Video-DPRP: A Differentially Private Approach for Visual Privacy-Preserving Video Human Activity Recognition

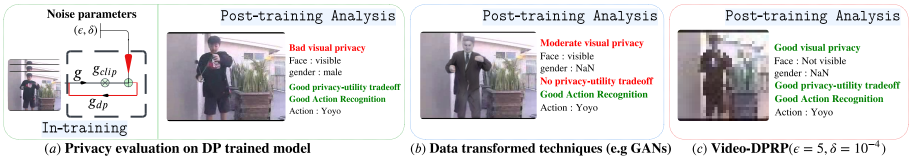
Allassan Tchangmena A Nken[1], Susan McKeever[3], Peter Corcoran[2], and Ihsan Ullah[1,4](✉)

[1] Visual Intelligence Lab, School of Computer Science, University of Galway, Ireland
[2] School of Electrical Engineering, University of Galway, Ireland
[3] School of Computer Science, Technological University Dublin, Ireland
[4] Insight Research Ireland Center for Data Analytics, University of Galway, Ireland
ihsan.ullah@universityofgalway.ie

Abstract. Considerable effort has been made in privacy-preserving video human activity recognition (HAR). Two primary approaches to ensure privacy preservation in Video HAR are differential privacy (DP) and visual privacy. Techniques enforcing DP during training provide strong theoretical privacy guarantees but offer limited capabilities for visual privacy assessment. Conversely methods, such as low-resolution transformations, data obfuscation and adversarial networks, emphasize visual privacy but lack clear theoretical privacy assurances. In this work, we focus on two main objectives: (1) leveraging DP properties to develop a model-free approach for visual privacy in videos and (2) evaluating our proposed technique using both differential privacy and visual privacy assessments on HAR tasks. To achieve goal (1), we introduce **Video-DPRP**: a **Video**-sample-wise **D**ifferentially **P**rivate **R**andom **P**rojection framework for privacy-preserved video reconstruction for HAR. By using random projections, noise matrices and right singular vectors derived from the singular value decomposition of videos, Video-DPRP reconstructs DP videos using privacy parameters (ϵ, δ) while enabling visual privacy assessment. For goal (2), using UCF101 and HMDB51 datasets, we compare Video-DPRP's performance on activity recognition with traditional DP methods, and state-of-the-art (SOTA) visual privacy-preserving techniques. Additionally, we assess its effectiveness in preserving privacy-related attributes such as facial features, gender, and skin color, using the PA-HMDB and VISPR datasets. Video-DPRP combines privacy-preservation from both a DP and visual privacy perspective unlike SOTA methods that typically address only one of these aspects. The source code is publicly available on GitHub [5].

Keywords: Activity Recognition · Differential Privacy · Visual Privacy.

[5] https://github.com/matzolla/Video-DPRP



(a) **Privacy evaluation on DP trained model** (b) **Data transformed techniques (e.g GANs)** (c) **Video-DPRP**($\epsilon = 5, \delta = 10^{-4}$)

Fig. 1. In (a), privacy is ensured during training (in-training) using differential privacy (DP), but not directly on the video itself. As a result visual privacy cannot be assessed. In (b), the video is transformed prior to training using either obfuscation methods or adversarial approaches, but the privacy-utility trade-off cannot be quantify as clearly as in DP. In (c) (ours), privacy is ensured using DP, directly on the video. This approach allows for visual privacy evaluation, where privacy-utility trade-off is quantified using the ϵ,δ parameters of DP.

1 Introduction

Privacy preservation is a critical research challenge in the field of video-based human activity recognition (HAR) and video analysis. Video HAR systems are increasingly used in settings like healthcare monitoring, smart homes and security [50,5,40]. However, these systems often capture sensitive personal information, creating a strong need for privacy measures to protect individuals' identities and personal activities from misuse or unauthorized access.

Current literature indicates that privacy preservation, in Video HAR can be achieved either at a model level or directly on the data by modifying its visual content. Model-based approaches usually ensure privacy by leveraging differential privacy (DP) [12,13,4,28]. This method provides a theoretical and empirical guarantee of privacy by incorporating noisy mechanisms into the training algorithms, using the privacy parameters ϵ and δ [1,32,35,9,10]. However, its effectiveness is limited when it comes to post-training privacy analysis such as visual privacy. In the context of video HAR, visual privacy can be define as a model's ability to recognize visual information such as faces, gender, or individuals performing activities. The underlying hypothesis is that diminished performance in these recognition tasks indicates higher visual privacy. As shown in Figure 1(a), models trained with DP cannot achieve this level of privacy because the data itself is not directly altered for DP; only the gradient's estimates g are adjusted during training.

Conversely, while some data-based approaches utilizing generative adversarial networks (GANs in Figure 1(b)) offer an affordable means of visual privacy assessment [31,38,18], the generated videos from these methods may still disclose sensitive visual content [41], as they are trained on unconstrained real-world data. Additionally GANs, including other video down-sampling and obfuscation approaches [39,38,21], lacks the rigorous mathematical privacy guarantees afforded by differential privacy. Theoretical privacy assurance is often overlooked in data-based methods, which typically rely on heuristic approaches, ad-hoc obfuscations, or data transformations. These methods lack transparency in how

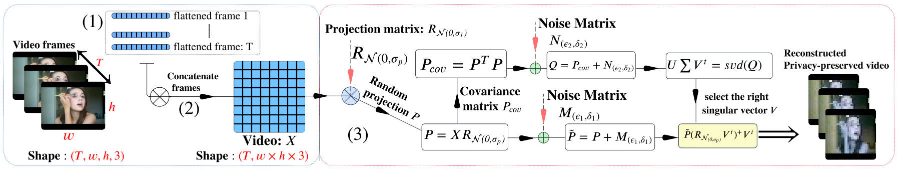

(1)
flattened frame 1
flattened frame: T
Video frames
T
h
w
Shape : $(T, w, h, 3)$
Concatenate frames
(2)
Video: X
Shape : $(T, w \times h \times 3)$
Projection matrix: $R_{\mathcal{N}(0,\sigma_1)}$
$R_{\mathcal{N}(0,\sigma_p)}$
Random projection P
(3)
$P_{cov} = P^T P$
Covariance matrix P_{cov}
$P = X R_{\mathcal{N}(0,\sigma_p)}$
Noise Matrix
$N_{(\epsilon_2,\delta_2)}$
$Q = P_{cov} + N_{(\epsilon_2,\delta_2)}$
Noise Matrix
$M_{(\epsilon_1,\delta_1)}$
$\tilde{P} = P + M_{(\epsilon_1,\delta_1)}$
$U \sum V^t = svd(Q)$
select the right singular vector V
$\tilde{P}(R_{\mathcal{N}(0,\sigma_p)} V^t)^+ V^t$
Reconstructed Privacy-preserved video

Fig. 2. Video-DPRP consists of the following components:(1) Each video frame is reshaped and flattened, then concatenated to form a video X of dimension $(T, w \times h \times 3)$. (2) A random projection matrix $\mathcal{R}_{\mathcal{N}(0,\sigma_p)}$ reduces X to a lower-dimensional space (T, k). (3) Noise is added to both the projected video and its covariance matrix, from which the right singular component V of the noisy covariance Q is used to reconstruct a differentially private video (see Section 3 for details).

privacy is preserved and can be vulnerable to reverse-engineering or sophisticated attacks [23], resulting in mere *security through obscurity*. In contrast, differential privacy is grounded in well-established mathematical principles that provide robust privacy guarantees, irrespective of an adversary's capabilities. Moreover, differential privacy offers clearer privacy explainability in terms of the chances of information leakage, quantified by the ϵ and δ parameters [33,6].

We identify two key limitations in previous privacy-preserving Video HAR studies: (1) Although DP models provide empirical and theoretical privacy guarantees during model training, their privacy-preserving effect does not extend beyond training. This limitation arises because the data itself remains unaltered, retaining visually sensitive content. Evaluating such data on visual privacy metrics is likely to yield poor results. (2) While some studies propose data-transformed methods for visual privacy evaluation, these approaches still fail to offer theoretical guarantees of privacy. Recent advancements in differential privacy and random projection present promising solutions. By leveraging a random projection matrix followed by the addition of a noise matrix to the projected data, previous work has demonstrated the feasibility of reconstructing differentially private tabular datasets and images [47,27,30,29,15]. However, applying differentially private random projections to a video dataset presents a significant challenge due to the added complexity introduced by the temporal dimension of videos.

In this work, we introduce Video-DPRP, a **Video**-sample-wise **D**ifferentially **P**rivate **R**andom **P**rojection framework tailored for visual privacy-preserved video reconstruction of HAR datasets. The framework unfolds in several stages: we begin by reshaping each video, as illustrated in Figure 2. Next, we apply a random projection to the reshaped video using a projection matrix, reducing its dimensionality while preserving its underlying structure. To ensure differential privacy, we add a noise matrix, calibrated with the (ϵ, δ) parameters, to both the projected video and its covariance matrix. Finally, by leveraging the right singular vectors from the singular value decomposition (SVD) of the noisy covariance matrix, we reconstruct a video sample, that is both visually and differentially private. **Ideally, a model trained with videos reconstructed using Video-DPRP is expected to exhibit both high-quality performance in video HAR and strong privacy preservation**. Our contributions are as follows:

- We introduce Video-DPRP, a differentially private approach for video reconstruction tailored for video HAR. Video-DPRP provides a theoretical guarantee of differential privacy, while also ensuring visual privacy.
- We evaluate the performance of Video-DPRP across both HAR and visual privacy-preserving attributes. For HAR evaluation, we use the UCF101 and HMDB51 datasets. To assess visual privacy-preserving attributes, we utilize the PA-HMDB and the VISPR datasets.

2 Related Work

Privacy-preserving video HAR. Privacy in the context of video HAR can be categorized into two main groups: visual privacy and differential privacy.
Visual privacy aims to obscure identifiable visual attributes in video content and can be categorized into 3 main groups: obfuscation, adversarial training and downsampling

Downsampling. As an example, Ryo et al. [39] proposed an inverse-super-resolution paradigm that learns an optimal set of transformations to generate low-resolution videos from high-resolution inputs. This approach utilizes a downsampling technique, similar to the method proposed by [7]. While this technique is effective, its major drawback lies in the trade-off between achieving accurate activity recognition and maintaining privacy preservation: a trade-off that could be better quantified with a rigorous mathematical bound on privacy.

Obfuscation. Ren et al. [38] presents a data obfuscation method for anonymizing facial images, using a learnable modifier. This approach employs an adversarial training setup, where a generator produces modified versions of facial images, and a discriminator attempts to identify facial features despite the modifications. The end result is a video anonymizer that performs pixel-level modifications to anonymize each person's face with minimal impact on action detection performance. Additional work on obfuscation has been conducted by Ilic et al., focusing on appearance-free action recognition using an optical-flow estimator [20] and selective video obfuscation using random noise [21]. However, obfuscation techniques have a limitation in that they require domain knowledge to effectively identify and obscure the region of interest.

Adversarial training. Beyond video down-sampling and obfuscation, some researchers have developed privacy optimization strategies using adversarial neural networks [46,36,8]. These strategies typically involve a cost function that is minimized for activity recognition, while simultaneously maximized for privacy preservation. A significant drawback of these techniques is their substantial computational resource requirements for reconstructing anonymized videos. In contrast, a more effective approach could be a model-free method capable of reconstructing videos at a considerably lower computational cost.

While visual privacy focuses on *hiding* identifiable visual attributes in sample videos, in differential privacy, a random noise is added to the gradient estimates during a model's training process. This noise is carefully calibrated to ensure that the model can still learn overall patterns and trends, while specific details that could identify a sample video are not leaked. It is important to note that the sample videos themselves are not directly modified; only their gradient estimates are altered during training. Figure $1(a)$ provides a clear illustration of training with differential privacy, specifically detailing a variant of stochastic gradient descent (SGD) known as differential private stochastic gradient descent (DP-SGD) [1]. DP-SGD differs from traditional SGD in that, after computing the per-sample gradient g it is clipped to a threshold value C, resulting in a clipped gradient g_{clip}. A Gaussian noise, calibrated with the DP parameters: ϵ, δ, is then carefully added to the clipped gradient producing the differentially private gradient g_{dp} (details about the DP parameters are provided in Section 3). Recently, Luo et al. [32] proposed Multi-Clip DP-SGD, a method designed to achieve video-level differential privacy in HAR. The DP framework is built such that, during model training, shorter video segments, or clips, are sampled from each video, and their gradients are computed and averaged across all the clips of the video. DP-SGD is then applied to the averaged gradient, ensuring differential privacy without additional privacy loss. Although the result is a differential private model, a significant challenge with DP-SGD and other DP learning algorithms is that privacy preservation is confined to the training phase, restricting further visual privacy assessments on the video data beyond training.

Differential Private Random Projection (DPRP). Previous research introduced DPRP primarily as a *data release* framework, for tabular data [47,15,25]. For instance Xu et al. [47] employed DPRP for the release of high-dimensional data, while Gondara et al. [15] adapted DPRP for smaller clinical datasets. In both scenarios, the original dataset is projected into a significantly lower-dimensional space using a random projection matrix, followed by the addition of a noise matrix. This noise matrix is calibrated with the (ϵ, δ) parameters to achieve differential privacy. In our approach, we apply DPRP on a per-video-sample basis rather than across the entire dataset, offering more granular privacy control and assessment on activity recognition.

3 Method Overview

We begin by introducing key concepts relevant to Video-DPRP, including an initial video transformation mechanism, the theoretical foundations of differential privacy, random projection, and the algorithmic framework of Video-DPRP. This section concludes with preliminary discussions of the privacy guarantees offered by Video-DPRP, which are further detailed in the Appendix.

3.1 Video Transformation

A sample video is structured as a $4D$ tensor, $(T,w,h,3)$, consisting of a $1D$ temporal dimension and $2D$ spatial dimensions. The temporal dimension is represented by the number of frames, T, in the video sequence, while the spatial dimensions are denoted by the pair (w, h), corresponding to the width and height of each frame. Moreover, each frame contains 3 color channels (red,green and blue). To facilitate our random projection strategy, we flattened the $2D$ spatial dimensions of each frame from $(1, w, h, 3)$ to $(1, w \times h \times 3)$. Next, we concatenate all the T flattened frames along the temporal axis (the first axis), resulting to $2D$ array X of dimension $(T, w \times h \times 3)$. This concatenation preserves the temporal sequence of the video, with each row of X corresponding to a flattened frame. This step is crucial for our subsequent methodology, and henceforth, we treat each video X as a $2D$ array.

3.2 Differential Privacy

We consider two sample videos, X and X', that differ by a single row, representing neighboring inputs. Intuitively, this means X and X' differ by one frame. Video-DPRP ensures that modifying the pixel values of a single frame does not pose a significant visual privacy risk, nor does it lead to a substantial drop in video HAR performance. This implies that even if an adversary knows the output video, they cannot infer sensitive information about the frame that was modified. We then give a formal definition introduced by Dwork et al. [11] and re-calibrated to our context:

Definition 1 (Differential Privacy). *A randomized mechanism $\mathcal{M}$, satisfies (ϵ, δ)-differential privacy if for any two input videos X and X', that differ in only one row (frame), and for all sets of possible outputs $O \in range(\mathcal{M})$, we have:*

$$\Pr[\mathcal{M}(X) \in O] \leq e^{\epsilon} \cdot \Pr[\mathcal{M}(X') \in O] + \delta$$

In other words, the outcomes of applying the random mechanism $\mathcal{M}$ to the two neighboring videos X and X' differ by at most a factor of e^{ϵ}. The privacy guarantee can fail with a probability of δ. When $\delta = 0$, the mechanism operates under pure ϵ-differential privacy.

3.3 Random Projection

Random projection is a dimensionality reduction technique that projects data from an initial dimension d to a lower dimension k, while preserving pairwise distances between data points (in our case, frames) using a projection matrix $\mathcal{R}$. To ensure that the pairwise distances between frames are preserved, the projection matrix must satisfy the Johnson-Lindenstrauss Lemma [24].

Lemma 1 (Johnson-Lindenstrauss [24]). *Let $\mathcal{S}$ be a set of n points such that $\mathcal{S} \subset \mathbb{R}^d$, with $\lambda > 0$ and $k = \frac{20 \log n}{\lambda^2}$. There exists a Lipschitz mapping $f : \mathbb{R}^d \to \mathbb{R}^k$ that distorts all pairwise distances by a factor of $1 \pm \lambda$. For any $x, y \in \mathbb{R}^d$, this mapping satisfies the following inequality:*

$$(1 - \lambda)\|x - y\|_2^2 \leq \|f(x) - f(y)\|_2^2 \leq (1 + \lambda)\|x - y\|_2^2$$

Contextually, for a given video X, the initial dimension is $d = w \times h \times 3$, where w and h are the width and height of the frames, respectively, and the set of n points corresponds to the number of frames T, as discussed earlier in section 3.1. To project the video $X^{T\times d}$, a random projection matrix $\mathcal{R}$ is required, such that the resulting projected video is $P = X\mathcal{R}$. A suitable random projection matrix that satisfies Lemma 1 is one whose entries are drawn from a normal distribution with mean $\mu = 0$ and variance $\sigma^2 = \frac{1}{k}$ (that is, $\mathcal{R} \sim \mathcal{N}(0, \frac{1}{\sqrt{k}})^{d\times k}$).

3.4 Video-DPRP Algorithm

The algorithmic framework of Video-DPRP is inspired by the influential work on the Johnson-Lindenstrauss transform [25], and recent developments in the release of small datasets [15].

Preliminary: Recall that each video is initially transformed into a $2D$ matrix of dimensions $(T \times d)$, where T, is the number of frames and $d = w \times h \times 3$ (with w being the width and h the height of a frame).

Privacy parameters: All our privacy parameters are derived from a single privacy parameter pair (ϵ, δ). To ensure that a given video remains differentially private without significantly compromising its utility, we avoid adding multiple independent noise matrices. Instead, we split the parameters into two sets: one set is used to make the random projection P differentially private (ϵ_1, δ_1) and the other set (ϵ_2, δ_2) to make the covariance matrix P_{cov}, differentially private. Each set is derived using the privacy budget allocator $b \in]0, 1[$ (see `lines 1-2`). The privacy budget is a parameter that controls the total amount of privacy loss allowed, balancing utility with privacy protection. The complete workflow of Video-DPRP is outlined in Algorithm 1. The time complexity of each step of the algorithm is highlighted in `blue`.

Algorithm 1 Video-DPRP

```
    Input: $D = \{X_1, X_2, ...., X_n\}, d \times k, \epsilon, \delta, b$
    /*The dataset D with n videos; the size of the projection matrix $d{\times}k$; the privacy parameters $\epsilon$ and $\delta$; the privacy budget allocator $b \in ]0,1[$ */
 1: $\epsilon_1, \delta_1 = \epsilon \times b, \delta \times b$
 2: $\epsilon_2, \delta_2 = \epsilon \times (1 - b), \delta \times (1 - b)$
 3: for $X^{T\times d} \in D$ do:
 4:     $\mathcal{R}\sim \mathcal{N}(0, \frac{1}{\sqrt{k}})^{d\times k}$/*projection matrix*/
 5:     $P = X\mathcal{R}$ /*Random projection:$O(Tdk)$*/
 6:     $\tilde{P} = P + M_{(\epsilon_1,\delta_1)}$ /*Noise addition:$O(Tk)$*/
 7:     $P_{cov} = P^t P$/*Covariance matrix:$O(Tk^2)$*/
 8:     $Q$=$P_{cov} + N_{(\epsilon_2,\delta_2)}$ /*Noise addition:$O(k^2)$*/
 9:     $U \sum V^t = \mathrm{SVD}(Q)$ /*Decomposition:$O(k^3)$*/
10:     $\tilde{X}$=$\tilde{P}\left(\mathcal{R}V^t\right)^+ V^t$/*reconstructed video:*/
    Output: reshaped video, reshape($\tilde{X}$)
```

To begin with, for each video $X^{T\times d}$ in the dataset D, we project the video into a lower-dimensional space, using the projection matrix $\mathcal{R}^{d\times k}$ (`lines 4-5`), which satisfies the JohnsonLindenstrauss Lemma 1. This result to a projected video P, of dimension $(T \times k)$. Here, k represents the number of dimensions for the random projection. At this stage, P still contains sensitive information from X and is therefore not differentially private. Differential privacy is ensured by adding a random noise matrix $M_{(\epsilon_1,\delta_1)}$ to the projected video (`line 6`), resulting to $\tilde{P}$. The entries of the random noise matrix are drawn from a Gaussian distribution with mean $\mu = 0$ and variance σ_1^2 ($M_{(\epsilon_1,\delta_1)} \sim \mathcal{N}(0, \sigma_1^2)^{T\times k}$). The variance σ_1^2 is determined using Theorem 1. Differentially private video reconstruction effectively begins at `line 7`, where the covariance matrix P_{cov} of the projected video P, is first computed as a necessary step for reconstruction. The use of the covariance matrix is motivated by principles similar to those in Principal Component Analysis, aiming to capture the most significant features of the video within the low-dimensional subspace. Similar to `line 5`, since P is not differentially private, its covariance matrix P_{cov} is also not. To achieve differential privacy, a random noise matrix $N_{(\epsilon_2,\delta_2)}$ is added to P_{cov}, resulting in a noisy covariance matrix Q (`line 8`) of dimension $(k \times k)$. In the same way, the entries of $N_{(\epsilon_2,\delta_2)}$ are drawn from a Gaussian distribution with mean $\mu = 0$ and variance σ_2^2 ($N(\epsilon_2, \delta_2) \sim \mathcal{N}(0, \sigma_2^2)^{k\times k}$). The variance σ_2^2 is determined using Theorem 2. To proceed, the noisy covariance matrix Q is subjected to a singular value decomposition (SVD), which decomposes Q into three matrices: $U\Sigma V^t$ (`line 9`), where U and V^t (denoting the transpose of V) are orthogonal matrices each of dimensions $(k \times k)$, and Σ is a diagonal matrix containing the singular values. Following the approach of [15], we only use the right singular component V^t, the random projection matrix $\mathcal{R}$, and the differentially private projected video $\tilde{P}$ for video reconstruction of $\tilde{X}$ (`line 10`). We use the Moore-Penrose pseudoinverse (denoted by `+`) of $\mathcal{R}V^t$ because $\mathcal{R}V^t$ is not a squared matrix and may not be invertible. $\tilde{X}$ has dimensions $(T \times d)$ and is ultimately reshaped back to its original video format $(T, w, h, 3)$.

Time complexity: Given that the algorithm processes n videos independently, the overall time complexity for the entire dataset D is $O(n(Tdk + Tk^2 + k^3))$. This complexity shows that the algorithm scales linearly with the number of videos n, and is influenced by both the number of frames T and the dimensionality d. The cubic term k^3 becomes dominant when the projection dimension k is large.

3.5 Privacy Guarantee of Video-DPRP

Differential privacy is applied at two stages in Algorithm 1: (i) to ensure that the projected video P is differentially private, and (ii) to make the covariance matrix P_{cov} differentially private. To establish the privacy guarantee of Video-DPRP, we must demonstrate that both stages meet differential privacy requirements. The proofs rely on two supporting theorems from [44,15], which are included here for completeness, with details provided in the appendix.

Theorem 1 (Privacy of projected video P). *Let $\epsilon_1 > 0$ and $0 < \delta_1 < \frac{1}{2}$. Consider a randomized Gaussian projection matrix $\mathcal{R} \sim \mathcal{N}(0, 1/\sqrt{k})^{d\times k}$. Then, the noisy projection $\tilde{P} = X\mathcal{R} + M_{(\epsilon_1,\delta_1)}$, where $M_{(\epsilon_1,\delta_1)}$ is a $(T \times k)$ Gaussian matrix with entries drawn from $\mathcal{N}(0, \sigma_1^2)$, is (ϵ_1, δ_1)-differentially private, with:*

$$\sigma_1 = \theta\sigma_p\sqrt{k + 2\sqrt{klog(2/\delta_1)} + 2log(2/\delta_1)}\sqrt{2(log(1/2\delta_1) + \epsilon_1)}/\epsilon_1$$

Where $\sigma_p = 1/\sqrt{k}$, and θ denotes the L_2 sensitivity bound of the input. The variables are consistent with those defined in Section 3.4 to maintain uniformity. **The L_2 sensitivity θ:** For the input $f(X) = X\mathcal{R}$ where X represents the video with pixel values ranging from $[0, 255]$ and $\mathcal{R}$ is a random matrix, the L_2 sensitivity θ is proportional to the maximum change in X, scaled by the norm of $\mathcal{R}$. This norm typically takes the value $1/\sqrt{k}$. Since the L_2 sensitivity of X is $|255 - 0|$, we define θ as $\theta = 255/\sqrt{k}$. Where $|.|$ denotes the absolute value. More details are provided in the appendix section.

Theorem 2 (Privacy of covariance matrix P_{cov}). *The mechanism defined by $Q = P_{cov} + N_{(\epsilon_2,\delta_2)}$, where $N_{(\epsilon_2,\delta_2)}$ is a Gaussian matrix with entries drawn from $\mathcal{N}(0, \sigma_2)$, is (ϵ_2, δ_2)-differentially private, provided that $\epsilon_2 > 0$ and $\delta_2 < 1/2$. Where $\sigma_2 = \theta\sqrt{\frac{\sqrt{2log(1.25)/\delta_2}}{\epsilon_2}}$.*

By applying the principle of sequential composition [13], each video X is (ϵ, δ)-differentially private as a result of the combination of two differentially private mechanisms in Algorithm 1. Where $\epsilon = \epsilon_1 + \epsilon_2$ and $\delta = \delta_1 + \delta_2$.

Table 1. Comparison with different visual privacy techniques, including data-obfuscation, adversarial training and video anonymization using GANS. cMAP and F1 metrics are for *privacy evaluation* while Top-1 is for *action evaluation*. Results are reported on UCF101 [42], HMDB51 [26],PA-HMDB [46] and VISPR [34]. The best results are in **bold-red**, while the second best are underlined.

	Raw Test set Top-1(↑)		**Reconstructed Test set** Top-1(↑)		**Raw Test set PA-HMDB**		**Raw Test set VISPR**	
Method	UCF101	HMDB51	UCF101	HMDB51	Top-1 (↑)	cMAP (↓)	cMAP (↓)	F1 (↓)
ISR$_{(32\times24)}$[39]	49.65±0.22	35.66±0.10	45.14±0.53	28.97±0.09	38.71±1.22	58.26±0.13	53.60±0.87	49.14±0.09
ISR$_{(16\times12)}$[39]	18.34±0.02	19.47±0.04	24.94±0.20	12.64±0.01	25.11±0.62	40.01±0.17	43.27±0.25	45.00±0.73
V-SAM[19]	17.32±0.30	14.72±0.12	10.02±1.48	12.03±0.91	15.31±0.54	40.39±0.38	44.64±0.09	**39.97**±0.16
Face Anonymizer[38]	32.05±0.49	19.04±0.24	21.62±0.35	21.13±0.69	17.04±0.03	41.18±1.09	44.00±0.63	51.43±0.39
SPAct[8]	60.82±0.33	**41.29**±0.01	-	-	44.13±0.73	60.55±0.75	56.71±0.18	47.61±0.11
ALF[46]	56.27±0.91	32.04±0.56	-	-	43.73±0.82	40.29±0.03	55.09±1.49	43.08±1.02
Deepprivacy[18]	16.72±0.36	11.54±0.05	14.95±0.58	11.69±0.90	18.77±0.13	39.76±0.83	42.06±0.28	41.27±0.50
Appearance free[20]	30.02±0.07	15.67±0.14	14.22±0.16	10.29±0.06	19.60±0.02	-	-	-
Selective privacy[21]	58.97±0.11	38.27±0.01	45.10±0.15	30.09±0.06	42.56±0.54	-	-	-
Face blurring[22]	51.07±0.63	37.98±0.21	40.01±0.64	28.81±0.01	37.00±0.44	42.13±0.68	47.04±0.33	52.34±0.27
Raw data (no privacy)	85.77±0.18	59.24±0.65	85.77±0.18	59.24±0.65	65.05±1.17	70.13±0.59	64.18±0.06	69.10±0.59
Video-DPRP$_{(\epsilon=2,\delta=10^{-4})}$	55.16±0.59	36.49±0.79	38.76±0.11	27.95±0.01	39.25±0.15	**39.42**±0.62	**41.89**±0.02	40.03±0.28
Video-DPRP$_{(\epsilon=5,\delta=10^{-4})}$	58.58±0.16	38.37±0.09	45.20±0.02	29.06±0.77	44.86±0.02	48.75±0.07	50.11±0.63	53.77±0.30
Video-DPRP$_{(\epsilon=8,\delta=10^{-4})}$	**61.69**±0.07	40.00±0.71	**50.00**±0.28	**32.13**±0.31	**51.07**±0.73	53.00±0.01	56.10±0.20	55.12±0.03

Table 2. Comparison with differential private training methods and Video-DPRP on action recognition, for $\epsilon \in \{2, 5, 8\}$ and δ=10^{-4}. The best results are in red , and the second best are underlined.

Method	**Raw Test set UCF101** (Top-1 (↑))			**Raw Test set HMDB51** (Top-1 (↑))			**Raw Test set PA-HMDB** (Top-1 (↑))		
	$\epsilon = 2$	$\epsilon = 5$	$\epsilon = 8$	$\epsilon = 2$	$\epsilon = 5$	$\epsilon = 8$	$\epsilon = 2$	$\epsilon = 5$	$\epsilon = 8$
DP-SGD[1]	25.54±0.33	37.24±0.26	45.32±0.86	14.18±0.06	30.09±0.18	32.16±1.85	15.34±2.15	25.70±1.97	29.90±0.04
MultiClip-DP$_{(3\ clips)}$[32]	44.07±0.18	70.03±0.13	72.03±0.71	36.11±0.25	48.00±0.05	50.98±0.66	37.06±0.24	45.16±0.05	52.73±0.46
Video-DPRP	55.16±0.59	58.58±0.16	61.69±0.07	36.49±0.79	38.37±0.09	40.00±0.71	39.25±0.15	44.86±0.02	51.07±0.73
Video-DPRP$_{(3\ clips)}$	60.11±0.10	70.87±0.03	74.06±0.38	42.72±0.44	49.63±0.95	51.63±0.53	41.08±0.04	48.17±0.42	54.98±0.86

4 Experiments

4.1 Datasets

We adopt **PA-HMDB** [46] and **VISPR** [34] for visual privacy assessment, and **UCF101** [42] and **HMDB51** [26] for HAR, as these are commonly used datasets in the literature.

PA-HMDB[46] is a dataset containing 515 videos with video-level action annotations and frame-wise visual privacy annotations, including privacy attributes such as *skin color, face, gender, nudity*, and *relationship*. The dataset covers 51 action classes.

VISPR [34] is an image dataset designed for visual privacy research. It contains various personal attributes similar to those in HMDB51. The dataset comprises $10,000$ training images, $4,100$ validation images, and $8,000$ test images.

UCF101 [42] and **HMDB51** [26] are both HAR datasets, containing 101 and 51 action classes, respectively. For both datasets, all results are reported on split-1, which includes $9,537$ training videos and $3,783$ test videos for UCF101, and $3,570$ training videos and $1,530$ test videos for HMDB51.

4.2 Implementation details

Many deep learning models incorporate Batch Normalization (Batch Norm) layers. However, such models are not compatible with differentially private training methods like DP-SGD [1] or MultiClip-DP-SGD [32] (abbreviated to MultiClip-DP in Table 2), as Batch Norm requires calculating the mean and standard deviation for each mini-batch, introducing dependencies between samples and violating the principles of differential privacy. For fair comparison across all our results in video HAR, we require a model with a different type of normalization layer. Therefore, we use the PyTorch implementation of the Multiscale Vision Transformer (MViT-B$_{(16\times4)}$) [14], which employs Layer Normalization [2] and is pre-trained on the large-scale Kinetics-400 dataset [3]. For each video, we randomly crop a clip consisting of 16 frames, with each frame resized to a shape of $(224, 224, 3)$. In the case of Video-DPRP$_{(3\text{ clips})}$ and MultiClip-DP$_{(3\text{ clips})}$ (see Table 2), we crop 3 clips and apply the same pre-processing as described above. The optimization is performed using stochastic gradient descent (SGD) with a learning rate of $lr = 0.01$, a batch size of 8, and 50 training epochs.

Set-up of Video-DPRP: We use video samples reconstructed by Video-DPRP as inputs for our training. In line with Algorithm 1, we set the dimensions of the projection matrix to $d \times k$, where $d = 320 \times 240 \times 3$ and $k = 32 \times 32 \times 3$. Note that d corresponds to the dimensions of a frame from the original video (as described in Section 3.4) and is therefore fixed to the value defined above by default. We set the privacy budget allocator b to 0.8, meaning that 80% of the privacy budget is allocated to making the random projection P, differentially private (see `line 6`) while the remaining 20% (i.e, $1 - b$) is used to ensure the covariance matrix P_{cov} is differentially private (see `line 8` of Algorithm 1).

For the differentially private training of DP-SGD [1] and MultiClip-DP$_{(3 \text{ clips})}$ [32], we use the PyTorch Opacus library [48], which includes a privacy budget accountant to track the differentially private parameters (ϵ, δ) during training. For fair comparison and simplicity across all differentially private techniques (i.e., DP-SGD [1], MultiClip-DP$_{(3 \text{ clips})}$ [32] and Video-DPRP), we set the privacy parameter δ=10^{-4} and only vary ϵ. All experiments were conducted on an NVIDIA RTX A6000 GPU. For a comparative analysis with state-of-the-art (SOTA) visual privacy techniques, such as **Obfuscation** and **Anonymization**, we replicate their techniques following the authors' descriptions.

Obfuscation methods: specifically, we compared our approach with: (1) a Inverse-Super-Resolution (ISR) [39] model, which initially downscales videos and further perform a set of transformations (rotations, cropping) to the downscaled videos ; (2) a Face Blurring [22] algorithm, that detects faces using YOLOv3 [37] and applies Gaussian blurring with a kernel size of $k = 21$, and a standard deviation $\sigma = 10$ for consistency with prior works [46,21]; and finally (3) a Appearance-Free Privacy model [20], which removes appearance cues on videos via optical flow warping [43]. For comparisons with **Anonymization techniques** we used: (4) DeepPrivacy [18] which perform a full-body video anonymization, (5) a video Face Anonymizer [38], and (6) a surface-adaptive modulation: V-SAM [19]. **Adversarial training strategies** include (7) a adversarial learning framework (ALF)[46], which utilizes a adversarial privacy budget, and (8) SPAct [8], which employs self-supervised learning with MViT-B(16×4) as the target classifier. We also implement differentially private training with (10) DP-SGD [1] and (11) MultiClip-DP(3 clips) [32] using a clipping norm of $C = 0.4$. Full SOTA experimental details are included in the appendix.

4.3 Evaluation Metrics and Protocols

Metrics: Action recognition evaluation is conducted using the Top-1 accuracy metric, following prior work [17,32,16]. For visual privacy recognition, considered as a multi-label image classification task due to the presence of multiple privacy attributes per image, we use the class-wise mean average precision (cMAP) [34] and the class-wise F1-score. All results are reported as percentages, averaged over three runs, with both their mean and variance provided. In our tables, ↑ denotes metrics where higher values are better, while ↓ indicates that lower values are better.

Protocols: Apart from **Adversarial training** methods, which ensure privacy directly during training, we apply two evaluation protocols for video HAR with visual privacy techniques. Protocol 1 evaluates on the raw test set X_{raw}^{test} of dataset $\mathrm{X} \in \{\textbf{UCF101}, \textbf{HMDB51}, \textbf{PA-HMDB}\}$, after training our model on the corresponding reconstructed train set $\mathrm{X}_{reconst}^{train}$ using a method $reconst \in \{\textbf{Obfuscation}, \textbf{Anonymization}, \textbf{Video-DPRP}\}$.

Here, **Obfuscation** and **Anonymization** refer to all obfuscation and anonymization techniques described in Section 4.2. It is important to note that for evaluation on the **PA-HMDB** dataset, we use **HMDB51**\\{**PA-HMDB**} as our training set. This means that all video samples present in **HMDB51** but not

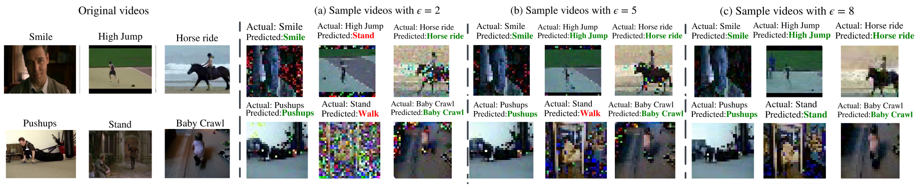

Original videos
Smile
High Jump
Horse ride
Pushups
Stand
Baby Crawl
(a) Sample videos with $\epsilon = 2$
Actual: Smile Predicted:Smile
Actual: High Jump Predicted:Stand
Actual: Horse ride Predicted:Horse ride
Actual: Pushups Predicted:Pushups
Actual: Stand Predicted:Walk
Actual: Baby Crawl Predicted:Baby Crawl
(b) Sample videos with $\epsilon = 5$
Actual: Smile Predicted:Smile
Actual: High Jump Predicted:High Jump
Actual: Horse ride Predicted:Horse ride
Actual: Pushups Predicted:Pushups
Actual: Stand Predicted:Walk
Actual: Baby Crawl Predicted:Baby Crawl
(c) Sample videos with $\epsilon = 8$
Actual: Smile Predicted:Smile
Actual: High Jump Predicted:High Jump
Actual: Horse ride Predicted:Horse ride
Actual: Pushups Predicted:Pushups
Actual: Stand Predicted:Stand
Actual: Baby Crawl Predicted:Baby Crawl

Fig. 3. Visual correlation: From left to right, we present video samples processed using Video-DPRP with varying $\epsilon = \{2, 5, 8\}$, while maintaining a fixed lower-dimensional space of $k = 32 \times 32 \times 3$, a privacy parameter of $\delta = 10^{-4}$, and a privacy budget of $b = 0.8$. Lower ϵ values introduce more *noise*, degrading reconstruction quality and leading to less accurate predictions. Incorrect classes are highlighted in **red**, while correct predictions are marked in **green**.

in **PA-HMDB** are used for training in this scenario. Protocol 2 evaluates on the reconstructed test set $X^{test}_{reconst}$ of dataset X, after training on $X^{train}_{reconst}$, using method *reconst*. Accordingly, no results are provided for adversarial training methods in the *reconstructed test set* column of Table 1. **Protocol 1** assesses the model's robustness in real-world scenarios where obfuscation or anonymization might not be applied, while testing on the reconstructed data (**Protocol 2**) measures performance consistency under privacy-preserving transformations, validating model adaptability across both standard and privacy-focused settings.

In Table 2, we restrict the analysis of video HAR to differentially private training methods: DP-SGD [1] and MultiClip-DP$_{(3\ \text{clips})}$ [32], alongside Video-DPRP, as these are the only methods that incorporate differential privacy. For visual privacy evaluation, we begin by training our model on the training set of **VISPR**, formulating the task as a multi-label image classification problem due to the multiple privacy attributes per image. We then use the annotated video frames from **PA-HMDB** as our test set. This is considered a cross-dataset evaluation protocol, as outlined in [8]. We also evaluate on the test set of **VISPR**, as reported in Table 1. We can observe from Table 1 that **Video-DPRP** provides competitive results, highlighted in **bold**, when compared to SOTA privacy-preserving techniques in both activity recognition and visual privacy preservation. Notably, the performance of **Video-DPRP** with ϵ=8 and δ=10^{-4} (i.e Video-DPRP$_{(\epsilon=8,\delta=10^{-4})}$, in Table 1) shows a significant improvement. However, Video-DPRP$_{(\epsilon=8,\delta=10^{-4})}$ shows a slight performance drop of **1.29%** in activity recognition on the HMDB51 dataset compared to SPAct[8], which achieved a baseline accuracy of **41.29%**. In terms of visual privacy, we observe that Video-DPRP achieved a cMAP score of **39.76%** on PA-HMDB51 and **42.06%** on VISPR (with ϵ=2), outperforming state-of-the-art methods such as ISR$_{(32\times24)}$[39] and V-SAM[19]. Despite yielding decent scores on visual privacy, anonymization methods such as DeepPrivacy[18], V-SAM[19] and Face Anonymizer [38] as well as obfuscation method like ISR$_{(16\times12)}$[39], struggle to achieve good utility performance on HAR, with results dropping as low as **12.00%**. Intuitively, DeepPrivacy[18], V-SAM[19] and Face Anonymizer[38]

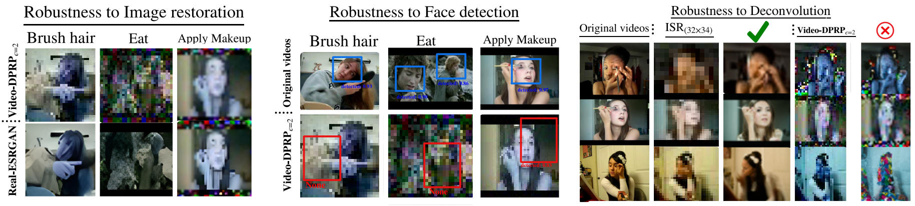

Robustness to Image restoration
Brush hair
Eat
Apply Makeup
Video-DPRP$_{\epsilon=2}$
Real-ESRGAN
Robustness to Face detection
Brush hair
Eat
Apply Makeup
Original videos
Video-DPRP$_{\epsilon=2}$
Robustness to Deconvolution
Original videos
ISR$_{(32\times34)}$
✓
Video-DPRP$_{\epsilon=2}$
⊗

Fig. 4. Robustness analysis. We evaluated the robustness of Video-DPRP with $\epsilon = 2$ against three types of attack: (1) Image Restoration (Left): Using Real-ESRGAN [45] a SOTA super-resolution generative adversarial network, we attempted to restore images from Video-DPRP$_{\epsilon=2}$. However, the image's quality remained significantly degraded (see second row). (2) Face Detection (Middle): We employed a pre-trained YOLOv3 [37] to detect faces in the original video (first row, **blue** bounding boxes) and in videos reconstructed using Video-DPRP$_{\epsilon=2}$(second row). The poor localization of the **red** bounding boxes highlights Video-DPRP's strong obfuscation effect. (3) Deconvolution Attacks (Right): We tested SUPIR [49], a SOTA deconvolution model on downscaled ISR [39] videos (16×12 resolution) and Video-DPRP$_{\epsilon=2}$. While SUPIR successfully recovers visual attributes from ISR videos (**green checkbox** column), it fails on Video-DPRP$_{\epsilon=2}$ (**red cross** column), demonstrating our method's resilience to deconvolution attacks. Full experimental details, along with additional results on face and skin-color detection, are provided in the appendix.

generate a *modified* version of the original video, which often fails to consistently preserve the motions of individuals involved in the activity. We conclude that while the above anonymization methods yield good privacy results, they may not be suitable for utility analysis in Video HAR. For obfuscation techniques, we argue that the visual content may be so *obscured* that models struggle to effectively identify activities. In contrast, Video-DPRP strikes a balance between utility and privacy, even for varying values of $\epsilon \in \{2, 5, 8\}$. We do not report the privacy results for Appearance-Free [20] and Selective Privacy [21], as both methods rely on optical flow between successive frames in videos for obfuscation, which is not applicable in our experiment since we use VISPR as the primary training set for privacy evaluation. In Figure 3, we present visual results of sample videos from UCF101 and HMDB51 with $\epsilon \in \{2, 5, 8\}$, along with their predicted classes. In Table 2, we use DP-SGD [1] as a baseline method and compare our results with MultiClip-DP$_{(3\text{ clips})}$ [32]. It is important to note that the results for MultiClip-DP$_{(3\text{ clips})}$ [32] are based on our own experiments, as the original code was not available. With 3 *clips* per sample video, Video-DPRP$_{(3\text{ clips})}$ provides competitive results when compared to MultiClip-DP$_{(3\text{ clips})}$, achieving Top-1 accuracy of **74.06%** on UCF101, **51.63%** on HMDB51 and **54.98%** on PA-HMDB with ϵ=8. Figure 4 presents a robustness analysis of Video-DPRP against image restoration, face detection, and deconvolution, simulating an adversarial scenario where an attacker attempts to compromise visual privacy from reconstructed videos. Additional analysis details are provided in the appendix.

5 Ablation Study

Keeping δ constant, we observe that increasing ϵ improves the action recognition performance of Video-DPRP but significantly reduces privacy, which is a typical behavior of differentially private algorithms. Video-DPRP is also influenced by two major components in its algorithm: the projection dimensionality k and the privacy budget b.

Effect of varying the dimensionality k: For a fixed ϵ=8, δ=10^{-4}, and a privacy budget of b=0.8, we observed that increasing the dimensionality k improves action recognition performance but results in a significant decrease in privacy, as shown in Table 3. This is because, the value of k, has a diminishing effect on the noise scale, σ_p= $1/\sqrt{k}$ and also on the L_2 sensitivity, $\theta = 255/\sqrt{k}$. As a result, when k increases, it substantially reduces the standard deviation value σ_1 in Theorem 1 and σ_2 in Theorem 2, leading to a decrease in the amount of noise added for differential privacy. Selecting an optimal k requires balancing performance gains with acceptable privacy levels for practical viability.

Effect of varying the privacy budget b: Recall that b, represents the privacy budget allocated to make the random projection differentially private, while $1 - b$, ensures the differential privacy of the covariance matrix (see Algorithn 1). To understand the effect of varying b, we fixed ϵ=8, δ=10^{-4} and k=$32 \times 32 \times 3$. Table 4 shows that increasing the privacy budget for random projection up to a value of **80%**, results in a less noisy random projection. Consequently, there is an increase in action recognition performance but with a substantial decrease in privacy. This suggests that the random projection plays a more critical role compared to the covariance matrix, in Video-DPRP.

Computational efficiency: We measured the time taken to reconstruct privacy-preserved videos from the UCF101 and HMDB51 datasets using different methods, as shown in Table 5. Although Video-DPRP has a polynomial time complexity as outlined in Section 3.4, it remains computationally efficient with an average reconstruction rate of ∼ **20 sec/Video** for both datasets. In contrast, V-SAM[19], Appearance-Free[20], and Face Blurring[22] incur additional computational overhead due to their use of surface-guided GANs, YOLO, and optical flow models, respectively.

	PA-HMDB	
Dimension k	**Action ↑**	**Privacy ↓**
20 × 20 × 3	24.60±1.67	**32.46**±0.75
24 × 32 × 3	28.03±0.07	40.17±1.22
32 × 32 × 3	51.07±0.73	53.00±0.01
50 × 50 × 3	67.18±0.49	56.96±0.07
64 × 80 × 3	**70.10**±0.14	60.18±0.33

Table 3. Action (Top-1) and privacy (cMAP) scores on **PA-HMDB**[46] for different lower dimensions k. The best result is highlighted in **bold**, and the second best is underlined.

	PA-HMDB	
Budget b	**Action ↑**	**Privacy ↓**
0.2	37.80±0.92	**29.02**±0.12
0.4	40.26±0.17	36.73±1.01
0.5	43.80±0.85	44.27±0.49
0.6	46.39±0.22	50.01±0.14
0.8	**51.07**±0.73	53.00±0.01

Table 4. Action (Top-1) and privacy (cMAP) scores on **PA-HMDB**[46] for different privacy budget b. The best result is highlighted in **bold**, and the second best is underlined.

	Reconstruction (sec/Video)	
Methods	**UCF101**	**HMBD51**
V-SAM[18]	33.12	35.07
ISR$_{(32\times24)}$[39]	**19.24**	**18.97**
Appearance free [20]	23.74	21.60
Face blurring [22]	26.08	24.49
Video-DPRP(ours)	20.32	19.84

Table 5. Reconstruction time per video (in seconds) for **UCF101** [42] and **HMDB51** [26]. The best (lowest) time is highlighted in **bold**, and the second best is underlined.

6 Discussion

While Video-DPRP offers strong theoretical and empirical guarantees, we acknowledge that its computational overhead particularly due to the SVD-based reconstruction step, has not been benchmarked against real-time video processing constraints. Although our reconstruction time ($\sim$ 20sec/video) is competitive compared to many SOTA visual privacy techniques (e.g., Face Blurring, V-SAM), it may still limit deployment in latency-sensitive or edge scenarios. Future work will focus on profiling runtime on embedded systems (e.g., Jetson Nano) and optimizing matrix operations, with the goal of enabling lightweight, real-time inference pipelines. We also plan to investigate approximations to SVD and reduced projection dimensionality k to balance speed, privacy, and utility.

7 Conclusion

This paper introduces Video-DPRP, a differentially private approach for constructing visual privacy-preserved videos for Human Activity Recognition (HAR). Video-DPRP aims to bridge the gap between visual privacy and utility by providing strong privacy guarantees through the mathematical properties of differential privacy and random projection. Our evaluation across multiple datasets demonstrate that Video-DPRP achieves competitive performance in activity recognition while maintaining robust privacy preservation compared to current state-of-the-art techniques.

Acknowledgments. This study was funded by the Research Ireland Centre for Research Training in Digitally-Enhanced Reality (D-real), under Grant No. 18/CRT/6224 and with the financial support of Insight Research Ireland Centre for Data Analytics under Grant number SFI/12/RC/2289_P2. For the purpose of open access, the author has applied for a CC BY public copyright license to any author-accepted manuscript version arising from this submission.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Abadi, M., Chu, A., Goodfellow, I., McMahan, H.B., Mironov, I., Talwar, K., Zhang, L.: Deep learning with differential privacy. In: Proceedings of the 2016 ACM SIGSAC conference on computer and communications security. pp. 308–318 (2016)
2. Ba, J.L.: Layer normalization. arXiv preprint arXiv:1607.06450 (2016)
3. Carreira, J., Zisserman, A.: Quo vadis, action recognition? a new model and the kinetics dataset. In: proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 6299–6308 (2017)
4. Chaudhuri, K., Monteleoni, C., Sarwate, A.D.: Differentially private empirical risk minimization. Journal of Machine Learning Research **12**(3) (2011)

5. Cristina, S., Despotovic, V., Pérez-Rodríguez, R., Aleksic, S.: Audio-and video-based human activity recognition systems in healthcare. IEEE Access (2024)
6. Cummings, R., Kaptchuk, G., Redmiles, E.M.: " i need a better description": An investigation into user expectations for differential privacy. In: Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security. pp. 3037–3052 (2021)
7. Dai, J., Saghafi, B., Wu, J., Konrad, J., Ishwar, P.: Towards privacy-preserving recognition of human activities. In: 2015 IEEE international conference on image processing (ICIP). pp. 4238–4242. IEEE (2015)
8. Dave, I.R., Chen, C., Shah, M.: Spact: Self-supervised privacy preservation for action recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 20164–20173 (2022)
9. Davody, A., Adelani, D.I., Kleinbauer, T., Klakow, D.: On the effect of normalization layers on differentially private training of deep neural networks. arXiv preprint arXiv:2006.10919 (2020)
10. Du, J., Li, S., Chen, X., Chen, S., Hong, M.: Dynamic differential-privacy preserving sgd. arXiv preprint arXiv:2111.00173 (2021)
11. Dwork, C., Kenthapadi, K., McSherry, F., Mironov, I., Naor, M.: Our data, ourselves: Privacy via distributed noise generation. In: Advances in Cryptology-EUROCRYPT 2006: 24th Annual International Conference on the Theory and Applications of Cryptographic Techniques, St. Petersburg, Russia, May 28-June 1, 2006. Proceedings 25. pp. 486–503. Springer (2006)
12. Dwork, C., McSherry, F., Nissim, K., Smith, A.: Calibrating noise to sensitivity in private data analysis. In: Theory of Cryptography: Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, March 4-7, 2006. Proceedings 3. pp. 265–284. Springer (2006)
13. Dwork, C., Roth, A., et al.: The algorithmic foundations of differential privacy. Foundations and Trends® in Theoretical Computer Science **9**(3–4), 211–407 (2014)
14. Fan, H., Xiong, B., Mangalam, K., Li, Y., Yan, Z., Malik, J., Feichtenhofer, C.: Multiscale vision transformers. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 6824–6835 (2021)
15. Gondara, L., Wang, K.: Differentially private small dataset release using random projections. In: Conference on Uncertainty in Artificial Intelligence. pp. 639–648. PMLR (2020)
16. Hara, K., Kataoka, H., Satoh, Y.: Learning spatio-temporal features with 3d residual networks for action recognition. In: Proceedings of the IEEE international conference on computer vision workshops. pp. 3154–3160 (2017)
17. Hara, K., Kataoka, H., Satoh, Y.: Can spatiotemporal 3d cnns retrace the history of 2d cnns and imagenet? In: Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. pp. 6546–6555 (2018)
18. Hukkelås, H., Lindseth, F.: Deepprivacy2: Towards realistic full-body anonymization. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. pp. 1329–1338 (2023)
19. Hukkelås, H., Smebye, M., Mester, R., Lindseth, F.: Realistic full-body anonymization with surface-guided gans. In: Proceedings of the IEEE/CVF Winter conference on Applications of Computer Vision. pp. 1430–1440 (2023)
20. Ilic, F., Pock, T., Wildes, R.P.: Is appearance free action recognition possible? In: European Conference on Computer Vision. pp. 156–173. Springer (2022)

21. Ilic, F., Zhao, H., Pock, T., Wildes, R.P.: Selective interpretable and motion consistent privacy attribute obfuscation for action recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 18730–18739 (2024)
22. Jaichuen, T., Ren, N., Wongapinya, P., Fugkeaw, S.: Blur & track: real-time face detection with immediate blurring and efficient tracking. In: 2023 20th International Joint Conference on Computer Science and Software Engineering (JCSSE). pp. 167–172. IEEE (2023)
23. Jang, J., Lyu, H., Hwang, S., Yang, H.J.: Unveiling hidden visual information: A reconstruction attack against adversarial visual information hiding. arXiv preprint arXiv:2408.04261 (2024)
24. Johnson, W.B.: Extensions of lipshitz mapping into hilbert space. In: Conference modern analysis and probability, 1984. pp. 189–206 (1984)
25. Kenthapadi, K., Korolova, A., Mironov, I., Mishra, N.: Privacy via the johnson-lindenstrauss transform. arXiv preprint arXiv:1204.2606 (2012)
26. Kuehne, H., Jhuang, H., Garrote, E., Poggio, T., Serre, T.: Hmdb: a large video database for human motion recognition. In: 2011 International conference on computer vision. pp. 2556–2563. IEEE (2011)
27. Lee, D., Yang, M.H., Oh, S.: Fast and accurate head pose estimation via random projection forests. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 1958–1966 (2015)
28. Levy, D., Sun, Z., Amin, K., Kale, S., Kulesza, A., Mohri, M., Suresh, A.T.: Learning with user-level privacy. Advances in Neural Information Processing Systems **34**, 12466–12479 (2021)
29. Li, P., Li, X.: Smooth flipping probability for differential private sign random projection methods. Advances in Neural Information Processing Systems **36** (2024)
30. Li, T., Wang, H., Zhuang, Z., Sun, J.: Deep random projector: Accelerated deep image prior. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 18176–18185 (2023)
31. Li, T., Lin, L.: Anonymousnet: Natural face de-identification with measurable privacy. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. pp. 0–0 (2019)
32. Luo, Z., Zou, Y., Yang, Y., Durante, Z., Huang, D.A., Yu, Z., Xiao, C., Fei-Fei, L., Anandkumar, A.: Differentially private video activity recognition. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 6657–6667 (2024)
33. Nanayakkara, P., Smart, M.A., Cummings, R., Kaptchuk, G., Redmiles, E.M.: What are the chances? explaining the epsilon parameter in differential privacy. In: 32nd USENIX Security Symposium (USENIX Security 23). pp. 1613–1630 (2023)
34. Orekondy, T., Schiele, B., Fritz, M.: Towards a visual privacy advisor: Understanding and predicting privacy risks in images. In: Proceedings of the IEEE international conference on computer vision. pp. 3686–3695 (2017)
35. Pichapati, V., Suresh, A.T., Yu, F.X., Reddi, S.J., Kumar, S.: Adaclip: Adaptive clipping for private sgd. arXiv preprint arXiv:1908.07643 (2019)
36. Pittaluga, F., Koppal, S., Chakrabarti, A.: Learning privacy preserving encodings through adversarial training. In: 2019 IEEE Winter Conference on Applications of Computer Vision (WACV). pp. 791–799. IEEE (2019)
37. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: Unified, real-time object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 779–788 (2016)

38. Ren, Z., Lee, Y.J., Ryoo, M.S.: Learning to anonymize faces for privacy preserving action detection. In: Proceedings of the european conference on computer vision (ECCV). pp. 620–636 (2018)
39. Ryoo, M., Rothrock, B., Fleming, C., Yang, H.J.: Privacy-preserving human activity recognition from extreme low resolution. In: Proceedings of the AAAI conference on artificial intelligence. vol. 31 (2017)
40. Shojaei-Hashemi, A., Nasiopoulos, P., Little, J.J., Pourazad, M.T.: Video-based human fall detection in smart homes using deep learning. In: 2018 IEEE International Symposium on Circuits and Systems (ISCAS). pp. 1–5. IEEE (2018)
41. Shokri, R., Stronati, M., Song, C., Shmatikov, V.: Membership inference attacks against machine learning models. In: 2017 IEEE symposium on security and privacy (SP). pp. 3–18. IEEE (2017)
42. Soomro, K.: Ucf101: A dataset of 101 human actions classes from videos in the wild. arXiv preprint arXiv:1212.0402 (2012)
43. Teed, Z., Deng, J.: Raft: Recurrent all-pairs field transforms for optical flow. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16. pp. 402–419. Springer (2020)
44. Tu, S.: Differentially private random projections
45. Wang, X., Xie, L., Dong, C., Shan, Y.: Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 1905–1914 (2021)
46. Wu, Z., Wang, H., Wang, Z., Jin, H., Wang, Z.: Privacy-preserving deep action recognition: An adversarial learning framework and a new dataset. IEEE Transactions on Pattern Analysis and Machine Intelligence **44**(4), 2126–2139 (2020)
47. Xu, C., Ren, J., Zhang, Y., Qin, Z., Ren, K.: Dppro: Differentially private high-dimensional data release via random projection. IEEE Transactions on Information Forensics and Security **12**(12), 3081–3093 (2017)
48. Yousefpour, A., Shilov, I., Sablayrolles, A., Testuggine, D., Prasad, K., Malek, M., Nguyen, J., Ghosh, S., Bharadwaj, A., Zhao, J., et al.: Opacus: User-friendly differential privacy library in pytorch. arXiv preprint arXiv:2109.12298 (2021)
49. Yu, F., Gu, J., Li, Z., Hu, J., Kong, X., Wang, X., He, J., Qiao, Y., Dong, C.: Scaling up to excellence: Practicing model scaling for photo-realistic image restoration in the wild. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 25669–25680 (2024)
50. Zhou, X., Liang, W., Kevin, I., Wang, K., Wang, H., Yang, L.T., Jin, Q.: Deep-learning-enhanced human activity recognition for internet of healthcare things. IEEE Internet of Things Journal **7**(7), 6429–6438 (2020)