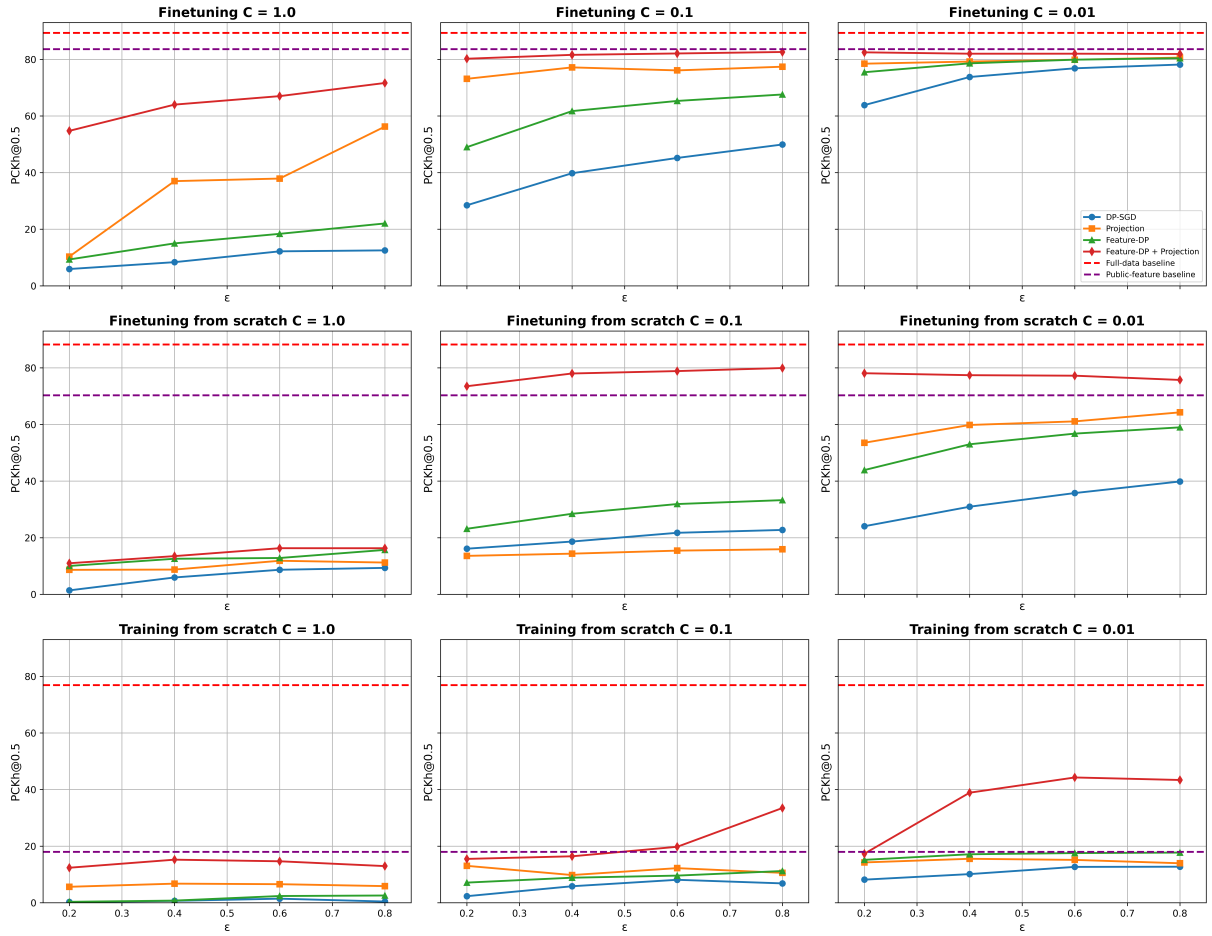


Figure 2: Comparison of PCKh@0.5 across private and non-private methodologies under different training strategies with varied privacy budget (ϵ) and clipping thresholds (C).

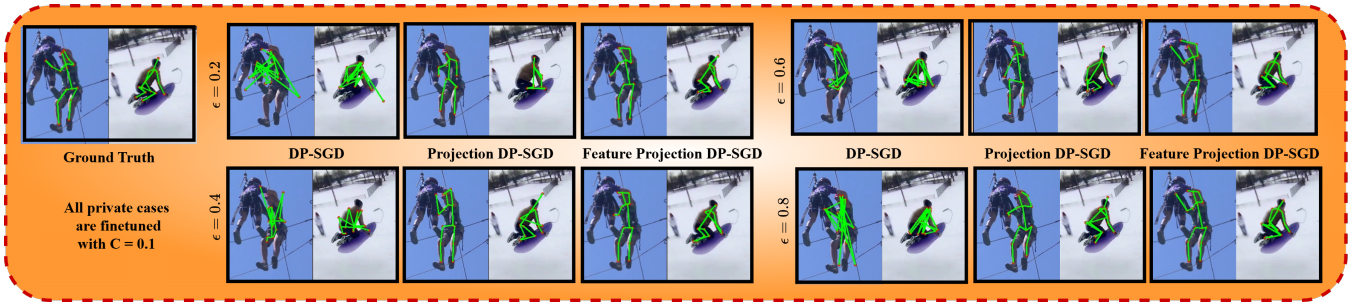


Figure 3: Depiction of qualitative results on DP-SGD, Projection DP-SGD and Feature Projection DP-SGD. We specifically show results on Finetuning with $C = 0.1$ at various privacy budgets.

dips in accuracy.

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 $\epsilon = 0.2$ (11.61% gain) and from 78.17% to 80.40% at $\epsilon = 0.8$ (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 $\epsilon = 0.8$, 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.

For completeness, detailed per-joint performance metrics, full qualitative results and additional experimental results are provided in the Supplementary Material.

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 82.61% mean PCKh@0.5 at $\epsilon = 0.8$, significantly narrowing the gap to non-private performance while exhibiting strong formal privacy guarantees. A key advantage of this method is that it requires no manual curation of private features, as we automatically protect the complete raw image, ensuring privacy preservation for both individuals and their personal spatial environments.

References

Abadi, M.; Chu, A.; Goodfellow, I.; McMahan, H. B.; Mironov, I.; Talwar, K.; and Zhang, L. 2016. Deep Learning with Differential Privacy. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, 308–318. ACM.

Abbasi, W.; Mori, P.; and Saracino, A. 2025. Trading-Off Privacy, Utility, and Explainability in Deep Learning-Based Image Data Analysis . *IEEE Transactions on Dependable and Secure Computing*, 22(01): 388–405.

Ahmad, S.; Morerio, P.; and Del Bue, A. 2024. Event anonymization: privacy-preserving person re-identification and pose estimation in event-based vision. *IEEE Access*.

Andriluka, M.; Pishchulin, L.; Gehler, P.; and Schiele, B. 2014. 2d human pose estimation: New benchmark and state of the art analysis. In *Proceedings of the IEEE Conference on computer Vision and Pattern Recognition*, 3686–3693.

Artacho, B.; and Savakis, A. 2020. Unipose: Unified human pose estimation in single images and videos. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 7035–7044.

Barattin, S.; Tzelepis, C.; Patras, I.; and Sebe, N. 2023. Attribute-preserving face dataset anonymization via latent code optimization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 8001–8010.

Bhargav Sivangi, F. D. 2024. Knowledge Distillation with Global Filters for Efficient Human Pose Estimation. In *British Machine Vision Conference(BMVC)*.

Boenisch, F.; Kowalczuk, A.; Dubinski, J.; Ghomi, A. A.; Sui, Y.; Stein, G.; Wu, J.; Cresswell, J. C.; and Dziedzic, A. 2024. Benchmarking Robust Self-Supervised Learning Across Diverse Downstream Tasks. *CoRR*, abs/2407.12588.

Bondugula, R. K.; Udgata, S. K.; and Sivangi, K. B. 2023. 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.

De, S.; Berrada, L.; Hayes, J.; Smith, S. L.; and Balle, B. 2022. Unlocking high-accuracy differentially private image classification through scale. *arXiv preprint arXiv:2204.13650*.

Deligianni, F.; Guo, Y.; and Yang, G.-Z. 2019. From emotions to mood disorders: A survey on gait analysis methodology. *IEEE journal of biomedical and health informatics*, 23(6): 2302–2316.

Dupuy, L.; Yovine, S.; Pan, F.; Basset, N.; and Dang, T. 2022. Towards Efficient Active Learning of PDFA. In *Proceedings of the 2022 International Conference on Learning Representations*. ICLR.

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

Dwork, C.; McSherry, F.; Nissim, K.; and Smith, A. 2006. Calibrating Noise to Sensitivity in Private Data Analysis. In Halevi, S.; and Rabin, T., eds., *Theory of Cryptography*, 265–284. Berlin, Heidelberg: Springer Berlin Heidelberg. ISBN 978-3-540-32732-5.

Fang, H.; Li, X.; Fan, C.; and Li, P. 2023. Improved Convergence of Differential Private SGD with Gradient Clipping. In *The Eleventh International Conference on Learning Representations*.

Geiping, J.; Bauermeister, H.; Dröge, H.; and Moeller, M. 2020. Inverting gradients-how easy is it to break privacy in federated learning? *Advances in neural information processing systems*, 33: 16937–16947.

Haim, N.; Vardi, G.; Yehudai, G.; michal Irani; and Shamir, O. 2022. Reconstructing Training Data From Trained Neural Networks. In Oh, A. H.; Agarwal, A.; Belgrave, D.; and Cho, K., eds., *Advances in Neural Information Processing Systems*.

Hardt, M.; and Talwar, K. 2010. On the geometry of differential privacy. In *Proceedings of the forty-second ACM symposium on Theory of computing*, 705–714.

Hatamizadeh, A.; Yin, H.; Molchanov, P.; Myronenko, A.; Li, W.; Dogra, P.; Feng, A.; Flores, M. G.; Kautz, J.; Xu, D.; et al. 2023. Do gradient inversion attacks make federated learning unsafe? *IEEE Transactions on Medical Imaging*, 42(7): 2044–2056.

Hesse, N.; Bodensteiner, C.; Arens, M.; Hofmann, U. G.; Weinberger, R.; and Sebastian Schroeder, A. 2018. 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*.

Hinojosa, C.; Niebles, J. C.; and Arguello, H. 2021. Learning Privacy-preserving Optics for Human Pose Estimation. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2553–2562.

Howard, A.; Sandler, M.; Chu, G.; Chen, L.-C.; Chen, B.; Tan, M.; Wang, W.; Zhu, Y.; Pang, R.; Vasudevan, V.; Le, Q. V.; and Adam, H. 2019. Searching for MobileNetV3. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*.

Hukkelås, H.; and Lindseth, F. 2023a. Deepprivacy2: Towards realistic full-body anonymization. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, 1329–1338.

Hukkelås, H.; and Lindseth, F. 2023b. Does image anonymization impact computer vision training? In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 140–150.

Hukkelås, H.; Smebye, M.; Mester, R.; and Lindseth, F. 2023. Realistic full-body anonymization with surface-guided GANs. In *Proceedings of the IEEE/CVF Winter conference on Applications of Computer Vision*, 1430–1440.

Jegorova, M.; Kaul, C.; Mayor, C.; O'Neil, A. Q.; Weir, A.; Murray-Smith, R.; and Tsaftaris, S. A. 2023. Survey: Leakage and Privacy at Inference Time. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(7): 9090–9108.

Ju, X.; Zeng, A.; Wang, J.; Xu, Q.; and Zhang, L. 2023. 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*, 618–629.

Kamel, A.; Sheng, B.; Li, P.; Kim, J.; and Feng, D. D. 2020. Hybrid refinement-correction heatmaps for human pose estimation. *IEEE Transactions on Multimedia*, 23: 1330–1342.

Kong, W.; and Munoz Medina, A. 2023. A unified fast gradient clipping framework for dp-sgd. *Advances in Neural Information Processing Systems*, 36: 52401–52412.

Lebensold, J.; Precup, D.; and Balle, B. 2024. On the privacy of selection mechanisms with gaussian noise. In *International Conference on Artificial Intelligence and Statistics*, 1495–1503. PMLR.

Li, J.; Bian, S.; Zeng, A.; Wang, C.; Pang, B.; Liu, W.; and Lu, C. 2021a. Human pose regression with residual log-likelihood estimation. In *Proceedings of the IEEE/CVF international conference on computer vision*, 11025–11034.

Li, Y.; Hao, M.; Di, Z.; Gundavarapu, N. B.; and Wang, X. 2021b. Test-time personalization with a transformer for human pose estimation. *Advances in Neural Information Processing Systems*, 34: 2583–2597.

Li, Y.; Yang, S.; Liu, P.; Zhang, S.; Wang, Y.; Wang, Z.; Yang, W.; and Xia, S.-T. 2022. Simcc: A simple coordinate classification perspective for human pose estimation. In *European Conference on Computer Vision*, 89–106. Springer.

Li, Y.; Zhang, S.; Wang, Z.; Yang, S.; Yang, W.; Xia, S.-T.; and Zhou, E. 2021c. Tokenpose: Learning keypoint tokens for human pose estimation. In *Proceedings of the IEEE/CVF International conference on computer vision*, 11313–11322.

Li, Z.; Ye, J.; Song, M.; Huang, Y.; and Pan, Z. 2021d. Online knowledge distillation for efficient pose estimation. In *Proceedings of the IEEE/CVF international conference on computer vision*, 11740–11750.

Lin, T.-Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; and Zitnick, C. L. 2014. Microsoft coco: Common objects in context. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*, 740–755. Springer.

Lu, P.; Jiang, T.; Li, Y.; Li, X.; Chen, K.; and Yang, W. 2024. 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)*, 1491–1500.

Mahloujifar, S.; Guo, C.; Suh, G. E.; and Chaudhuri, K. 2025. Machine Learning with Privacy for Protected Attributes. In *2025 IEEE Symposium on Security and Privacy (SP)*, 2640–2657. IEEE.

Mao, W.; Ge, Y.; Shen, C.; Tian, Z.; Wang, X.; Wang, Z.; and den Hengel, A. v. 2022. Poseur: Direct human pose regression with transformers. In *European conference on computer vision*, 72–88. Springer.

Martinez-Martin, N.; Luo, Z.; Kaushal, A.; Adeli, E.; Haque, A.; Kelly, S. S.; Wieten, S.; Cho, M. K.; Magnus, D.; Fei-Fei, L.; et al. 2021. Ethical issues in using ambient intelligence in health-care settings. *The lancet digital health*, 3(2): e115–e123.

Mironov, I. 2017. Rényi differential privacy. In *2017 IEEE 30th computer security foundations symposium (CSF)*, 263–275. IEEE.

Nie, X.; Feng, J.; Zhang, J.; and Yan, S. 2019. Single-stage multi-person pose machines. In *Proceedings of the IEEE/CVF international conference on computer vision*, 6951–6960.

Ruiz-Zafra, A.; Precioso, D.; Salvador, B.; Lubián-López, S. P.; Jiménez, J.; Benavente-Fernández, I.; Pigueiras, J.; Gómez-Ullate, D.; and Gontard, L. C. 2023. 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.

Shi, W.; Cui, A.; Li, E.; Jia, R.; and Yu, Z. 2021. Selective differential privacy for language modeling. *arXiv preprint arXiv:2108.12944*.

Shokri, R.; Stronati, M.; Song, C.; and Shmatikov, V. 2017. Membership inference attacks against machine learning models. In *2017 IEEE symposium on security and privacy (SP)*, 3–18. IEEE.

Toshev, A.; and Szegedy, C. 2014. Deeppose: Human pose estimation via deep neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1653–1660.

Wang, B.; Gu, Q.; Boedihardjo, M.; Wang, L.; Barekat, F.; and Osher, S. J. 2020. DP-LSSGD: A Stochastic Optimization Method to Lift the Utility in Privacy-Preserving ERM.

In Lu, J.; and Ward, R., eds., *Proceedings of The First Mathematical and Scientific Machine Learning Conference*, volume 107 of *Proceedings of Machine Learning Research*, 328–351. PMLR.

Wang, C.; Zhang, F.; Zhu, X.; and Ge, S. S. 2022a. Low-resolution human pose estimation. *Pattern Recognition*, 126: 108579.

Wang, Y.; Li, M.; Cai, H.; Chen, W.-M.; and Han, S. 2022b. Lite Pose: Efficient Architecture Design for 2D Human Pose Estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 13126–13136.

Wu, K.; Zhang, J.; Peng, H.; Liu, M.; Xiao, B.; Fu, J.; and Yuan, L. 2022. TinyViT: Fast Pretraining Distillation for Small Vision Transformers. In Avidan, S.; Brostow, G.; Cissé, M.; Farinella, G. M.; and Hassner, T., eds., *Computer Vision – ECCV 2022*, 68–85. Cham: Springer Nature Switzerland. ISBN 978-3-031-19803-8.

Xiao, B.; Wu, H.; and Wei, Y. 2018. Simple baselines for human pose estimation and tracking. In *Proceedings of the European conference on computer vision (ECCV)*, 466–481.

Xu, Y.; Zhang, J.; Zhang, Q.; and Tao, D. 2022. Vitpose: Simple vision transformer baselines for human pose estimation. *Advances in Neural Information Processing Systems*, 35: 38571–38584.

Ye, S.; Zhang, Y.; Hu, J.; Cao, L.; Zhang, S.; Shen, L.; Wang, J.; Ding, S.; and Ji, R. 2023. Distilpose: Tokenized pose regression with heatmap distillation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2163–2172.

Yu, D.; Naik, S.; Backurs, A.; Gopi, S.; Inan, H. A.; Kamath, G.; Kulkarni, J.; Lee, Y. T.; Manoel, A.; Wutschitz, L.; et al. 2021. Differentially private fine-tuning of language models. *arXiv preprint arXiv:2110.06500*.

Yu, L.; Liu, L.; Pu, C.; Gursoy, M. E.; and Truex, S. 2019. Differentially private model publishing for deep learning. In *2019 IEEE symposium on security and privacy (SP)*, 332–349. IEEE.

Zakariyya, I.; Tran, L.; Sivangi, K. B.; Henderson, P.; and Deligianni, F. 2025. Differentially Private Integrated Decision Gradients (IDG-DP) for Radar-based Human Activity Recognition. In *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*.

Zhang, S.; Qiang, B.; Yang, X.; Wei, X.; Chen, R.; and Chen, L. 2023. Human Pose Estimation via an Ultra-Lightweight Pose Distillation Network. *Electronics*, 12(12): 2593.

Zheng, C.; Wu, W.; Chen, C.; Yang, T.; Zhu, S.; Shen, J.; Kehtarnavaz, N.; and Shah, M. 2023. Deep learning-based human pose estimation: A survey. *ACM Computing Surveys*, 56(1): 1–37.

Zhou, Y.; Wu, S.; and Banerjee, A. 2021. Bypassing the Ambient Dimension: Private {SGD} with Gradient Subspace Identification. In *International Conference on Learning Representations*.

Zhou, Y.; Wu, Z. S.; and Banerjee, A. 2020. Bypassing the ambient dimension: Private sgd with gradient subspace identification. *arXiv preprint arXiv:2007.03813*.

Zhu, L.; Liu, Z.; and Han, S. 2019. Deep leakage from gradients. *Advances in neural information processing systems*, 32.

Implementation Details

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) (Andriluka et al. 2014) as an evaluation metric.

All models are trained under differential privacy constraints using DP-SGD with various clipping norms and privacy budgets (ϵ). Each model undergoes training for a total of 25 epochs, as we empirically observed no significant performance improvements when extending training beyond this duration under DP constraints. Throughout all experiments, we maintain a fixed input resolution of 256×192 pixels to ensure consistency across experiments and enable comparison with prior work.

For the privacy parameter settings, we use three gradient clipping norms $C = \{1.0, 0.1, 0.01\}$, with target privacy budgets of $\epsilon = \{0.2, 0.4, 0.6, 0.8\}$. We adopt Renyi Differential Privacy (RDP) (Mironov 2017) for privacy accounting with the privacy parameter $\delta = 4e - 5$.

For the projection method, we randomly select 100 samples from the training dataset of MPII as S_{pub} (ensuring no image overlap with the private data) with the remaining data forming the private training set S_{priv}. The default projection dimension K is set to 50 for all experiments.

To generate the public feature map, we employ Gaussian blur as ψ with a kernel size of $(25, 25)$ and standard deviation $\sigma_X = 10$, which effectively suppresses facial and body structure details. We deliberately blur the entire image rather than selectively masking human regions, as this approach provides comprehensive privacy protection by obscuring not only human identities but also contextual environmental details.

We chose to blur the complete image owing to the technicalities of the dataset as MPII has multiple instances of humans from similar image.

Per Joint PCKh@0.5 on MPII and HumanART

Tables below report per-joint PCKh@0.5 scores on MPII under various DP-SGD settings. Smaller clipping norms (C) and higher privacy budgets (ϵ) consistently improve pose accuracy. While we provide detailed results on the MPII dataset across various DP-SGD settings, including training from scratch, fine-tuning, and projection-based setups under different clipping norms and privacy budgets, we only report a subset of results for the Human-Art dataset. This is because certain settings, especially those involving large clipping norms or training without pretraining, result in extremely low utility. In many of these cases, the models fail to converge or produce meaningful predictions, with PCKh scores often dropping below 5%. Since these settings do not offer useful insights for practical analysis, we exclude them from the main tables and focus on configurations that yield more stable and interpretable performance

Table 2: MPII Results: DP-SGD.

Privacy Parameter(ϵ)	Head	Shoulder	Elbow	Wrist	Hip	Knee	Ankle	Mean	Mean@0.1
Finetuning									
C = 1.0									
$\epsilon = 0.2$	0.00	5.15	13.31	6.00	10.20	2.18	0.21	5.94	0.29
$\epsilon = 0.4$	0.99	6.27	14.83	7.68	12.97	2.38	3.26	8.36	0.34
$\epsilon = 0.6$	0.31	13.06	21.56	8.43	19.16	3.00	3.59	12.19	0.58
$\epsilon = 0.8$	2.97	10.43	19.16	7.06	22.31	3.75	8.12	12.53	0.61
C = 0.1									
$\epsilon = 0.2$	30.73	35.51	31.19	13.43	37.96	12.59	13.30	28.46	1.61
$\epsilon = 0.4$	43.01	49.92	47.98	20.20	50.67	18.22	21.56	39.78	2.43
$\epsilon = 0.6$	53.10	59.51	52.00	24.55	58.54	22.65	22.65	45.18	2.86
$\epsilon = 0.8$	64.56	64.44	53.09	28.29	60.05	30.29	28.74	49.93	3.42
C = 0.01									
$\epsilon = 0.2$	78.14	83.36	65.21	47.49	69.98	47.35	39.02	63.85	5.68
$\epsilon = 0.4$	83.83	88.88	77.02	63.83	74.87	62.50	52.12	73.77	9.30
$\epsilon = 0.6$	87.11	90.05	78.71	70.05	75.37	66.61	59.42	76.85	11.05
$\epsilon = 0.8$	87.79	90.32	78.95	71.99	77.91	69.07	61.55	78.17	11.91
Finetuning from scratch									
C = 1.0									
$\epsilon = 0.2$	4.09	2.96	0.82	1.51	0.71	0.75	0.59	1.39	0.06
$\epsilon = 0.4$	6.51	7.51	3.44	3.67	1.66	14.04	1.06	5.95	0.22
$\epsilon = 0.6$	16.17	11.79	7.41	3.92	9.23	9.27	1.82	8.67	0.37
$\epsilon = 0.8$	8.94	17.05	8.40	3.79	10.04	8.38	2.95	9.34	0.39
C = 0.1									
$\epsilon = 0.2$	12.48	20.87	15.70	10.91	23.39	13.96	8.36	16.11	0.68
$\epsilon = 0.4$	16.68	22.69	21.15	11.10	24.10	12.98	9.38	18.64	0.80
$\epsilon = 0.6$	23.26	28.07	21.70	13.06	26.92	14.57	9.73	21.73	1.07
$\epsilon = 0.8$	28.58	29.14	21.36	13.18	27.40	14.31	9.09	22.74	1.10
C = 0.01									
$\epsilon = 0.2$	28.89	31.52	22.43	12.11	27.37	16.72	10.39	24.05	1.16
$\epsilon = 0.4$	49.25	43.19	26.28	15.01	33.58	18.05	15.28	30.94	1.86
$\epsilon = 0.6$	60.20	50.68	30.99	16.99	37.84	20.15	20.34	35.77	2.25
$\epsilon = 0.8$	62.28	54.33	36.48	21.66	43.36	22.43	21.42	39.86	2.86
Training from scratch									
C = 1.0									
$\epsilon = 0.2$	0.14	0.05	0.37	0.02	0.64	0.26	0.35	0.30	0.02
$\epsilon = 0.4$	1.71	0.07	0.07	0.15	0.28	3.28	0.02	0.68	0.03
$\epsilon = 0.6$	0.03	4.64	0.00	1.25	0.90	3.10	0.05	1.45	0.07
$\epsilon = 0.8$	0.31	0.00	1.76	0.26	0.00	0.00	0.90	0.44	0.02
C = 0.1									
$\epsilon = 0.2$	1.09	0.03	2.97	7.01	1.28	1.23	0.50	2.33	0.09
$\epsilon = 0.4$	0.14	5.04	6.49	4.69	18.87	0.67	0.24	5.84	0.23
$\epsilon = 0.6$	0.10	9.51	8.01	9.58	22.90	2.08	1.23	8.12	0.33
$\epsilon = 0.8$	0.24	8.07	5.25	9.46	17.03	2.04	2.39	6.85	0.28
C = 0.01									
$\epsilon = 0.2$	0.31	10.31	10.07	4.90	13.33	3.41	1.77	8.17	0.36
$\epsilon = 0.4$	1.33	15.61	9.17	8.96	19.92	3.87	1.32	10.13	0.48
$\epsilon = 0.6$	6.79	17.24	13.50	9.60	21.53	5.74	2.13	12.68	0.56
$\epsilon = 0.8$	13.27	16.76	13.38	9.68	19.87	5.48	2.17	12.74	0.54

Table 3: MPII Results: DP-SGD with Projection.

Privacy Parameter(ϵ)	Head	Shoulder	Elbow	Wrist	Hip	Knee	Ankle	Mean	Mean@0.1
Finetuning									
C = 1.0									
$\epsilon = 0.2$	3.48	14.79	6.85	8.38	17.36	8.85	8.62	10.34	0.42
$\epsilon = 0.4$	54.67	55.45	32.54	23.70	37.23	21.44	13.30	36.99	2.33
$\epsilon = 0.6$	55.97	49.10	31.99	28.22	39.67	24.10	14.34	37.90	2.31
$\epsilon = 0.8$	71.18	76.19	55.51	46.60	53.97	39.13	32.97	56.26	4.19
C = 0.1									
$\epsilon = 0.2$	88.85	89.44	75.78	68.46	61.21	63.83	54.68	73.13	10.56
$\epsilon = 0.4$	88.44	89.84	78.92	72.62	70.21	69.45	61.08	77.17	12.55
$\epsilon = 0.6$	90.31	90.20	79.27	71.43	66.02	68.75	58.01	76.11	12.21
$\epsilon = 0.8$	91.51	90.39	79.51	72.84	71.23	67.86	59.78	77.41	12.88
C = 0.01									
$\epsilon = 0.2$	92.02	90.78	79.10	72.47	72.72	70.74	64.29	78.48	13.67
$\epsilon = 0.4$	91.81	90.74	79.92	72.04	75.42	71.79	65.78	79.23	13.49
$\epsilon = 0.6$	92.29	91.78	80.48	73.75	74.16	72.29	67.88	79.89	14.28
$\epsilon = 0.8$	92.29	91.49	80.86	74.52	75.32	73.91	69.77	80.63	14.61
Finetuning from scratch									
C = 1.0									
$\epsilon = 0.2$	0.48	8.93	8.86	8.24	20.72	3.77	1.75	8.64	0.31
$\epsilon = 0.4$	4.13	14.88	14.32	6.99	3.41	5.80	3.73	8.74	0.38
$\epsilon = 0.6$	3.48	16.34	9.90	12.35	13.21	13.48	10.23	11.85	0.56
$\epsilon = 0.8$	2.69	15.73	11.20	11.79	19.65	6.83	1.94	11.22	0.51
C = 0.1									
$\epsilon = 0.2$	4.40	18.99	16.99	9.42	18.07	10.70	6.78	13.58	0.63
$\epsilon = 0.4$	12.45	15.30	17.25	10.90	21.50	9.61	7.01	14.36	0.61
$\epsilon = 0.6$	15.59	15.64	16.70	10.50	19.49	14.19	7.98	15.43	0.66
$\epsilon = 0.8$	13.34	23.30	15.51	10.08	20.06	8.68	9.05	15.92	0.66
C = 0.01									
$\epsilon = 0.2$	82.44	69.58	49.75	43.23	43.31	39.11	36.56	53.54	5.87
$\epsilon = 0.4$	83.77	75.70	55.19	50.44	52.12	46.12	45.35	59.82	7.87
$\epsilon = 0.6$	86.02	74.25	61.31	52.96	51.39	46.75	45.42	61.09	8.61
$\epsilon = 0.8$	87.14	77.77	63.32	56.18	57.14	50.92	48.89	64.28	9.68
Training from scratch									
C = 1.0									
$\epsilon = 0.2$	0.07	1.77	10.07	11.07	12.74	1.37	0.02	5.65	0.26
$\epsilon = 0.4$	1.36	6.98	17.69	9.77	5.24	0.95	0.07	6.76	0.28
$\epsilon = 0.6$	0.07	1.00	14.47	12.44	6.42	7.19	1.87	6.57	0.31
$\epsilon = 0.8$	0.17	7.24	11.71	5.91	2.68	4.27	1.23	5.89	0.24
C = 0.1									
$\epsilon = 0.2$	1.19	18.05	12.78	9.53	22.66	7.64	5.95	13.05	0.56
$\epsilon = 0.4$	0.75	13.08	6.17	5.55	21.91	3.36	5.50	9.80	0.41
$\epsilon = 0.6$	1.98	21.28	8.16	11.05	19.70	4.11	4.72	12.24	0.56
$\epsilon = 0.8$	1.30	13.20	4.48	8.31	22.62	6.73	4.27	10.57	0.43
C = 0.01									
$\epsilon = 0.2$	5.15	15.10	15.07	12.39	19.61	10.28	8.83	14.26	0.65
$\epsilon = 0.4$	9.21	20.60	15.78	10.67	22.33	13.06	7.01	15.54	0.70
$\epsilon = 0.6$	17.09	19.70	6.89	10.50	22.36	15.88	10.51	15.15	0.62
$\epsilon = 0.8$	9.48	19.55	16.12	10.91	22.45	6.61	3.57	13.96	0.60

Table 4: MPII Results: Feature DP-SGD.

Privacy Parameter(ϵ)	Head	Shoulder	Elbow	Wrist	Hip	Knee	Ankle	Mean	Mean@0.1
Finetuning									
C = 0.01									
$\epsilon = 0.2$	87.04	89.28	75.78	66.54	75.89	65.04	58.48	75.47	10.64
$\epsilon = 0.4$	90.76	91.30	78.51	69.95	78.55	68.85	63.86	78.60	12.71
$\epsilon = 0.6$	91.47	91.95	79.63	71.99	79.87	71.00	65.45	79.90	13.62
$\epsilon = 0.8$	92.16	92.15	79.85	72.81	80.41	71.55	66.30	80.41	14.22
C = 0.1									
$\epsilon = 0.2$	52.42	65.78	57.32	27.43	53.31	36.05	26.03	48.99	3.53
$\epsilon = 0.4$	78.04	81.62	64.53	40.43	67.70	45.14	39.58	61.73	5.44
$\epsilon = 0.6$	78.48	84.80	68.48	45.60	69.88	49.89	45.11	65.32	6.47
$\epsilon = 0.8$	82.74	85.21	70.29	48.50	71.39	54.50	47.85	67.60	7.03
C = 1.0									
$\epsilon = 0.2$	1.60	9.90	7.86	11.57	14.25	3.00	7.91	9.32	0.33
$\epsilon = 0.4$	11.49	17.05	8.30	11.08	23.89	8.20	9.57	15.01	0.59
$\epsilon = 0.6$	14.84	25.00	12.36	13.62	27.07	10.84	11.29	18.37	0.81
$\epsilon = 0.8$	21.69	29.28	16.16	14.56	26.55	15.66	15.45	22.05	1.01
Finetuning from scratch									
C = 0.01									
$\epsilon = 0.2$	71.15	57.17	38.71	22.32	47.24	28.39	29.12	43.89	3.48
$\epsilon = 0.4$	78.75	69.40	47.15	33.34	54.86	38.87	37.72	53.00	5.10
$\epsilon = 0.6$	82.44	72.69	52.00	38.19	59.22	41.69	41.14	56.78	6.07
$\epsilon = 0.8$	83.80	74.81	54.88	41.29	60.93	44.35	42.80	58.98	6.70
C = 0.1									
$\epsilon = 0.2$	33.94	30.42	19.36	12.83	27.19	15.03	10.82	23.15	1.14
$\epsilon = 0.4$	40.86	37.75	22.57	15.08	31.71	18.09	15.94	28.46	1.53
$\epsilon = 0.6$	47.68	43.27	25.09	16.31	35.35	19.56	17.78	31.90	1.92
$\epsilon = 0.8$	52.25	45.31	27.27	16.57	35.90	19.28	18.47	33.27	2.21
C = 1.0									
$\epsilon = 0.2$	12.14	8.85	8.76	10.54	11.72	9.47	4.04	10.03	0.39
$\epsilon = 0.4$	10.06	13.33	10.74	9.92	20.84	12.17	5.48	12.57	0.51
$\epsilon = 0.6$	7.03	12.65	17.11	11.29	21.67	15.35	3.19	12.84	0.52
$\epsilon = 0.8$	13.47	19.74	15.95	9.58	22.95	12.65	8.90	15.69	0.60
Training from scratch									
C = 0.01									
$\epsilon = 0.2$	6.92	18.29	17.40	10.96	22.62	14.89	3.71	15.17	0.60
$\epsilon = 0.4$	10.57	22.61	18.41	12.15	23.09	14.33	6.83	17.14	0.71
$\epsilon = 0.6$	11.39	22.18	17.16	11.91	24.23	15.90	8.86	17.57	0.82
$\epsilon = 0.8$	14.09	21.31	19.52	11.31	24.32	16.44	7.35	17.74	0.80
C = 0.1									
$\epsilon = 0.2$	4.23	0.56	7.82	8.64	19.44	1.91	1.96	7.13	0.26
$\epsilon = 0.4$	0.48	12.62	10.69	8.12	21.15	2.36	0.33	8.87	0.37
$\epsilon = 0.6$	0.48	15.08	10.64	9.00	19.72	4.33	1.20	9.60	0.41
$\epsilon = 0.8$	0.61	13.09	12.94	10.74	22.73	4.94	1.37	11.22	0.48
C = 1.0									
$\epsilon = 0.2$	0.31	0.32	0.14	0.00	0.03	0.06	0.12	0.35	0.02
$\epsilon = 0.4$	1.84	3.63	0.02	0.43	0.00	0.00	0.09	0.78	0.02
$\epsilon = 0.6$	0.00	2.62	0.05	4.61	8.34	0.12	0.00	2.39	0.10
$\epsilon = 0.8$	0.14	0.02	0.07	0.00	15.22	0.02	0.26	2.57	0.12

Table 5: MPII Results: Feature Projection DP-SGD.

Privacy Parameter(ϵ)	Head	Shoulder	Elbow	Wrist	Hip	Knee	Ankle	Mean	Mean@0.1
Finetuning									
C = 0.01									
$\epsilon = 0.2$	94.17	92.70	82.17	73.29	80.79	76.45	72.60	82.50	19.65
$\epsilon = 0.4$	94.13	92.70	81.46	73.10	80.51	75.44	71.00	82.01	19.23
$\epsilon = 0.6$	94.10	92.70	81.47	72.79	80.87	75.32	71.19	82.01	19.48
$\epsilon = 0.8$	94.27	92.63	81.58	72.50	80.37	75.56	70.88	81.91	19.22
C = 0.1									
$\epsilon = 0.2$	93.79	91.85	79.07	71.56	77.43	73.32	68.92	80.24	15.20
$\epsilon = 0.4$	94.03	92.56	80.72	73.19	79.85	74.95	69.96	81.60	17.03
$\epsilon = 0.6$	93.86	92.82	81.17	74.70	79.78	75.68	70.78	82.09	17.74
$\epsilon = 0.8$	94.78	93.21	82.31	74.51	80.63	76.04	71.30	82.62	18.64
C = 1.0									
$\epsilon = 0.2$	73.36	65.88	53.72	40.79	56.14	43.44	34.10	54.75	4.05
$\epsilon = 0.4$	86.53	82.17	63.52	54.50	57.68	49.20	43.01	64.02	6.09
$\epsilon = 0.6$	86.66	86.06	66.70	51.94	67.47	51.40	47.02	67.02	6.73
$\epsilon = 0.8$	91.13	89.06	73.91	63.39	62.92	59.60	51.75	71.66	9.37
Finetuning from scratch									
C = 0.01									
$\epsilon = 0.2$	92.91	88.94	76.55	67.33	77.81	71.11	65.99	78.11	16.82
$\epsilon = 0.4$	92.74	88.55	75.80	65.63	77.48	70.02	65.37	77.41	16.23
$\epsilon = 0.6$	93.21	88.62	75.61	65.15	77.79	69.33	64.52	77.23	16.37
$\epsilon = 0.8$	92.29	87.11	74.11	63.13	76.53	67.46	62.78	75.74	15.83
C = 0.1									
$\epsilon = 0.2$	91.95	85.61	71.11	59.78	73.84	64.58	61.36	73.51	14.33
$\epsilon = 0.4$	93.76	88.65	76.97	67.07	77.27	70.30	66.23	78.01	16.93
$\epsilon = 0.6$	93.49	89.08	77.09	67.81	79.09	72.23	67.48	78.86	16.93
$\epsilon = 0.8$	94.03	90.46	78.56	68.58	80.68	72.54	69.08	79.95	17.83
C = 1.0									
$\epsilon = 0.2$	4.20	8.02	13.07	9.20	22.99	7.64	7.02	10.98	0.46
$\epsilon = 0.4$	10.54	17.53	13.19	11.50	20.60	7.41	5.12	13.49	0.53
$\epsilon = 0.6$	15.52	20.67	14.91	12.11	21.64	14.99	8.10	16.29	0.78
$\epsilon = 0.8$	20.36	20.92	12.63	10.95	24.98	10.07	8.46	16.29	0.73
Training from scratch									
C = 0.01									
$\epsilon = 0.2$	16.75	19.53	17.62	12.41	24.84	14.35	9.99	17.34	0.71
$\epsilon = 0.4$	62.65	52.11	34.07	19.03	41.09	27.30	23.74	38.90	3.43
$\epsilon = 0.6$	71.66	60.36	38.86	22.51	46.65	31.51	26.26	44.28	4.39
$\epsilon = 0.8$	67.84	58.95	38.06	21.86	44.95	31.99	27.66	43.39	4.12
C = 0.1									
$\epsilon = 0.2$	14.02	17.56	14.69	10.86	20.65	15.84	7.96	15.50	0.72
$\epsilon = 0.4$	11.02	19.51	16.62	11.02	23.61	14.16	10.16	16.42	0.65
$\epsilon = 0.6$	18.21	24.56	19.24	12.34	26.62	15.68	12.49	19.78	0.93
$\epsilon = 0.8$	53.27	46.01	29.37	17.27	35.47	21.20	17.17	33.49	2.33
C = 1.0									
$\epsilon = 0.2$	1.13	14.74	9.56	8.19	22.14	14.47	3.54	12.39	0.54
$\epsilon = 0.4$	17.77	14.81	15.85	9.46	22.16	14.83	4.06	15.23	0.70
$\epsilon = 0.6$	18.49	21.08	15.39	11.12	22.54	5.14	5.48	14.67	0.66
$\epsilon = 0.8$	5.83	14.96	14.88	10.66	22.21	13.04	3.83	12.97	0.55

Table 6: HumanART Results: DP-SGD.

Privacy Parameter(ϵ)	Head	Shoulder	Elbow	Wrist	Hip	Knee	Ankle	Mean	Mean@0.1	inkoti
Finetuning										
C = 0.01										
$\epsilon = 0.2$	29.1	67.6	20.0	14.0	30.9	34.0	71.6	28.6	22.0	35.5
$\epsilon = 0.4$	40.7	74.7	39.3	23.8	42.7	45.8	77.1	46.7	32.9	47.9
$\epsilon = 0.6$	37.5	74.7	33.6	20.9	39.4	42.1	77.6	41.4	29.1	43.8
$\epsilon = 0.8$	39.0	75.9	36.0	22.2	40.9	43.5	78.3	43.6	30.1	45.3
C = 0.1										
$\epsilon = 0.2$	3.5	19.0	0.0	0.9	4.0	8.7	35.6	0.8	5.0	9.2
$\epsilon = 0.4$	7.7	32.4	0.5	2.3	8.5	14.6	47.7	4.0	8.7	15.3
$\epsilon = 0.6$	10.4	39.3	1.5	3.4	11.3	17.1	51.9	6.2	10.5	17.9
$\epsilon = 0.8$	12.0	43.3	2.4	3.9	13.0	18.4	54.1	7.3	11.4	19.2
Finetuning from scratch										
C = 0.01										
$\epsilon = 0.2$	0.9	5.8	0.0	0.2	1.0	3.8	19.7	0.1	2.6	3.9
$\epsilon = 0.4$	1.3	8.0	0.0	0.4	1.4	5.0	23.7	0.3	3.5	5.2
$\epsilon = 0.6$	1.7	10.7	0.0	0.5	1.9	6.1	27.5	0.6	4.2	6.4
$\epsilon = 0.8$	2.1	12.9	0.0	0.8	2.3	6.8	29.6	0.7	5.1	7.0
C = 0.1										
$\epsilon = 0.2$	0.0	0.4	0.0	0.0	0.1	0.8	5.7	0.0	0.5	0.8
$\epsilon = 0.4$	0.2	1.9	0.0	0.1	0.3	2.3	13.4	0.1	1.4	2.4
$\epsilon = 0.6$	0.5	3.6	0.0	0.1	0.6	3.3	18.6	0.0	2.2	3.4
$\epsilon = 0.8$	0.6	4.3	0.0	0.1	0.7	3.5	20.0	0.0	2.4	3.7

Table 7: HumanART Results: DP-SGD with projection.

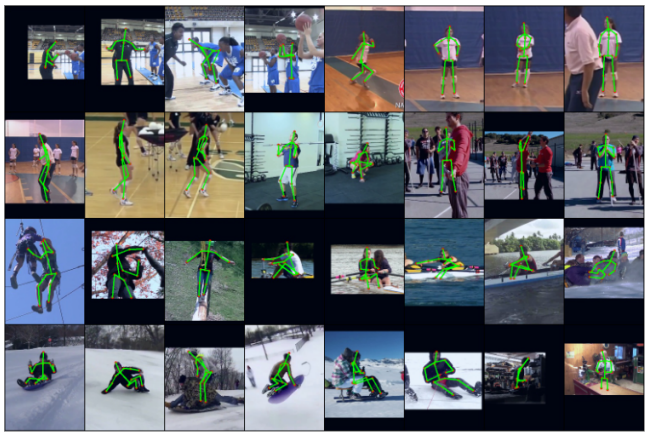
Privacy Parameter(ϵ)	Head	Shoulder	Elbow	Wrist	Hip	Knee	Ankle	Mean	Mean@0.1	inkoti
Finetuning										
C = 0.01										
$\epsilon = 0.2$	38.3	73.9	35.6	21.7	40.4	43.6	76.9	43.6	30.8	45.8
$\epsilon = 0.4$	34.7	73.4	29.1	18.3	36.6	39.7	76.0	37.9	26.9	41.4
$\epsilon = 0.6$	40.3	75.8	38.4	22.5	42.6	46.1	78.5	47.0	33.1	47.8
$\epsilon = 0.8$	38.7	74.3	36.3	21.1	40.8	44.3	77.3	44.5	32.1	45.9
C = 0.1										
$\epsilon = 0.2$	31.6	69.1	24.7	17.0	33.3	36.8	72.7	33.1	26.0	38.2
$\epsilon = 0.4$	32.9	70.4	27.0	18.0	34.6	38.1	73.2	35.2	27.2	39.5
$\epsilon = 0.6$	33.6	71.3	26.9	18.4	35.4	38.6	74.3	35.2	28.2	40.0
$\epsilon = 0.8$	33.5	69.4	28.4	18.4	35.3	38.7	72.6	36.8	27.1	40.2
C = 1.0										
$\epsilon = 0.2$	1.1	6.7	0.0	0.3	1.3	4.8	21.9	0.2	3.2	5.0
$\epsilon = 0.4$	11.7	41.7	2.0	3.8	12.8	18.1	52.4	7.8	11.1	19.0
$\epsilon = 0.6$	10.7	39.4	1.9	2.9	11.7	17.1	51.2	6.8	9.7	18.0
$\epsilon = 0.8$	23.5	60.4	14.2	11.1	25.2	30.2	66.2	24.2	20.2	31.5
Finetuning from scratch										
C = 0.01										
$\epsilon = 0.2$	1.7	8.5	0.4	0.4	1.9	4.9	19.5	1.1	2.9	5.2
$\epsilon = 0.4$	3.5	15.2	0.3	1.1	3.9	8.4	29.0	2.6	5.8	8.7
$\epsilon = 0.6$	2.1	10.2	0.1	0.7	2.4	5.9	22.2	1.4	3.4	6.2
$\epsilon = 0.8$	1.4	7.3	0.1	0.5	1.6	4.3	18.2	0.6	2.4	4.5
C = 0.1										
$\epsilon = 0.2$	0.5	2.8	0.1	0.2	0.6	2.2	10.9	0.1	1.5	2.3
$\epsilon = 0.4$	0.6	4.0	0.0	0.3	0.7	2.5	12.9	0.1	1.7	2.6
$\epsilon = 0.6$	1.2	5.7	0.5	0.4	1.5	3.4	15.3	0.6	1.7	3.7
$\epsilon = 0.8$	1.1	4.2	1.0	0.3	1.2	3.1	14.0	0.5	1.8	3.2
C = 1.0										
$\epsilon = 0.2$	0.1	0.9	0.0	0.0	0.2	0.9	5.6	0.0	0.6	0.9
$\epsilon = 0.4$	0.3	2.0	0.1	0.0	0.4	1.9	11.1	0.1	1.0	2.0
$\epsilon = 0.6$	0.8	4.6	0.0	0.1	0.9	3.1	16.1	0.1	1.7	3.3
$\epsilon = 0.8$	0.2	1.3	0.1	0.0	0.3	1.3	7.5	0.1	0.9	1.4
Training from scratch										
C = 0.01										
$\epsilon = 0.2$	0.12	0.68	0.06	0.0	0.17	1.06	5.7	0.1	0.5	1.1
$\epsilon = 0.4$	0.0	0.09	0.0	0.0	0.0	0.09	0.71	0.0	0.10	0.09
$\epsilon = 0.6$	0.31	2.1	0.0	0.07	0.4	1.8	11.7	0.01	1.18	1.93
$\epsilon = 0.8$	0.11	0.82	0.0	0.04	0.2	1.2	7.0	0.02	1.15	1.20
C = 0.1										
$\epsilon = 0.2$	0.05	0.29	0.0	0.06	0.38	2.5	0.0	0.3	0.4	1.24
$\epsilon = 0.4$	0.12	0.91	0.0	0.05	0.16	1.35	7.58	0.04	1.3	1.37
$\epsilon = 0.6$	0.19	1.27	0.0	0.04	0.23	1.6	9.06	0.03	1.02	1.6
$\epsilon = 0.8$	0.02	0.11	0.0	0.0	0.03	2.09	0.0	0.25	0.32	0.31

Table 8: HumanART Results: Feature Projection DP-SGD.

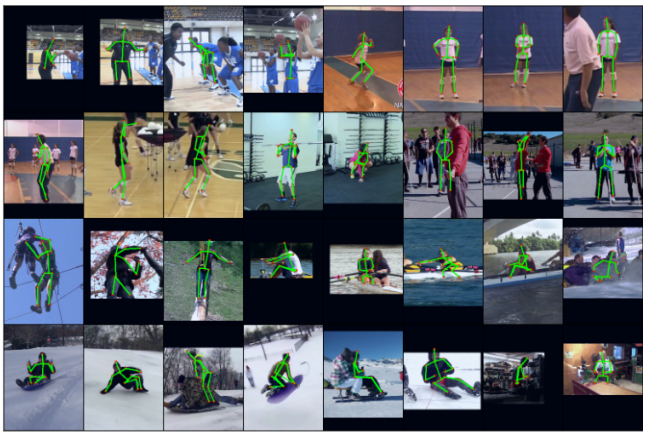
Privacy Parameter(ϵ)	Head	Shoulder	Elbow	Wrist	Hip	Knee	Ankle	Mean	Mean@0.1	inkoti
Finetuning										
C = 0.01										
$\epsilon = 0.2$	32.6	71.8	26.4	17.1	34.6	38.1	75.2	35.1	26.6	39.6
$\epsilon = 0.4$	37.3	74.6	34.2	20.7	39.5	42.6	77.7	42.4	30.4	44.2
$\epsilon = 0.6$	39.3	75.9	37.6	22.2	41.7	44.5	78.8	45.4	31.4	46.2
$\epsilon = 0.8$	40.5	76.9	39.4	23.1	42.7	45.6	79.4	47.0	32.5	47.3
C = 0.1										
$\epsilon = 0.2$	8.7	37.0	1.1	3.1	9.4	14.1	48.7	2.9	9.6	14.7
$\epsilon = 0.4$	13.5	47.2	2.1	4.8	14.6	18.8	55.9	6.8	12.3	19.6
$\epsilon = 0.6$	15.8	51.9	3.5	5.9	17.0	21.1	59.1	9.3	14.1	22.0
$\epsilon = 0.8$	17.1	53.5	4.6	6.6	18.4	22.4	60.6	10.8	14.9	23.3
Finetuning from scratch										
C = 0.01										
$\epsilon = 0.2$	4.8	23.0	0.2	0.9	5.5	10.4	38.4	2.4	6.3	10.9
$\epsilon = 0.4$	7.7	30.0	1.7	1.7	8.5	13.6	43.1	5.0	8.5	14.2
$\epsilon = 0.6$	8.8	32.8	2.3	2.0	9.7	14.8	45.2	6.3	9.3	15.4
$\epsilon = 0.8$	9.5	34.5	2.6	2.2	10.6	15.7	46.9	7.1	10.1	16.3
C = 0.1										
$\epsilon = 0.2$	0.2	1.9	0.0	0.0	0.3	2.3	13.6	0.0	1.4	2.5
$\epsilon = 0.4$	0.7	4.9	0.0	0.0	0.8	3.2	18.3	0.09	1.7	3.3
$\epsilon = 0.6$	1.0	7.0	0.0	0.2	1.1	3.9	21.0	0.1	2.7	4.0
$\epsilon = 0.8$	1.3	9.3	0.0	0.2	1.5	4.4	23.1	0.1	2.8	4.6

Table 9: HumanART Results: Feature Projection DP-SGD plus projection.

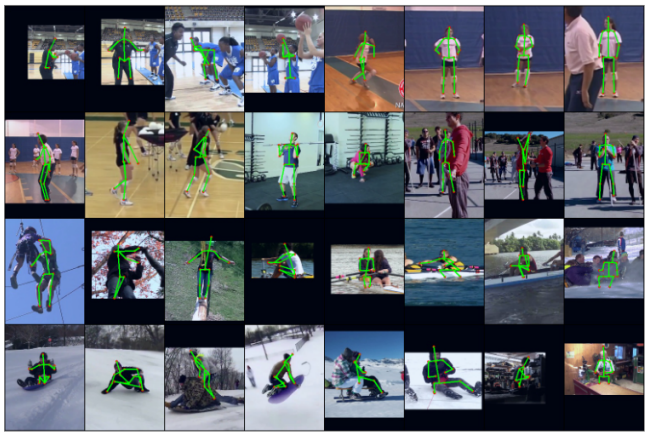
Privacy Parameter(ϵ)	Head	Shoulder	Elbow	Wrist	Hip	Knee	Ankle	Mean	Mean@0.1	inkoti
Finetuning										
C = 0.01										
$\epsilon = 0.2$	50.6	80.2	53.4	31.3	52.8	55.3	82.4	59.1	41.3	57.2
$\epsilon = 0.4$	51.6	80.4	55.0	33.5	53.8	56.2	82.7	60.3	42.2	58.1
$\epsilon = 0.6$	50.9	80.2	53.6	32.5	53.2	55.9	82.7	59.9	42.1	57.8
$\epsilon = 0.8$	51.6	80.2	54.7	33.6	53.7	56.2	82.5	60.2	42.3	58.1
C = 0.1										
$\epsilon = 0.2$	40.4	75.5	38.9	24.4	42.3	45.4	78.0	46.0	33.0	47.1
$\epsilon = 0.4$	41.1	76.8	40.2	24.6	43.3	46.8	79.5	48.2	34.4	48.4
$\epsilon = 0.6$	42.2	75.7	42.2	25.2	44.3	47.5	78.6	49.3	34.4	49.2
$\epsilon = 0.8$	43.0	77.0	42.9	25.7	45.1	48.4	79.9	50.3	35.7	50.1
C = 1.0										
$\epsilon = 0.2$	26.8	62.1	19.0	13.5	28.4	32.9	67.0	28.6	23.2	34.2
$\epsilon = 0.4$	32.4	70.5	25.4	19.0	34.2	39.0	74.7	36.6	29.9	40.3
$\epsilon = 0.6$	33.9	70.6	28.7	18.3	35.8	40.1	74.4	38.5	28.8	41.6
$\epsilon = 0.8$	35.4	72.1	30.7	20.2	37.2	40.7	75.1	38.6	29.9	42.1
Finetuning from scratch										
C = 0.01										
$\epsilon = 0.2$	45.3	76.9	46.4	29.5	47.3	50.4	79.7	53.0	38.0	52.1
$\epsilon = 0.4$	45.1	76.6	46.5	28.0	47.2	50.5	79.1	53.5	36.8	52.3
$\epsilon = 0.6$	43.8	75.2	44.5	28.5	45.6	49.6	78.0	52.2	38.3	51.1
$\epsilon = 0.8$	46.0	76.7	47.5	29.3	48.1	51.4	79.7	54.5	38.0	53.2
C = 0.1										
$\epsilon = 0.2$	6.8	24.4	2.3	2.8	7.4	11.6	35.9	5.2	8.8	12.0
$\epsilon = 0.4$	22.0	53.7	14.5	12.0	23.3	27.7	60.3	22.1	20.6	28.7
$\epsilon = 0.6$	29.1	62.4	24.0	17.7	30.8	34.9	67.8	32.0	26.2	36.1
$\epsilon = 0.8$	33.0	68.0	27.9	21.5	34.4	38.6	72.0	36.5	30.0	39.8



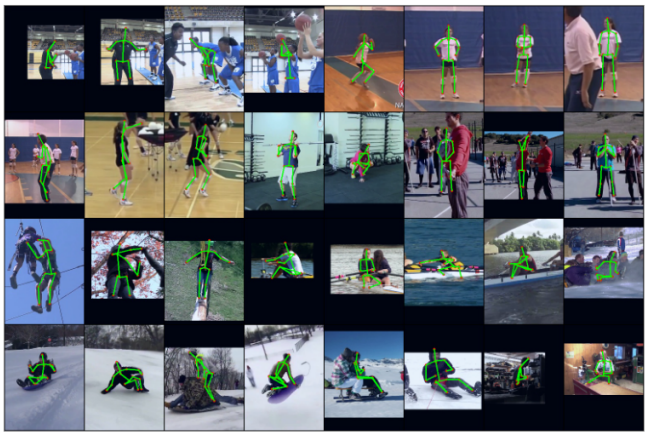
(a) Ground Truth



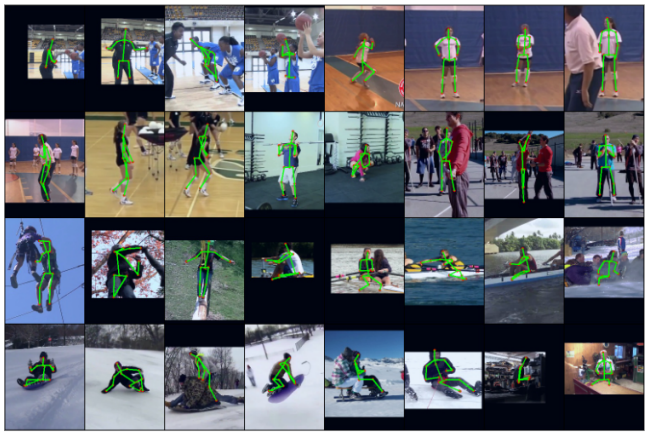
(d)
$\epsilon = 0.6$



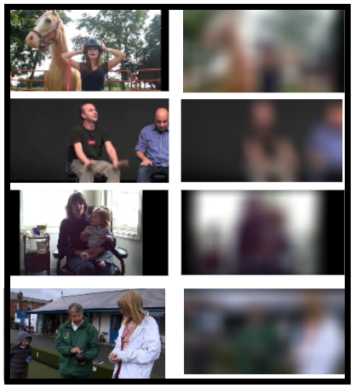
(b)
$\epsilon = 0.2$



(e)
$\epsilon = 0.8$



(c)
$\epsilon = 0.4$



(f) Raw vs Public Images

Figure 4: Figures (a-e)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. (f) Representation of Raw (Private) image compared to public feature (gaussian blurred).