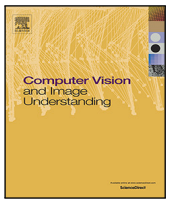
Computer Vision and Image Understanding 265 (2026) 104675

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

journal homepage: www.elsevier.com/locate/cviu



Securing workers and workspaces: Contextual privacy for vision-based ergonomics ☆

Sander De Coninck [a,*], Emilio Gamba [b], Bart Van Doninck [b], Abdellatif Bey-Temsamani [b], Thorsten Cardoen [a], Sam Leroux [a], Pieter Simoens [a]

[a] IDLab, Department of Information Technology at Ghent University – imec, Technologiepark 126, Ghent, B-9052, Belgium
[b] Flanders Make, corelab ProductionS, Gaston Geenslaan 8, Heverlee, 3001, Belgium

ARTICLE INFO

Keywords:
Contextual privacy protection
Visual privacy
Human pose estimation
Ergonomic analysis
Generative adversarial privacy
Industry 5.0

ABSTRACT

Multi-camera computer vision in industry offers advantages but poses risks to worker privacy and intellectual property through exposure of sensitive contextual information. Existing privacy methods often inadequately protect background details crucial in manufacturing. This issue is prominent in applications like automated ergonomic assessment, where visual data for posture analysis can reveal sensitive workplace information. We propose a system for simultaneous personal privacy and enhanced contextual intellectual property protection, featuring a novel probabilistic obfuscation technique. Our edge-based Generative Adversarial Privacy system employs a modified obfuscator that learns to inject controlled, pixel-wise random noise, particularly into non-critical background regions. This more effectively obscures IP-sensitive environmental details before data transmission for central analysis (e.g., pose estimation). Our approach, validated in a multi-camera ergonomic study, effectively protects worker privacy and contextual IP (metrics-evaluated) and maintains 3D pose accuracy for reliable ergonomic assessment. This work provides a solution for deploying vision systems in sensitive industrial settings by holistically addressing privacy requirements through an advanced, adaptive obfuscation strategy.

1. Introduction

The proliferation of single and multi-camera computer vision systems within industrial environments, driven by advancements in Industry 4.0 and 5.0, promises significant gains in areas such as process optimization, quality control, safety monitoring, and worker assistance (Pasquadibisceglie et al., 2022; Zhou et al., 2023). However, the deployment of pervasive visual sensing technologies introduces a critical tension: maximizing operational benefits while safeguarding sensitive information. Capturing detailed visual data inevitably raises concerns about worker privacy and the potential exposure of valuable intellectual property (IP), including proprietary machine designs, unique manufacturing workflows, or confidential component details. Navigating this challenge is further complicated by stringent data protection regulations like the GDPR (EU, 2016) and CCPA (Bukaty, 2019), which impose strict limitations on the collection and processing of personal data.

Ergonomic risk assessment in manufacturing serves as a compelling example of this dilemma. Musculoskeletal disorders (MSDs), often stemming from repetitive tasks and awkward postures, represent a major burden on worker well-being and organizational finances (Bevan, 2015; WHO et al., 2003). While vision-based systems offer a non-intrusive alternative to manual assessments or cumbersome wearable sensors (Stefana et al., 2021; Yu et al., 2019; Yan et al., 2017) for monitoring worker posture and movement, they inherently capture the very contextual and personal information that requires protection. Companies seek the benefits of automated ergonomic analysis but cannot risk violating worker trust or leaking competitive intelligence embedded within the visual data stream.

Performing full, computationally demanding state-of-the-art pose estimation directly at the edge to avoid transmitting raw video is often impractical due to the high computational cost (Zheng et al., 2023; Wang et al., 2022). Furthermore, deploying, maintaining, and updating these complex models across numerous distributed devices presents significant operational challenges (Leroux et al., 2022). Crucially, processing data centrally allows leveraging powerful, potentially proprietary or off-the-shelf, pose estimation models without needing to deploy confidential model architectures or weights onto edge devices. This not only sidesteps potential IP concerns regarding the models themselves but

☆ This article is part of a Special issue entitled: 'ACVR 2024' published in Computer Vision and Image Understanding.

* Corresponding author.
E-mail address: sander.deconinck@ugent.be (S. De Coninck).

https://doi.org/10.1016/j.cviu.2026.104675
Received 16 May 2025; Received in revised form 16 December 2025; Accepted 29 January 2026
Available online 4 February 2026

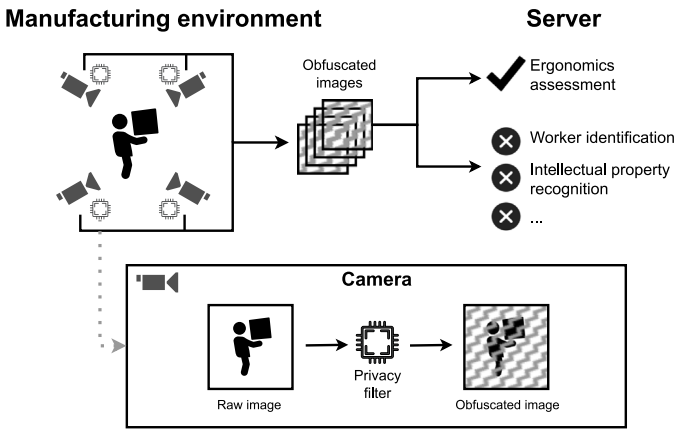


Fig. 1. Configuration of our privacy- and IP-aware system applied to ergonomics. Camera footage is transformed and obfuscated on the camera, protecting both individuals and the background context, rendering it suitable only for keypoint estimation. Obfuscated images are then transmitted for ergonomic assessment.

also significantly reduces development costs and effort associated with creating and optimizing custom models for resource-constrained edge environments. Consequently, transmitting visual data (after appropriate protection) for central analysis remains a more feasible and cost-effective strategy, reinforcing the need for robust edge-based privacy mechanisms.

Existing approaches to privacy preservation in visual data often fall short in these high-stakes industrial contexts. Simple methods like blurring detected individuals (Zhou and Pun, 2020) are prone to detection failures and can inadvertently leave sensitive information exposed. While more advanced techniques like inpainting, replacement, or synthetic abstraction (Himmi et al., 2022; Uittenbogaard et al., 2019; Shetty et al., 2018; Hukkelås and Lindseth, 2023a,c) offer better human privacy, they typically overlook the crucial need to protect sensitive background elements or contextual IP. These methods often fail to provide the comprehensive protection required when both people and proprietary environments must be secured.

To address these challenges, we propose a privacy- and IP-aware computer vision system for ergonomic assessment. Our core idea is to transform raw video data directly at the edge, potentially within a secure environment on the camera itself, before any transmission or external access. This transformation obfuscates visual identifiers (e.g., faces, skin tones) and sensitive background details, retaining only the essential information required for the downstream task, in this case, human pose estimation for ergonomic analysis. An overview of our system configuration is shown in Fig. 1.

This paper significantly extends our preliminary work presented in De Coninck et al. (2025). While the foundational concept of using Generative Adversarial Privacy (GAP) (Huang et al., 2017; De Coninck et al., 2024) for privacy in ergonomics was introduced earlier, this work makes several key advancements. We place a much stronger emphasis on contextual privacy, explicitly addressing the protection of intellectual property (IP) embedded in the background. To achieve this, we introduce an updated probabilistic obfuscation process specifically redesigned to enhance protection against the inference of sensitive environmental details. Furthermore, we employ new privacy metrics tailored to evaluate the effectiveness of this enhanced contextual protection alongside personal privacy.

We validate these advancements in a multi-camera ergonomic assessment scenario. Video streams from multiple viewpoints are independently obfuscated at the edge using our enhanced method. The privacy-protected streams are then used for 2D pose estimation, which are subsequently fused into 3D poses for calculating ergonomic risk scores. We utilize the Rapid Entire Body Assessment (REBA) (Hignett and McAtamney, 2000; Joshi and Deshpande, 2019), which provides a score from 1 to 10 that can be used to assign an ergonomic risk label. Our results demonstrate that this extended approach effectively safeguards both workers and sensitive industrial environments while maintaining the accuracy needed for reliable ergonomic analysis, evaluated using our refined metrics.

The main contributions of this paper are:

- An enhanced system architecture integrating edge-based GAP, specifically focusing on simultaneous personal privacy and robust contextual/IP protection in industrial vision applications.
- A redesigned stochastic obfuscation variant tailored to significantly improve contextual privacy compared to our prior method.
- Validation using new privacy metrics alongside utility metrics, demonstrating superior protection, particularly for background IP, on a relevant case study dataset.

The remainder of this paper is structured as follows: Section 2 describes related works. In Section 3, we introduce the architecture of our privacy-aware system. Section 4 details the data collection and ergonomic scoring methodology. Experimental results are presented in Section 5, followed by conclusions and future directions in Section 6.

2. Related works

2.1. Privacy-aware solutions for computer vision

The increasing deployment of computer vision (CV) systems necessitates methods to preserve visual privacy within image and video data. Research in this area often focuses on transforming the visual input to obscure sensitive information, such as identities, while attempting to retain utility for downstream tasks like object detection or action recognition (Zhao et al., 2025). This presents a fundamental trade-off between privacy protection and data usefulness. A primary approach involves data obfuscation, where sensitive image regions are altered. While simple methods like blurring or pixelation exist, they often degrade utility significantly and may not offer robust protection (Lee and You, 2024; Hukkelås and Lindseth, 2023b). More advanced techniques focus on inpainting or replacement. These methods typically detect sensitive entities (e.g., faces, bodies) and substitute these regions with synthetically generated content, often produced by Generative Adversarial Networks (GANs) (Shetty et al., 2018; Hukkelås and Lindseth, 2023a; Uittenbogaard et al., 2019) or diffusion models (Malm et al., 2024; Patwari et al., 2024). The goal is usually to create visually plausible outputs that preserve the overall scene context while removing identifiable features. However, the efficacy of these replacement methods hinges on accurate detection; missed detections lead to privacy leaks, while false positives can hinder utility. Furthermore, many generative anonymization techniques prioritize realistic identity removal, potentially overlooking other sensitive contextual information present in the scene (Patwari et al., 2024). To better balance privacy and utility, some methods pursue utility-aware privacy preservation. For example, Huang et al. (2025) proposed jointly optimizing an anonymization network with a specific utility network (e.g., pose estimation) to minimize task performance degradation. While potentially effective, this strategy typically requires access to and modification of the utility model's training process, hindering its use with pre-trained or proprietary black-box models. Ilic et al. (2024) has proposed human-designed obfuscation templates that selectively hide sensitive regions while preserving context and temporal consistency, achieving strong performance without retraining and improving interpretability in action recognition tasks (Dave et al., 2022).

2.2. Context-aware visual privacy

While significant attention has been paid to anonymizing primary subjects like faces, researchers have also recognized that contextual information within visual data can inadvertently leak sensitive details. Seemingly innocuous background elements or surrounding objects can reveal private information about individuals, locations, or activities. Efforts to address this have primarily focused on identifying such sensitive elements. Notable examples include datasets curated for this purpose, such as the Visual Redactions Dataset (Orekondy et al., 2018) for object redaction targets, Biv Priv Seg (Tseng et al., 2025) for segmenting private image content, and VizWiz Priv (Gurari et al., 2019), which examines privacy needs in images from visually impaired users. Further emphasizing the privacy risks inherent in visual context, recent research demonstrates that powerful Vision-Language Models can accurately infer personal attributes purely from contextual cues within images, even when the individuals themselves are not depicted (Tömekçe et al., 2024). Complementing these findings, the Multi-P2 A benchmark (Zhang et al., 2025) specifically addresses the need for comprehensive privacy assessment in Large Vision-Language Models, evaluating their awareness of sensitive input and their propensity for privacy leakage across a wide range of categories, notably extending beyond personal privacy to include trade secrets and state secrets. However, despite the development of these valuable resources for detection and assessment, there is a relative scarcity of advanced techniques specifically designed to actively mitigate privacy risks stemming from these broader contextual cues.

3. Privacy-aware ergonomics

Ergonomic monitoring plays a critical role in promoting sustainable manufacturing practices and safeguarding worker well-being. Nevertheless, conventional approaches such as continuous video surveillance raise considerable privacy concerns, and wearable sensor systems may compromise comfort and impede workers' operational efficiency. To address these limitations, we present a camera-based ergonomic assessment system that emphasizes privacy preservation without sacrificing analytical accuracy.

The proposed system employs a multi-camera configuration that captures workers from four distinct viewpoints within the manufacturing environment. To ensure privacy protection, each camera incorporates an on-device image obfuscation mechanism prior to data transmission. Specifically, the system utilizes a Generative Adversarial Privacy (GAP) framework such as proposed by De Coninck et al. (2024), as described in Section 3.1, which transforms the original video feed into a privacy-preserving representation. This transformation retains only the visual features necessary for human pose estimation, thereby preventing the extraction of extraneous personal or contextual information.

The obfuscated video frames are subsequently transmitted to a centralized processing server. There, 2D keypoints are extracted using YOLO11-Pose (Jocher et al., 2023), a state-of-the-art human pose estimation model. Keypoints from the various camera perspectives are matched and triangulated to reconstruct a three-dimensional (3D) pose of the worker using a Direct Linear Transformation (DLT) formulation solved in a least-squares sense (Hartley, 2003). The resulting 3D keypoints serve as input for the computation of the REBA score (Hignett and McAtamney, 2000), providing a quantitative measure of ergonomic risk associated with the task being performed.

REBA is one of the most commonly used ergonomic assessment tools among practitioners (Joshi and Deshpande, 2019). It evaluates the risk of a given posture based on joint angles of body segments such as the neck, trunk, legs, and upper and lower arms, and in its full form also incorporates task-level factors including repetitive movements, load handling, and force exertion. Joint angles are defined relative to the trunk; therefore, reliable trunk detection is essential for computing a REBA score, and a valid score cannot be produced if the trunk is not detected. When keypoints required to compute a specific joint angle are missing, that angle is excluded from the computation, which can lower the overall score. In the full REBA method, scores range from 1 to 15; however, because our evaluation is based on single images and does not include motion, load-handling factors, or wrist posture, the Activity Score is fixed to 0 and the total obtainable score ranges from 1 to 10. These scores can be grouped into four risk categories: no, low, medium, and high risk, as shown in Table 1. We measure accuracy both on the score itself and on these risk categories.

Table 1
REBA risk analysis based on the REBA score.

REBA score	Risk analysis
1	No risk
2–3	Low risk
4–7	Medium risk
8–10	High risk

The overall system architecture is depicted in Fig. 2. Subsequent sections elaborate on the GAP framework and introduce a novel variant that incorporates stochastic noise to further enhance contextual privacy guarantees.

3.1. Generative adversarial training

Our core privacy mechanism employs the Generative Adversarial Privacy (GAP) framework as implemented by De Coninck et al. (2024), utilizing two adversarially trained networks: an Obfuscator (O) and a Deobfuscator (D). The Obfuscator transforms an input camera frame X into an obfuscated version $X_{obf} = O(X)$. The Deobfuscator attempts to reconstruct the original image X from the obfuscated version, producing an estimate $\hat{X} = D(X_{obf})$.

The Obfuscator O is trained to minimize a composite loss function:

$$L_{obf}(X) = L_{pose}(X_{obf}) - \alpha \|X - \hat{X}\|_2^2 \tag{1}$$

where $L_{pose}(X_{obf})$ is the task loss of the downstream pose estimator (YOLOv11) applied to the obfuscated image X_{obf}, ensuring utility for ergonomic analysis. Crucially, the pose estimation network's weights are kept fixed during this adversarial training; only the Obfuscator and Deobfuscator are optimized. This ensures that the generated X_{obf} remains compatible with the pre-trained pose estimation model.

The second term, $\alpha\|X-\hat{X}\|_2^2$, represents the weighted reconstruction loss, where $\hat{X} = D(X_{obf})$. Minimizing L_{obf} encourages the Obfuscator to keep L_{pose} low (maintaining utility) while simultaneously maximizing the reconstruction error $\|X - \hat{X}\|_2^2$ (hindering reconstruction, enhancing privacy). The reconstruction error serves as a proxy for information removal. By training the Obfuscator to produce images that the Deobfuscator cannot accurately restore, we compel it to discard information from the original scene. This process is general in nature, thereby reducing the presence of sensitive details alongside other reconstructable information. The hyperparameter $\alpha > 0$ balances this trade-off.

Concurrently, the Deobfuscator D is trained to minimize its reconstruction loss, aiming to make its output $\hat{X}$ as close as possible to the original image X. This loss is implemented as a pixel-wise Mean Squared Error (MSE):

$$L_{deobf}(X, \hat{X}) = \|X - \hat{X}\|_2^2 \tag{2}$$

This adversarial process compels the Obfuscator O to learn a transformation yielding X_{obf} that preserves essential pose features while discarding other sensitive information, thereby preventing D from accurately generating $\hat{X} \approx X$.

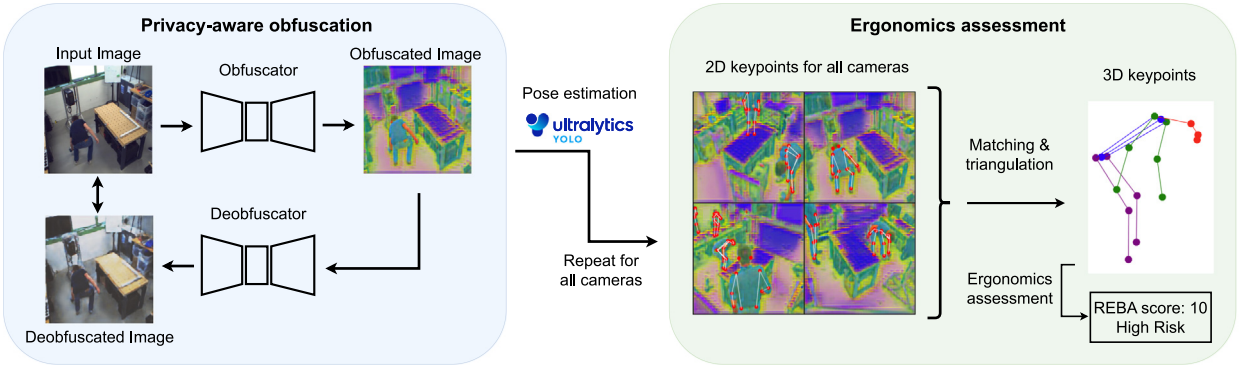


Fig. 2. Workflow of our privacy-aware ergonomics assessment system. The images are obfuscated using a Generative Adversarial Privacy scheme as described in De Coninck et al. (2024). This obfuscation process is applied to each camera, and the resulting frames are used to detect 2D keypoints with an unmodified YOLO model. After matching keypoints across views and performing triangulation, a 3D representation of the worker is constructed, enabling the calculation of a REBA score.

3.2. Probabilistic obfuscation variant for contextual privacy

While the standard GAP approach (Section 3.1) provides a strong baseline for privacy, we introduce a novel variant in this paper, specifically designed to enhance privacy protection for contextual information within the scene, such as background elements or surrounding equipment, which might inadvertently reveal sensitive operational details or location specifics. The standard approach optimizes privacy globally across the image; our variant introduces controlled, targeted randomness to potentially allow for stronger obfuscation in non-critical areas without significantly degrading pose estimation utility.

Our proposed modification extends the obfuscator network O. Instead of directly outputting the obfuscated image, the modified obfuscator outputs parameters defining a distribution for each pixel. Specifically, for an input image X, the obfuscator generates a mean tensor μ and a standard deviation tensor σ:

$$(\mu, \sigma) = O(X) \tag{3}$$

Here, μ represents the base structure of the obfuscated image (as would have been generated by the original obfuscator architecture), while σ defines a pixel-wise standard deviation, shared across all channels for a given pixel (i, j). Random noise ξ is then sampled independently for each pixel location (i, j, k), with k indicating the channel, from a Gaussian distribution governed by the learned standard deviation:

$$\text{Sample } \xi_{ijk} \sim \mathcal{N}(0, \sigma_{ij}) \quad \text{independently for each } i, j, k \tag{4}$$

The final, stochastically obfuscated image X_{obf} is then constructed by adding this sampled noise to the mean tensor and applying a sigmoid to bring the image to a $[0, 1]$ range:

$$X_{obf} = \text{sigmoid}(\mu + \xi) \tag{5}$$

This process is illustrated in Fig. 3. For implementation of the random sampling, we use the reparameterization trick (Kingma et al., 2013), which enables gradient-based optimization through stochastic nodes by expressing the sampled noise as a deterministic function of the mean and a random variable.

To encourage the network to utilize this noise injection for privacy, particularly in contextual areas, we modify the obfuscator's loss function. The objective now includes a term explicitly promoting higher variance (i.e., larger σ) where possible. The updated loss function for the obfuscator O is:

$$L_{obf}(X) = L_{pose}(X_{obf}) - \alpha \|X - \hat{X}\|_2^2 - \beta \|\xi\|_1 \tag{6}$$

where $X_{obf} = \mu + \xi$ is the noise-injected obfuscated image, and $\|\xi\|_1 = \sum_{i,j,k} |\xi_{ijk}|$ is the L1 norm of the sampled noise tensor. The new hyperparameter $\beta \geq 0$ controls the strength of the noise magnitude maximization term. Since L_{obf} is minimized, the term $-\beta\|\xi\|_1$ incentivizes the network to produce outputs (μ, σ) that lead to sampling larger magnitude noise ξ.

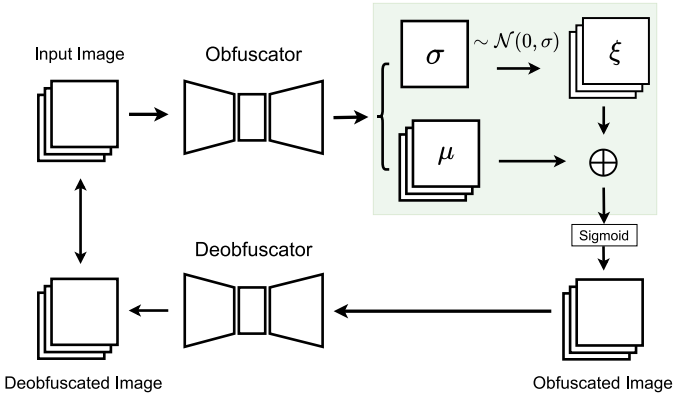


Fig. 3. Architecture of the proposed probabilistic obfuscator variant, with the expansion highlighted in green. The network outputs parameters μ and σ, noise ξ is sampled based on σ, and added to μ to produce the final obfuscated image X_{obf}.

Our hypothesis is that the combined loss function will drive the network to learn a spatially varying σ. The $L_{pose}(X_{obf})$ term penalizes noise injection that disrupts keypoint detection, likely constraining σ (and thus the expected magnitude of ξ) to be small in regions critical for pose estimation. Conversely, the $-\beta\|\xi\|_1$ term encourages larger noise magnitude. We expect this pressure to preferentially increase σ in regions where noise has less impact on L_{pose} (e.g., background), thereby enhancing privacy for contextual elements by actively maximizing the added distortion in those areas.

4. Manufacturing assembly task dataset

To evaluate our proposed approach, we collected a new dataset specifically designed to incorporate elements relevant to ergonomic assessment, personal privacy, and intellectual property protection within a simulated manufacturing assembly context. The primary objectives in creating and analyzing this dataset were threefold:

1. To capture realistic assembly actions exhibiting a range of ergonomic risk levels suitable for REBA analysis.
2. To include human operators and incidental bystanders, presenting challenges for personal privacy preservation.

3. To embed objects and environmental details representative of potential visual intellectual property (e.g., specific tools, brands, setup) that require contextual privacy protection.

Given the inclusion of identifiable participants and potentially sensitive contextual elements, this dataset was developed specifically for the internal validation presented in this study and can only be distributed upon request due to privacy and ethical considerations. This section details the data collection process and the characteristics of the resulting dataset relevant to these objectives.

4.1. Data collection setup and procedure

The core activity involved participants assembling an aluminum frame using profiles and angle brackets within a designated work cell. This cell contained a workbench (height: 86 cm), a tool cabinet, and component storage drawers at knee height.

- Ergonomic Relevance: The task was explicitly designed to include actions likely to yield medium-to-high REBA scores, such as bending to retrieve components from low drawers and reaching across the workbench (see Fig. 5).
- IP/Contextual Elements: To simulate protectable visual IP, specific identifiable tools were intentionally included and used within the workspace. Examples include a drill identifiable as Makita© brand, a Facom© toolbox, and a specific type of wrench (see examples in Fig. 4). These items were present both as static elements but were also moved around and used during the scenes, representing typical challenges for contextual privacy methods aiming to obfuscate sensitive background information.
- Participants: The recordings involved 20 unique individuals (colleagues volunteering for the study), acting as primary operators (working alone or in pairs) or as incidental bystanders appearing in the background, necessitating personal privacy considerations.

The activities were recorded using four synchronized 5MP RGB cameras[1] positioned for diverse viewpoints. A total of 15 distinct assembly sessions were captured across two different physical setups to enhance environmental diversity. Frames were extracted at 5 fps. The resulting video data was partitioned chronologically into training (11 sessions), validation (2 sessions), and test sets (2 sessions) for developing and evaluating both the privacy mechanisms and the pose estimation models. Partitioning by session ensures that each actor appears in only one set, preventing any overlap between training, validation, and test data.

For the specific purpose of evaluating IP protection techniques, the IP-sensitive objects (like the branded tools) were manually labeled with bounding boxes in a subset of 6 sessions, which were then split into 4 training, 1 validation, and 1 test session according to the overall data split.

4.2. Baseline analysis: Ergonomics and visibility

We employed a human pose estimation algorithm (Yolo11-pose) to detect 2D keypoints in the video frames from the test set, subsequently reconstructing 3D poses. These 3D keypoints served as input for a REBA analysis, providing an objective ergonomic score for observed postures (an example reconstruction and analysis is shown in Fig. 5). The distribution of REBA scores across the dataset (Fig. 6) confirms the presence of postures spanning all risk levels (no, low, medium and high risk), although low and medium risks were most prevalent. High-risk postures were captured but occurred less frequently. This baseline analysis quantifies the ergonomic characteristics we aim to assess accurately even after applying privacy transformations.

[1] Daheng MER2-503-23GC-P camera with Daheng LCM-5MP-08MM-F1.4-1.5-ND1 lens.

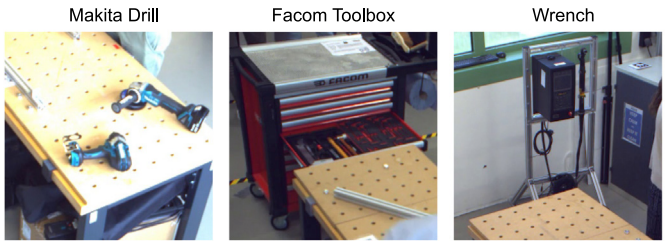


Fig. 4. Examples of objects included in the dataset to represent potentially sensitive visual intellectual property: a Makita-branded drill, a Facom toolbox, and an electric wrench.

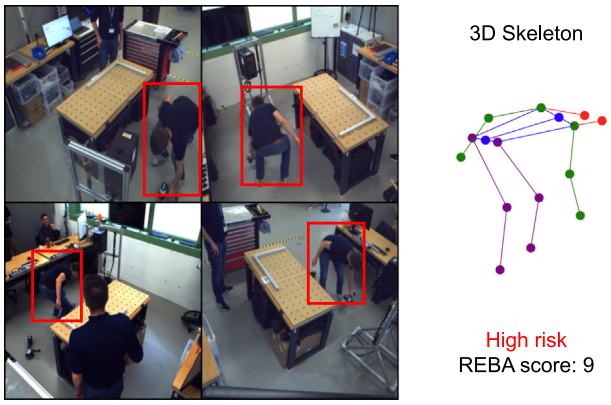
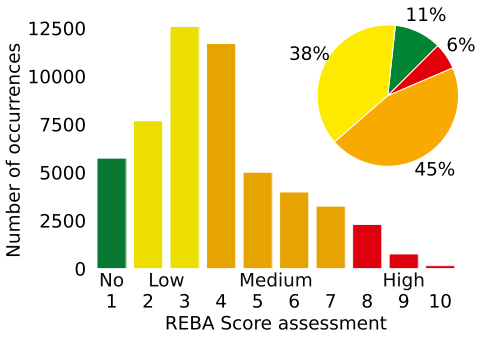


Fig. 5. Example risk situation with constructed 3D skeleton and REBA analysis from the baseline data.

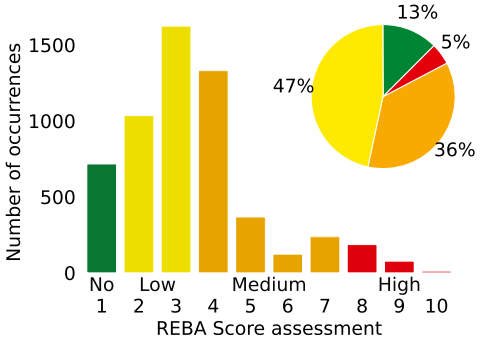
The overall REBA score is a composite measure derived from partial scores assigned to different body segments (e.g., neck, trunk, upper/lower limbs). Fig. 7 illustrates the distribution of these constituent partial scores within the training dataset. This analysis reveals segment-specific risk patterns; for instance, the neck and upper limbs (particularly involving elbow position) frequently contribute higher partial scores, indicative of more demanding postures, compared to the lower body segments, which predominantly exhibit lower-risk scores. This detailed breakdown helps identify which body parts are most often subjected to ergonomic strain during the observed tasks.

Furthermore, understanding the inherent limitations of the multi-camera setup itself is crucial for interpreting results. The multi-camera configuration inherently leads to variable keypoint visibility due to occlusions and differing viewing angles. Fig. 8 illustrates this by showing the average number of cameras detecting each specific anatomical keypoint. Notably, distal keypoints (e.g., head, hands) are detected by fewer cameras on average compared to central torso keypoints.

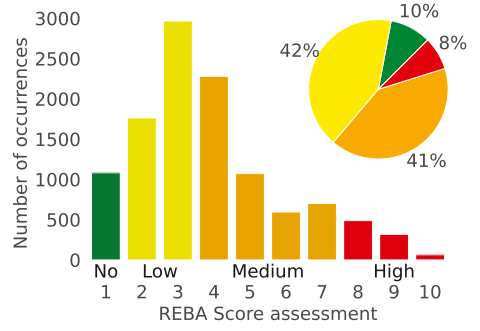
Complementing this view, Fig. 9 presents the distribution of the total number of detected keypoints per person instance across simultaneous camera views (1 to 4). Quantitative analysis of this distribution reveals significant limitations in comprehensive visibility: on average, only 3.04 keypoints per person are visible across all four cameras concurrently. The relatively high standard deviation of 3.93 can be explained by the fact that individuals are often not visible in all four views simultaneously. Even considering keypoints visible in at least two cameras, the average increases to only 7.32. This finding underscores that the 3D pose reconstructions used for REBA scoring in this setup are frequently derived from partial keypoint data, representing a baseline challenge independent of any privacy obfuscation.



(a) Train.



(b) Validation.



(c) Test.

Fig. 6. Overview of baseline REBA score distributions in the training, validation, and test sets, confirming coverage of different ergonomic risk levels.

4.3. Limitations

It is important to note the limitations stemming from our dataset's scope. Participants were colleagues who volunteered for this study, and the data was collected in a controlled, simulated environment. Environmental variation was introduced through two distinct locations with different layouts and lighting conditions; however, the overall scale of the study (20 participants) remains relatively limited. Evaluating the proposed approach on larger and more diverse datasets, particularly those collected in live industrial setting, would be a valuable next step. Furthermore, this study focuses on postural risk (REBA) and does not analyze factors like repetition, force, or load handling, thus limiting the maximum possible REBA score observed (scores above 10 are not possible without load/coupling factors). The primary aim was to create a dataset suitable for evaluating privacy techniques against a standard postural analysis method while including representative personal and contextual privacy challenges.

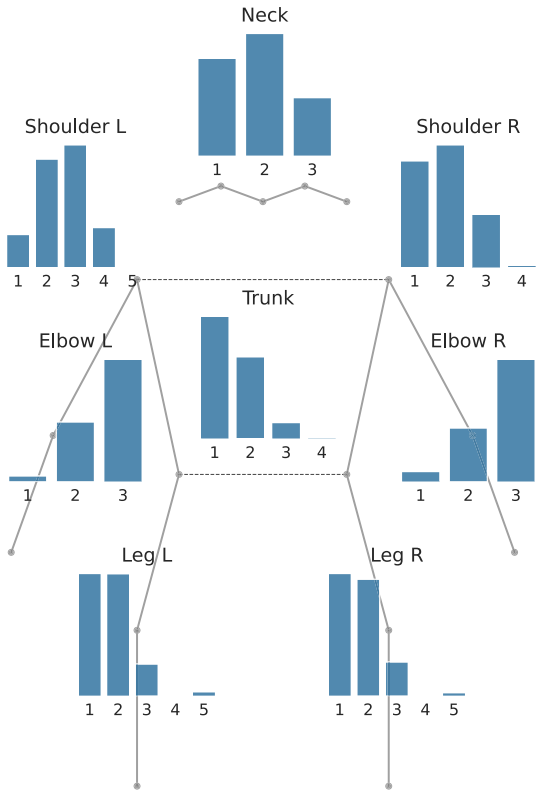


Fig. 7. Distribution of baseline partial REBA scores by body segment for the training dataset. Note that the ranges of values can differ per segment.

5. Experimental evaluation

5.1. Evaluation metrics

We evaluate our approach by assessing its effectiveness in enabling ergonomic evaluations while also measuring its ability to protect both personal privacy and IP-sensitive aspects.

5.1.1. Ergonomic evaluation

We measure the capability to do ergonomic evaluations through keypoint estimation metrics and REBA calculation capacity. In human keypoint estimation tasks, Average Precision (AP) is commonly used to evaluate the performance of a model. The AP is calculated by measuring the overlap between the predicted and ground truth keypoints using the Object Keypoint Similarity (OKS) (Lin et al., 2015) metric. The OKS serves a role similar to the Intersection over Union (IoU) used in object detection. For each object, ground truth keypoints are given as $[x_1, y_1, v_1, \ldots, x_k, y_k, v_k]$, where (x_i, y_i) are the keypoint coordinates, and v_i is a visibility flag. The person's scale s is defined as the square root of the person's segment area. Predicted keypoints have the same format but do not require visibility prediction.

The Object Keypoint Similarity (OKS) is calculated as:

$$\text{OKS} = \frac{\sum_i \left[\exp\left(-\frac{d_i^2}{2s^2\kappa_i^2}\right)\delta(v_i > 0)\right]}{\sum_i \delta(v_i > 0)} \tag{7}$$

where d_i is the Euclidean distance between the predicted and ground truth keypoints, κ_i is a keypoint-specific scale factor, and $\delta(v_i > 0)$ is an indicator function that includes only visible or labeled keypoints in the calculation.

In addition to keypoint estimation accuracy, we evaluate the system's performance using the Rapid Entire Body Assessment (REBA) score and risk level assessment. For our evaluation, we measure the

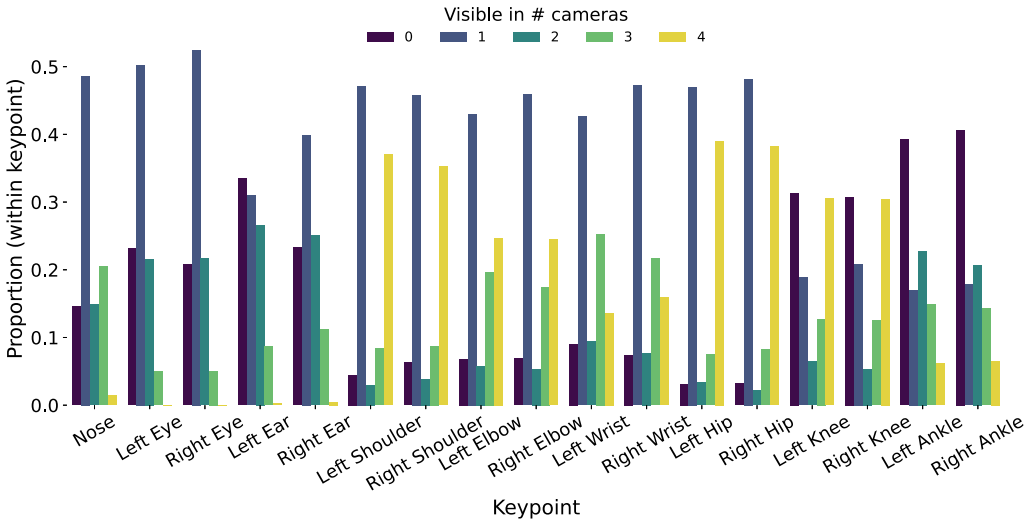


Fig. 8. Average number of cameras (out of 4) simultaneously detecting each anatomical keypoint across the dataset. Distal keypoints (e.g., head, hands) show lower average visibility compared to central torso keypoints.

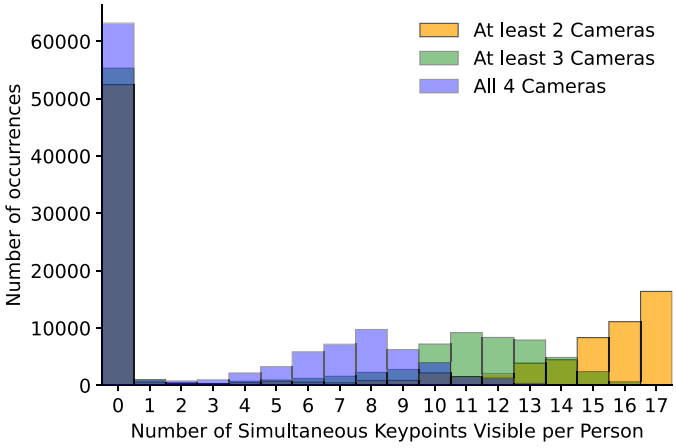


Fig. 9. Distribution of the total number of detected keypoints per person instance observed simultaneously by two, three, or all four camera views.

accuracy of the REBA score itself and the correctness of the resulting risk category classification.

5.1.2. Privacy evaluation

To quantitatively evaluate the extent of privacy preservation, we adopt PerceptAnon (Patwari et al., 2024), a metric explicitly designed to align with human perceptions of anonymity. Traditional image quality measures such as SSIM (Wang et al., 2004), PSNR, and VIF (Sheikh and Bovik, 2006), as well as perceptual similarity metrics like LPIPS (Zhang et al., 2018) and SemSim, a privacy metric focused on reconstruction attacks (Sun et al., 2024), often fail to capture how individuals assess privacy, particularly when contextual information is present (Sun et al., 2024; Patwari et al., 2024).

PerceptAnon addresses this limitation by being trained directly on human-labeled data reflecting perceived anonymity. It evaluates anonymization with respect to both identifiable personal features and contextual cues in the scene. The framework provides two key evaluation modes:

- **HA1**: Measures the absolute level of perceived anonymity for a given image.
- **HA2**: Assesses the relative change in perceived anonymity between an obfuscated image and its original counterpart.

Both metrics produce scores between 0 and 10, with 10 indicating the highest privacy protection. In our work, we primarily focus on the HA2 metric. This choice allows for a more direct comparison of the change in perceived anonymity across different techniques, including those like DeepPrivacy2 that aim for visual realism in the anonymized output, as HA1 (viewing the anonymized image in isolation) might not fully capture the anonymization effect for such methods without the original context.

We evaluate the system's ability to obscure sensitive contextual information, specifically focusing on the ease of automated detection of IP-sensitive objects (e.g., branded tools) present in our dataset (Section 4). While we recognize that machine-based object detection is not equivalent to human visual perception, we use its performance as a robust proxy to quantify the amount of identifiable information remaining after obfuscation. A significant reduction in detection accuracy, particularly against an adapted adversary, serves as a strong indicator of the difficulty in reliably identifying specific IP. We evaluate this under two adversarial scenarios:

Scenario 1: Generic adversary. An object detector D_{orig}, trained solely on original images $\mathcal{D}_{orig} = \{(X_i, Y_i^{ip})\}$, is tested directly on obfuscated images X_{obf}^{test}. This evaluates if obfuscation prevents zero-shot transfer from unobfuscated training data.

Scenario 2: Informed adversary. A detector D_{obf} is trained or fine-tuned on obfuscated image-label pairs $\mathcal{D}_{obf} = \{(X_{obf,i}, Y_i^{ip})\}$ and tested on X_{obf}^{test}. This assesses the difficulty of detecting objects even after adapting the detector to the obfuscated domain. If significant identifying features remained, even in a distorted form, a model re-trained with ground-truth labels would learn to exploit them. Therefore, poor performance in this scenario indicates a fundamental removal of identifying information.

Detection performance is measured using mean Average Precision (mAP) at an IoU threshold of 0.50 (mAP@0.50) and mAP averaged for IoU thresholds from 0.50 to 0.95 (mAP from 0.50 to 0.95), along with per-class Average Precision (AP). Significant reductions in these metrics for obfuscated images compared to originals indicate effective IP information hiding.

5.2. Experiment setup and baselines

We train all models on our own dataset which we labeled automatically using YOLO, more specifically the Yolov11x-pose which we ran on all frames after resizing to 640 × 640. We trained for 20 epochs with

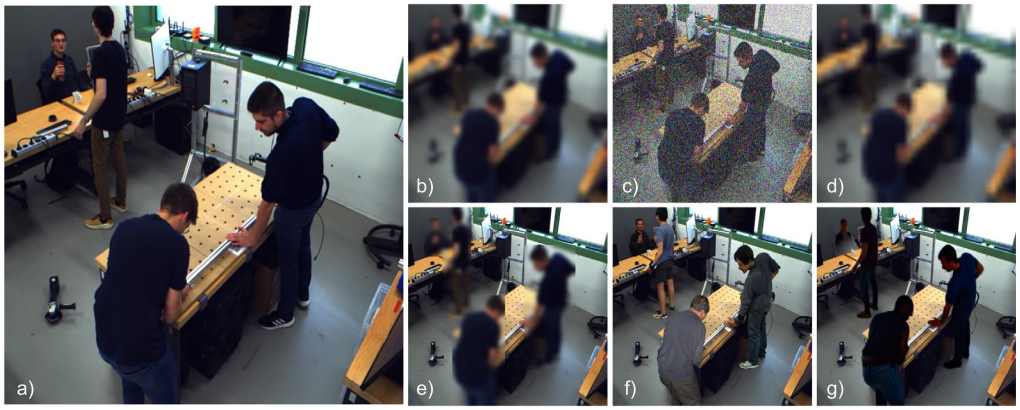


Fig. 10. Visualization of the baseline obfuscation techniques. From left to right, top to bottom: (a) original, (b) gaussian blur ($k = 51$), (c) additive gaussian noise ($\sigma = 0.05$), (d) pixelation ($f = 16$), (e) two-step blur ($k = 41$), (f) DeepPrivacy2, (g) RAD.

batch size 8. Both the obfuscator and deobfuscator are optimized using the AdamW optimizer (Loshchilov and Hutter, 2019) with an initial learning rate of 1×10^{-3}, which is decreased by a factor of 10 every 10 epochs. The training was conducted on a Tesla V100 GPU with automatic mixed precision to enhance efficiency. After training, the deobfuscator is discarded, and only the obfuscator is used for inference. We use the ultralytics yolov8 data augmentations, which includes a mixture of mosaic, mixup (Zhang et al., 2017) and cutmix (Yun et al., 2019) transformations as well as random color augmentations. We use $\alpha = 20$ when training the base obfuscator, and $\alpha = 20, \beta = 100$ when training the stochastic obfuscator.

Baselines. We compare our base GAP (Section 3.1) and stochastic obfuscator (Section 3.2) against several baselines using the metrics from Section 5.1. These include:

- **Fixed Methods:** Applied uniformly across the image to evaluate performance across different utility-privacy trade-offs.
  - Gaussian Blurring: Kernel sizes $k \in \{21, 35, 51, 75\}$.
  - Additive Gaussian Noise: Standard deviations $\sigma \in \{0.01, 0.02, 0.03, 0.05\}$.
  - Pixelation: Downscale factors $f \in \{2, 4, 8, 16\}$.
- **Learning-based Methods:**
  - Two-Step Detection + Blurring: Employs a pre-trained object detector (YOLOv11) followed by Gaussian blurring (kernel $k \in \{11, 21, 31, 41\}$) applied only within detected person bounding boxes.
  - DeepPrivacy2 (Hukkelås and Lindseth, 2023a): A GAN-based person anonymization technique, using the publicly available pre-trained model.
  - RAD (Malm et al., 2024): A diffusion-based person anonymization technique, using its public pre-trained model. Owing to its computational intensity, RAD's evaluation was conducted on a reduced dataset of 260 images (1/50th of the total test images).

This selection provides comparison points against both simple transformations at various strengths and more sophisticated privacy-enhancing techniques. A visualization of these baseline techniques to an image can be seen in Fig. 10. The baselines play to different strengths, while DeepPrivacy2 generates the most realistic and privacy-preserving images for humans, the Gaussian blur, pixelation and noise adding provide contextual privacy protection as well.

5.3. Visual analysis of obfuscated frames

We visually assess the obfuscation techniques in Fig. 11, comparing the original GAP obfuscator (Section 3.1) and our stochastic variant (Section 3.2). Consistent with the adversarial training objectives, Fig. 11(a) shows that individuals remain recognizable as human figures for pose estimation utility, while distorted features hinder identification, enhancing privacy. The impact on context differs: our stochastic variant (Section 3.2) introduces more pronounced alterations to non-human elements (e.g., tools in Fig. 11(a) become almost unrecognizable), aligning with its specific design goal of enhancing contextual privacy compared to the original method.

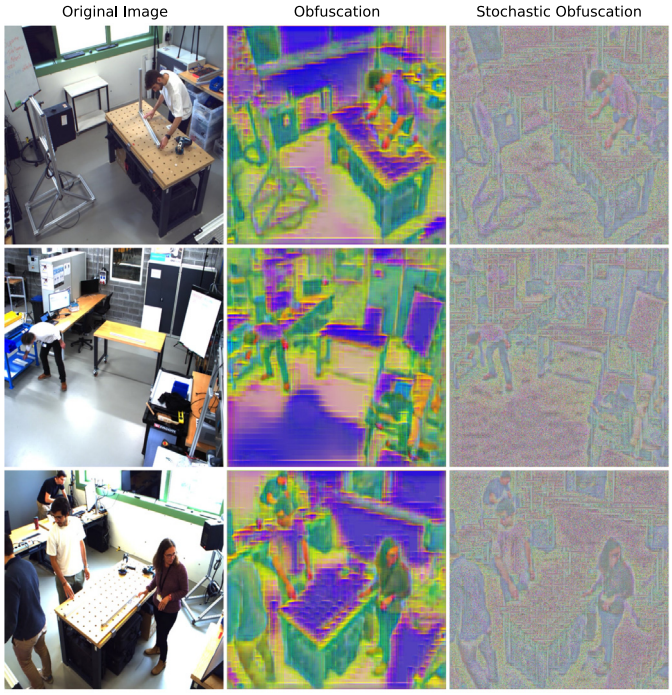
Visualizing the sampled noise maps ξ (Fig. 11(b), brighter areas indicate higher applied noise magnitude) offers insight into the stochastic variant's mechanism. These maps confirm our hypothesis by showing significantly more noise concentrated in background regions than on workers. This demonstrates that the obfuscator, guided by the loss function in Eq. (6), successfully learns to apply noise differentially, preserving keypoint-critical regions while maximizing distortion elsewhere for contextual privacy. Although effective for separating foreground subjects, the resulting spatial noise pattern sometimes resembles a general edge detector more than precise object boundaries.

5.4. Comparative evaluation of privacy-utility trade-off

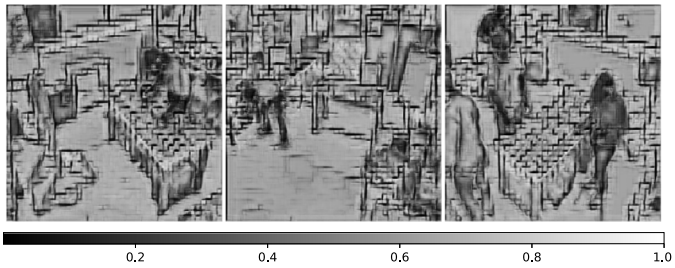
We now empirically evaluate the trade-off between preserving contextual privacy and maintaining utility for ergonomic assessment. Leveraging the metrics defined previously (Section 5.1.1), we plot the person keypoint estimation performance (mAP50, indicating utility for ergonomics) against the PerceptAnon HA2 score (measuring perceptual difference as a proxy for contextual privacy) on the test set. Fig. 12 presents this comparison for our proposed methods: the baseline GAP Obfuscator: (Section 3.1) and the Stochastic Obfuscator (Section 3.2) – alongside several baseline techniques (Gaussian Blur, Additive Noise, Pixelation at varying intensities), a Two-Step blurring obfuscator, and the advanced DeepPrivacy2 and RAD methods.

The results clearly highlight the effectiveness of our Stochastic Obfuscator. As shown in Fig. 12, it achieves by far the highest PerceptAnon HA2 score (HA2 $\sim$ 8.0), indicating a significantly greater perceptual difference from the original images compared to all other methods, while simultaneously maintaining strong utility with an mAP50 score of approximately 0.78. This demonstrates its capability to substantially enhance contextual privacy, aligning with the design goals outlined in Section 3.2, without crippling the downstream ergonomic analysis task.

Our Base Obfuscator provides a moderate balance, yielding a mAP50 of $\sim$0.62 and an HA2 score of $\sim$4.9. While traditional methods like Gaussian Blur can achieve slightly higher HA2 scores (up to $\sim$5.4),



(a) Example obfuscations using the original obfuscator architecture, and the stochastic obfuscation variant.



(b) Pixel maps for the stochastic obfuscator, indicating regions where noise is preferentially added. Pixels at value 1 indicate maximal noise addition.

Fig. 11. Visual results of the obfuscation process.

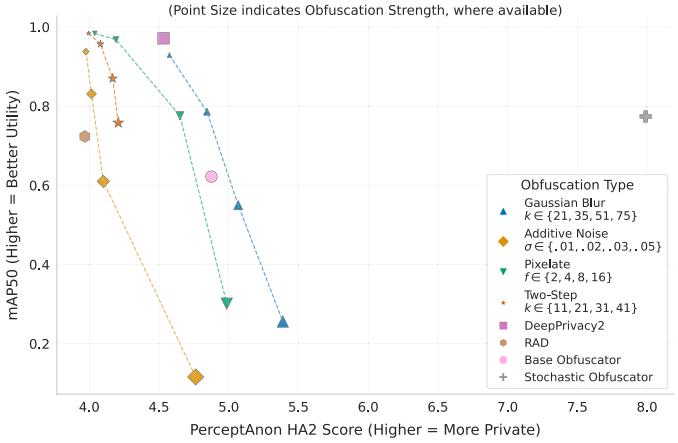


Fig. 12. Comparison of the privacy-utility trade-off. Person mAP50 (Utility for ergonomics) vs. PerceptAnon HA2 (Contextual privacy proxy). Higher values are better for both axes. Our Stochastic Obfuscator (gray cross) demonstrates a significantly enhanced privacy level (HA2 $\sim$ 8.0) while maintaining high utility (mAP50 $\sim$ 0.78), outperforming other methods in achieving this balance.

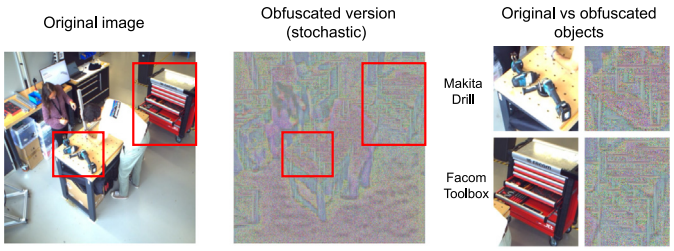


Fig. 13. An example highlighting the obfuscation of IP-sensitive objects. After obfuscation, items are difficult to discern.

this only occurs at the expense of drastically reduced utility (mAP50 $\sim$ 0.25). At comparable utility levels (mAP50 $\sim$0.6), the Base Obfuscator offers slightly better or similar HA2 performance compared to the baselines (Noise, Pixelate, Blur, Two-Step), which generally exhibit a less favorable privacy-utility curve.

Among the advanced learning-based methods, DeepPrivacy2 excels in preserving utility, showing minimal drop in mAP50 ($\sim$0.98). However, its corresponding HA2 score ($\sim$4.7) is relatively low, falling below both our Base and Stochastic Obfuscators. Similarly, the RAD method (brown hexagon) achieves a utility of mAP50 $\sim$0.74 but with an even lower HA2 score of $\sim$4.0. These findings, particularly for methods like DeepPrivacy2 and RAD which focus on individual anonymization, highlight that they are less effective at broader contextual privacy compared to our Stochastic Obfuscator. This underscores the distinct advantage of our Stochastic approach for applications demanding strong contextual privacy in addition to task utility. Overall, the Stochastic Obfuscator demonstrates a superior position on the privacy-utility spectrum for the target application of privacy-aware ergonomic monitoring.

5.5. Detection accuracy of IP sensitive aspects

To test the ability of the anonymization methods to mask IP-sensitive aspects (e.g., tool brands), we performed two experiments evaluating object detection performance against both a generic adversary (detector trained only on original images, denoted 'Orig') and an informed adversary (detector trained or fine-tuned on anonymized images, denoted 'Anon'). The results, presented as mean Average Precision (mAP50), can be seen in Table 2.

The findings highlight the effectiveness of the Stochastic Obfuscator in hindering the detection of IP-sensitive items. Against the generic adversary ('Orig' columns), the Stochastic Obfuscator provides excellent protection, drastically reducing the overall mAP50 to 0.01. This level of protection is comparable to the most aggressive blurring, noising, and the Base Obfuscator, likely benefiting significantly from the domain shift between original and obfuscated data, as evidenced by the generally low 'Orig' mAP scores for most obfuscation techniques.

Against the more challenging informed adversary ('Anon' columns), which mitigates the data shift effect, the Stochastic Obfuscator still demonstrates strong performance. It achieves the second-lowest overall mAP50 (0.67), surpassed only by heavy pixelation ($f = 16, mAP50 = 0.58$). This indicates that the Stochastic Obfuscator removes significant task-relevant information pertaining to contextual IP items, likely due to its mechanism of injecting noise preferentially into background regions (as discussed in Section 3.2). However, we note that even with substantial obfuscation, objects sometimes remain discernible, potentially influenced by the limited size and the stationary nature of objects within our dataset. The Facom toolbox, for example, has the highest mAP of the objects, and is the largest, most stationary item. An example of our obfuscation applied on the objects can be seen in Fig. 13.

Table 2
IP Object detection performance (mAP50, mAP50-95) and pose estimation performance when applying different potential privacy-preserving techniques compared to the baseline (Original). The 'Anon' indicates an informed adversary, which is trained on obfuscated data, whereas 'Orig' indicates a generic adversary trained only on the original data. Metrics are evaluated overall and on individual object classes. Bold entries highlight the method resulting in the lowest mAP score for each column, except for the last, where it indicates the highest value.

Metric	mAP50-95		mAP50		FacomToolbox		MakitaDrill		Wrench		Pose mAP50
Scenario	Anon	Orig	Anon	Orig	Anon	Orig	Anon	Orig	Anon	Orig	
Blurred											
k = 21	0.56	0.34	0.80	0.49	0.69	0.47	0.49	0.46	0.49	0.09	0.93
k = 35	0.53	0.14	0.81	0.22	0.71	0.27	0.44	0.16	0.46	**0.00**	0.79
k = 51	0.49	0.04	0.77	0.09	0.61	0.13	0.47	**0.00**	0.39	**0.00**	0.55
k = 75	0.46	0.02	0.71	0.06	0.67	0.05	0.43	**0.00**	0.27	**0.00**	0.26
Noised											
σ = .10	0.63	0.31	0.87	0.50	0.67	0.46	0.70	0.13	0.51	0.34	0.94
σ = .20	0.59	0.04	0.80	0.07	0.73	0.02	0.58	**0.00**	0.46	0.09	0.83
σ = .30	0.56	0.02	0.80	0.03	0.72	**0.01**	0.55	**0.00**	0.41	0.04	0.61
σ = .50	0.53	**0.00**	0.77	**0.01**	0.70	**0.01**	0.49	**0.00**	0.39	**0.00**	0.12
Pixelated											
f = 2	0.71	0.68	0.91	0.88	0.78	0.80	0.74	0.72	0.60	0.53	**0.98**
f = 4	0.59	0.61	0.83	0.86	0.73	0.74	0.52	0.62	0.50	0.46	0.96
f = 8	0.46	0.44	0.76	0.69	0.70	0.64	0.40	0.39	0.29	0.29	0.78
f = 16	**0.31**	0.13	**0.58**	0.29	0.64	0.29	**0.14**	0.03	**0.15**	0.07	0.30
Stochastic obfuscator	0.44	**0.00**	0.67	**0.01**	**0.58**	**0.01**	0.44	**0.00**	0.29	**0.00**	0.78
Base obfuscator	0.53	0.01	0.75	0.02	0.69	**0.01**	0.53	0.01	0.36	**0.00**	0.62
Two-Step	0.58	0.63	0.83	0.82	0.67	0.75	0.61	0.63	0.47	0.51	0.87
DeepPrivacy2	0.66	0.68	0.87	0.86	0.79	0.77	0.60	0.74	0.58	0.53	0.97
Original (Ref.)	–	0.71	–	0.89	–	0.80	–	0.78	–	0.54	–

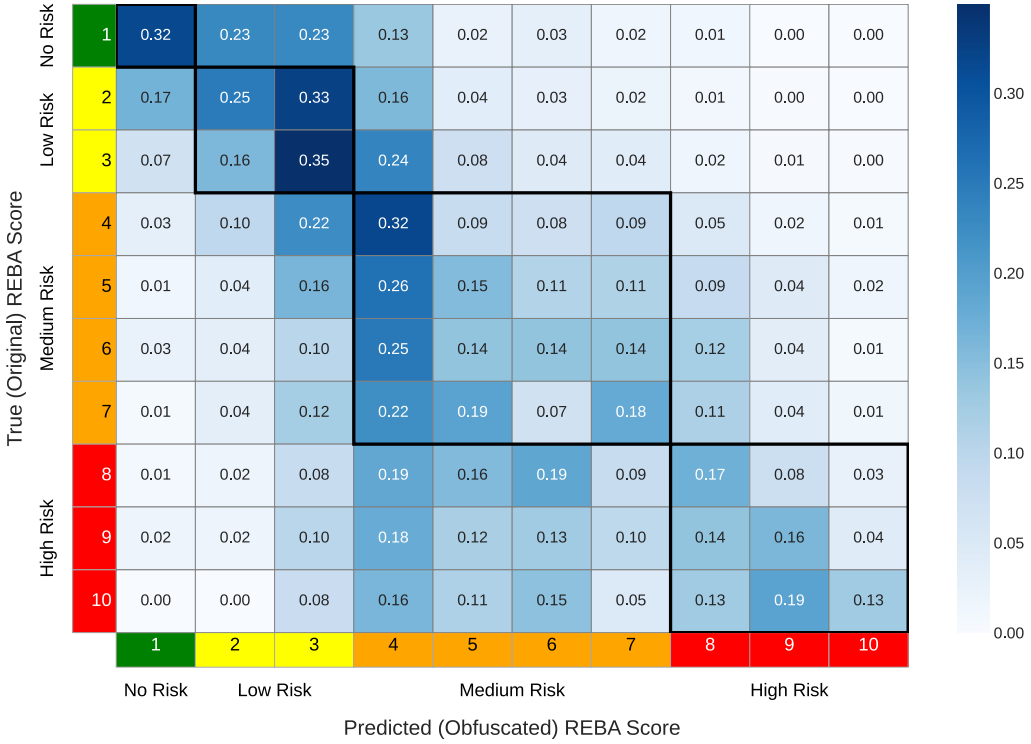


Fig. 14. Comparison of REBA scores and risk categories calculated on original and obfuscated data.

Conversely, techniques applying obfuscation only to people (Two-Step Obfuscator and DeepPrivacy2[2]) expectedly offer minimal protection for these contextual IP items, yielding high mAP scores (0.83 and 0.87 respectively) even against the generic adversary, as they do not alter the regions containing the tools. Overall, these experiments suggest the Stochastic Obfuscator effectively balances utility (as shown in Section 5.4) with robust protection against the detection of sensitive contextual objects, presenting a strong alternative to overly aggressive techniques or those neglecting contextual privacy.

5.6. REBA assessment on obfuscated videos

Finally, we evaluate the performance of ergonomic assessments when conducted on obfuscated video streams. While keypoint estimation accuracy provides an initial indicator, REBA score calculation relies on precise 3D keypoint reconstruction from multiple camera

[2] We did not perform this experiment with RAD due to the high computational cost of performing anonymization on all IP-labeled data.

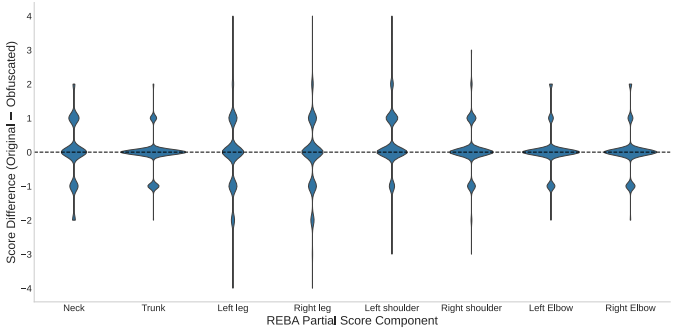


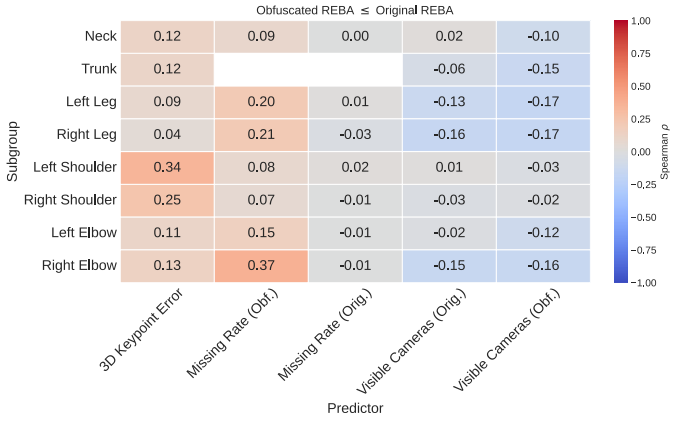
Fig. 15. Difference between original and obfuscated partial scores. A negative difference indicates the obfuscated score was higher than the original.

angles. Even minor deviations in keypoint locations, potentially introduced by the obfuscation process, can influence the final 3D skeleton and, consequently, the REBA score. Conversely, the multi-view setup can sometimes offer resilience if a keypoint is poorly detected in a single view. Fig. 14 presents a confusion matrix comparing the REBA scores and their corresponding risk categories (No, Low, Medium and High Risk) derived from original and obfuscated video data. It is important to note that in 5% of cases, a valid REBA assessment could not be completed from the obfuscated frames due to a person not being identified on 2 cameras after obfuscation; the results presented here pertain to the 95% of instances where analysis was successful. The comparison reveals a strong correspondence: 27% of obfuscated REBA scores exactly matched the original scores, and 53% fell within the same broader risk category. Overall, the Mean Absolute Difference (MAD) between original and obfuscated REBA scores was 1.46 ± 1.4, whereas the median was 1. Assessments based on obfuscated data tend to be most accurate for lower-risk scenarios. Higher-risk scenarios pose more challenges, exhibiting larger discrepancies between original and obfuscated REBA scores. This is likely attributable to the lower prevalence of high-risk postures in our training dataset (as detailed in Section 4), meaning the obfuscator had fewer such examples to learn from, potentially impacting its ability to precisely preserve keypoint information critical for these less common, more extreme postures.

To further investigate these differences, Fig. 15 illustrates the distribution of differences (Original Score — Obfuscated Score) for the constituent partial REBA scores. For many body components, the difference is frequently zero, indicating no change in the partial score after obfuscation. However, the 'Neck' component notably shows a wider distribution of differences and a higher likelihood of score changes. This increased sensitivity for the neck may be due to a combination of factors. Firstly, as REBA requires fine-grained distinctions for neck posture assessment, small variations in keypoint locations can map to different partial scores. Secondly, the baseline analysis in Section 4 (Fig. 7) showed that neck postures frequently contribute to higher risk scores in the original data, potentially placing them near scoring thresholds more often. Furthermore, related keypoints such as those on the head can exhibit lower average camera visibility (Fig. 8), meaning their 3D reconstruction might be more sensitive to any noise introduced by the obfuscation process, thereby impacting neck angle calculation. Additionally, the negative skew in the difference for several components (e.g., Neck, Left leg, Right leg, Left shoulder) indicates that obfuscated partial scores tend to be slightly higher on average than the original scores for these body parts (since a negative Original — Obfuscated value implies Obfuscated Score > Original Score). This suggests the obfuscation might, in some cases, lead to a more conservative (higher) risk assessment for these specific segments.

To investigate the causes of observed differences in REBA subscores, we examine several potentially influential factors. The primary source of error is obfuscation-induced degradation, quantified by the

(a) Potential overestimation (Obfuscated ≥ original).



(b) Potential underestimation (Obfuscated ≤ original).

Fig. 16. Spearman's ρ correlation between individual error factors and the absolute difference in REBA sub-scores ($|\text{REBA}_{\text{obf}} - \text{REBA}_{\text{orig}}|$), stratified by cases where obfuscation increases (a) or decreases (b) estimated risk. Overestimation correlates more with 3D spatial error, underestimation with missing keypoints. Correlations are generally low ($|\rho| < 0.4$), suggesting discrepancies arise from multiple factors. Trunk sub-scores require valid detection of all trunk keypoints; if any are missing, no REBA score can be computed, so missing-rate statistics and corresponding score differences are omitted.

average 3D keypoint error between original and obfuscated frames, as well as the rate of missing keypoints in the obfuscated output and the amount of obfuscated camera view a person is detected in. We also consider pre-existing challenges, such as the number of cameras a person was detected in and occlusions in the source footage, measured by the missing keypoint rate in the original frames.

For all factors, we compute Spearman's ρ correlation with the difference in REBA sub-scores. We apply the Benjamini–Hochberg procedure (Benjamini and Hochberg, 1995) to control the false discovery rate. To capture the distinct mechanisms driving score deviations, we evaluate correlations separately for potential underestimation (Obfuscated Score ≤ Original) and overestimation (Obfuscated Score ≥ Original). The results are presented in Fig. 16.

The analysis reveals a distinct asymmetry in the sources of error. In scenarios involving potential overestimation (Fig. 16(a)), the error is driven primarily by spatial misalignment rather than detection failure. Here, missing keypoints have negligible influence, while 3D keypoint error shows the strongest consistent correlations, particularly for the Neck ($\rho = 0.31$) and Left Leg ($\rho = 0.30$). Conversely, scenarios involving potential underestimation (Fig. 16(b)) exhibit moderate positive correlations with the rate of missing keypoints in the obfuscated data, most

Fig. 17. Visualization of significant discrepancies between original and obfuscated REBA predictions. Each row corresponds to a single subject assessment. The obfuscated model predictions are overlaid on the original images for comparison.

notably for the Right Elbow ($\rho = 0.37$) and Right Leg ($\rho = 0.21$). This indicates that when the obfuscator fails to detect a body part, extreme postures are often unregistered, resulting in a lower risk score. It is important to note, however, that the overall correlation magnitudes remain generally low ($|\rho| < 0.4$). This suggests that while these factors are the primary respective drivers, significant risk assessment errors likely result from the compound effect of multiple interacting factors rather than any single source of degradation.

To contextualize these statistical patterns, Fig. 17 presents a visual analysis of three representative high-error scenarios, where each row depicts the assessment for a single subject. The top row illustrates a severe underestimation (ΔREBA $= -8$) involving an unusual, deep forward bend. Consistent with the correlations observed in Fig. 16(b), this error is driven by compounded visibility issues: the subject is undetected in the obfuscated output (Camera 1) or even the original footage (Camera 3), while the remaining view (Camera 2) suffers from severe keypoint misalignment. The middle row demonstrates a case of overestimation (ΔREBA $= +7$). Here, the obfuscated predictions (Cameras 1, 2 and 4) are not missing but are spatially scattered or "jittery". This misalignment creates artificial acute angles, directly reflecting the relationship between 3D spatial error and score inflation identified in Fig. 16(a). Finally, the bottom row depicts a complex underestimation (ΔREBA $= -7$) involving partial occlusion. The error here arises from a difficult viewing angle compounded by obfuscation noise, reinforcing our conclusion that significant discrepancies are often the result of compound factors rather than a single isolated variable.

6. Conclusion and future work

This paper addressed the critical challenge of deploying computer vision for industrial ergonomic assessment while safeguarding not only personal privacy but also sensitive contextual information and intellectual property (IP). We introduced a privacy-aware system centered on edge-based Generative Adversarial Privacy (GAP), featuring a novel stochastic obfuscation variant specifically engineered to enhance contextual privacy. This was achieved through an adaptive, pixel-wise noise injection strategy, designed to more effectively obscure IP-sensitive environmental details and background elements before data transmission for central analysis.

Experimental evaluations on a new dataset demonstrated our stochastic obfuscator's superior privacy-utility trade-off (PerceptAnon HA2 $\sim$ 8.0, keypoint mAP50 $\sim$ 0.78), outperforming baselines in comprehensive scene privacy. It also effectively obscured IP-sensitive objects against both generic (mAP50 $\sim$ 0.01) and informed (mAP50 $\sim$ 0.67) adversaries. Crucially, REBA ergonomic assessments on obfuscated data remained largely viable, with an average score difference of 1.46 ± 1.4 (median 1) and most assessments falling within the same risk category, despite some partial score sensitivities and a 5% analysis failure rate. This demonstrates a practical pathway for automated ergonomic monitoring that holistically addresses the dual requirements of personal and contextual data protection.

Future works include the implementation of our technique in a secure environment on the cameras themselves, such as within a trusted execution environment, and expanding the dataset using data augmentation to create more cases of persons in high-risk ergonomic situations.

Limitations include the controlled dataset, the scope of REBA factors analyzed, and the specific nature of IP-sensitive objects. Future work will focus on dataset expansion (particularly increasing the diversity of participants and environments), investigating REBA assessment failures, refining the stochastic obfuscation (e.g., with semantic understanding or adaptive strength), addressing motion-based privacy leakage (such as preventing object identification during movement), and validating the system in real-world industrial settings.

In conclusion, our work offers a contribution towards intelligent ergonomic assessment tools that consider the protection of both individual privacy and corporate contextual information. By specifically targeting the obfuscation of sensitive background details as well as personal identifiers, this research aims to support safer, more sustainable, and confidentiality-conscious manufacturing practices within the Industry 5.0 framework.

CRediT authorship contribution statement

Sander De Coninck: Writing – original draft, Visualization, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Emilio Gamba:** Writing – review & editing, Software, Resources, Investigation, Data curation, Conceptualization. **Bart Van Doninck:** Writing – review & editing, Validation, Supervision, Resources, Data curation, Conceptualization. **Abdellatif Bey-Temsamani:** Writing – review & editing, Validation, Resources, Funding acquisition. **Thorsten Cardoen:** Writing – review & editing, Software, Resources. **Sam Leroux:** Writing – review & editing, Supervision, Methodology, Formal analysis. **Pieter Simoens:** Writing – review & editing, Supervision, Project administration, Methodology, Formal analysis.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used Google AI Studio and ChatGPT to improve the readability of the text and aid in the design of results figures. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the published article.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Sander De Coninck receives funding from the Special Research Fund of Ghent University under grant no. BOF22/DOC/093. This research is done in the framework of the Flanders AI Research Program (https://www.flandersairesearch.be/en) financed by EWI (Economie Wetenschap & Innovatie), and Flanders Make (https://www.flandersmake.be/en), the strategic research Centre for the Manufacturing Industry who owns the Operator 4.0/5.0 infrastructure.

Data availability

The original recordings cannot be shared publicly to protect participant privacy. The anonymized data will be made available on request.

References

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B Stat. Methodol. 57, 289–300.

Bevan, S., 2015. Economic impact of musculoskeletal disorders (msds) on work in Europe. Best Pr. Res. Clin. Rheumatol. 29, 356–373.

Bukaty, P., 2019. The California Consumer Privacy Act (CCPA): An Implementation Guide. IT Governance Publishing, URL: http://www.jstor.org/stable/j.ctvjghvnn.

Dave, I.R., Chen, C., Shah, M., 2022. Spact: Self-supervised privacy preservation for action recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. CVPR, pp. 20164–20173.

De Coninck, S., Gamba, E., Van Doninck, B., Bey-Temsamani, A., Leroux, S., Simoens, P., 2025. Enabling privacy-aware ai-based ergonomic analysis. Procedia CIRP 136, 371–376. http://dx.doi.org/10.1016/j.procir.2025.08.065, URL: https://www.sciencedirect.com/science/article/pii/S2212827125008182. 35th CIRP Design 2025.

De Coninck, S., Wang, W.C., Leroux, S., Simoens, P., 2024. Privacy-preserving visual analysis: training video obfuscation models without sensitive labels. Appl. Intell. 1–12.

EU, 2016. Regulation (eu) 2016/679 of the European parliament and of the council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (general data protection regulation). L 119, pp. 1–88, Off. J. Eur. Union.

Gurari, D., Li, Q., Lin, C., Zhao, Y., Guo, A., Stangl, A., Bigham, J.P., 2019. Vizwiz-priv: A dataset for recognizing the presence and purpose of private visual information in images taken by blind people. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 939–948.

Hartley, R., 2003. Multiple View Geometry in Computer Vision, vol. 665, Cambridge University Press.

Hignett, S., McAtamney, L., 2000. Rapid entire body assessment (reba). Appl. Ergon. 31, 201–205. http://dx.doi.org/10.1016/S0003-6870(99)00039-3.

Himmi, S., Ilter, O., Pailleau, F., Siegwart, R., Bescos, B., Cadena, C., 2022. Don't share my face: Privacy preserving inpainting for visual localization. In: 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems. IROS, IEEE, pp. 12506–12511.

Huang, C., Kairouz, P., Chen, X., Sankar, L., Rajagopal, R., 2017. Context-aware generative adversarial privacy. Entropy 19, 656.

Huang, W., Ni, Y., Rezvani, A., Jeong, S., Chen, H., Liu, Y., Wen, F., Imani, M., 2025. Recoverable anonymization for pose estimation: a privacy-enhancing approach. In: 2025 IEEE/CVF Winter Conference on Applications of Computer Vision. WACV, IEEE, pp. 5239–5249.

Hukkelås, H., Lindseth, F., 2023a. Deepprivacy2: Towards realistic full-body anonymization. In: 2023 IEEE/CVF Winter Conference on Applications of Computer Vision. WACV, pp. 1329–1338. http://dx.doi.org/10.1109/WACV56688.2023.00138.

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

Hukkelås, H., Lindseth, F., 2023c. Does image anonymization impact computer vision training? In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. pp. 140–150.

Ilic, F., Zhao, H., Pock, T., Wildes, R.P., 2024. Selective interpretable and motion consistent privacy attribute obfuscation for action recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. CVPR, pp. 18730–18739.

Jocher, G., Qiu, J., Chaurasia, A., 2023. Ultralytics YOLO. URL: https://github.com/ultralytics/ultralytics.

Joshi, M., Deshpande, V., 2019. A systematic review of comparative studies on ergonomic assessment techniques. Int. J. Ind. Ergon. 74, 102865. http://dx.doi.org/10.1016/j.ergon.2019.102865.

Kingma, D.P., Welling, M., et al., 2013. Auto-encoding variational bayes.

Lee, J.H., You, S.J., 2024. Balancing privacy and accuracy: Exploring the impact of data anonymization on deep learning models in computer vision. IEEE Access 12, 8346–8358.

Leroux, S., Simoens, P., Lootus, M., Thakore, K., Sharma, A., 2022. Tinymlops: Operational challenges for widespread edge ai adoption. In: 2022 IEEE International Parallel and Distributed Processing Symposium Workshops. IPDPSW, pp. 1003–1010. http://dx.doi.org/10.1109/IPDPSW55747.2022.00160.

Lin, T.Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C.L., Dollár, P., 2015. Microsoft coco: Common objects in context. arXiv:1405.0312.

Loshchilov, I., Hutter, F., 2019. Decoupled weight decay regularization. URL: https://arxiv.org/abs/1711.05101. arXiv:1711.05101.

Malm, S., Rönnbäck, V., Håkansson, A., Le, M.h., Wojtulewicz, K., Carlsson, N., 2024. Rad: Realistic anonymization of images using stable diffusion. Proceedings of the 23rd Workshop on Privacy in the Electronic Society. Association for Computing Machinery, New York, NY, USA, pp. 193–211. http://dx.doi.org/10.1145/3689943.3695048.

Orekondy, T., Fritz, M., Schiele, B., 2018. Connecting pixels to privacy and utility: Automatic redaction of private information in images. In: Conference on Computer Vision and Pattern Recognition. CVPR.

Pasquadibisceglie, V., Appice, A., Castellano, G., Malerba, D., 2022. A multi-view deep learning approach for predictive business process monitoring. IEEE Trans. Serv. Comput. 15, 2382–2395. http://dx.doi.org/10.1109/TSC.2021.3051771.

Patwari, K., Chuah, C.N., Lyu, L., Sharma, V., 2024. Perceptanon: exploring the human perception of image anonymization beyond pseudonymization for gdpr. In: Forty-First International Conference on Machine Learning.

Sheikh, H.R., Bovik, A.C., 2006. Image information and visual quality. IEEE Trans. Image Process. 15, 430–444.

Shetty, R.R., Fritz, M., Schiele, B., 2018. Adversarial scene editing: Automatic object removal from weak supervision. Adv. Neural Inf. Process. Syst. 31.

Stefana, E., Marciano, F., Rossi, D., Cocca, P., Tomasoni, G., 2021. Wearable devices for ergonomics: A systematic literature review. Sens. 21, 777.

Sun, X., Gazagnadou, N., Sharma, V., Lyu, L., Li, H., Zheng, L., 2024. Privacy assessment on reconstructed images: are existing evaluation metrics faithful to human perception? Adv. Neural Inf. Process. Syst. 36.

Tömekçe, B., Vero, M., Staab, R., Vechev, M., 2024. Private attribute inference from images with vision-language models. Advances in Neural Information Processing Systems. Curran Associates, Inc., pp. 103619–103651, URL: https://proceedings.neurips.cc/paper_files/paper/2024/file/bb97e9a7c811904c9b01f51fde66edcf-Paper-Conference.pdf,

Tseng, Y.Y., Sharma, T., Zhang, L., Stangl, A., Findlater, L., Wang, Y., Gurari, D., 2025. Biv-priv-seg: Locating private content in images taken by people with visual impairments. In: Proceedings of the Winter Conference on Applications of Computer Vision. WACV, pp. 430–440.

Uittenbogaard, R., Sebastian, C., Vijverberg, J., Boom, B., Gavrila, D.M., With, P.H.d., 2019. Privacy protection in street-view panoramas using depth and multi-view imagery. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. CVPR.

Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P., 2004. Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. 13, 600–612.

Wang, Y., Li, M., Cai, H., Chen, W.M., Han, S., 2022. Lite pose: Efficient architecture design for 2d human pose estimation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 13126–13136.

WHO, et al., 2003. The burden of musculoskeletal conditions at the start of the new millennium: report of a who scientific group.

Yan, X., Li, H., Wang, C., Seo, J., Zhang, H., Wang, H., 2017. Development of ergonomic posture recognition technique based on 2d ordinary camera for construction hazard prevention through view-invariant features in 2d skeleton motion. Adv. Eng. Inform. 34, 152–163.

Yu, Y., Yang, X., Li, H., Luo, X., Guo, H., Fang, Q., 2019. Joint-level vision-based ergonomic assessment tool for construction workers. J. Constr. Eng. Manag. 145, 04019025.

Yun, S., Han, D., Oh, S.J., Chun, S., Choe, J., Yoo, Y., 2019. Cutmix: Regularization strategy to train strong classifiers with localizable features. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 6023–6032.

Zhang, J., Cao, X., Han, Z., Shan, S., Chen, X., 2025. Multi-p^{2}a: A multi-perspective benchmark on privacy assessment for large vision-language models. URL: https://arxiv.org/abs/2412.19496. arXiv:2412.19496.

Zhang, H., Cisse, M., Dauphin, Y.N., Lopez-Paz, D., 2017. mixup: Beyond empirical risk minimization. arXiv preprint arXiv:1710.09412.

Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O., 2018. The unreasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 586–595.

Zhao, R., Zhang, Y., Wang, T., Wen, W., Xiang, Y., Cao, X., 2025. Visual content privacy protection: A survey. ACM Comput. Surv. 57, http://dx.doi.org/10.1145/3708501.

Zheng, C., Wu, W., Chen, C., Yang, T., Zhu, S., Shen, J., Kehtarnavaz, N., Shah, M., 2023. Deep learning-based human pose estimation: A survey. ACM Comput. Surv. 56, http://dx.doi.org/10.1145/3603618.

Zhou, J., Pun, C.M., 2020. Personal privacy protection via irrelevant faces tracking and pixelation in video live streaming. IEEE Trans. Inf. Forensics Secur. 16, 1088–1103.

Zhou, L., Zhang, L., Konz, N., 2023. Computer vision techniques in manufacturing. IEEE Trans. Syst. Man Cybern.: Syst. 53, 105–117. http://dx.doi.org/10.1109/TSMC.2022.3166397.