Supplementary Material: Evaluation of Human Visual Privacy: Three-Dimensional Approach and Benchmark Dataset

Sara Abdulaziz
Eindhoven University of Technology
Eindhoven, The Netherlands
s.e.a.m.abdulaziz@tue.nl

Giacomo D'Amicantonio
Eindhoven University of Technology
Eindhoven, The Netherlands
g.d.amicantonio@tue.nl

Egor Bondarev
Eindhoven University of Technology
Eindhoven, The Netherlands
e.bondarev@tue.nl

S1. Supplementary Overview

Section S2: HR-VISPR details
Section S3: Implementation details
Section S4: Additional results

S2. HR-VISPR Dataset

HR-VISPR retains most of the human-related attributes in VISPR [S13], and introduces new labels to cover the soft-biometric attributes, such as the clothing. We present examples of the newly introduced private attributes in HR-VISPR in Fig. S1. The overall privacy attribute distribution in the dataset is shown in Fig. S2, whereas HR-VISPR utility object distribution is shown in Fig. S3.

S3. Implementation Details

This section presents the implementation and training details for the evaluation metrics. Table S1 shows the implementation details of the 11 anonymization methods selected for evaluation on HR-VISPR by the proposed framework.

S3.1. Privacy Metric

A ResNet50 backbone was adopted for the multi-label classifier. The model was trained on the HR-VISPR at a 224 × 224 resolution for 100 epochs. The starting learning rate was $1e-3$, following a linear warmup and a scheduler that drops 1/5 with loss stagnation [S5], [S3]. The batch size was set to 32, and we used Adam optimizer. To compensate for the class imbalance in this multi-label classification setup, we applied class-wise loss weighting (W_c), where weights are inversely proportional to class frequency and normalized by the number of classes. The weights are computed as $W_c = \frac{N}{N_c \times C}$, where N, N_c, and C denote the total number of samples, the number of samples per class, and the total number of classes, respectively.

The dataset was augmented for training the classifier similarly to prior works [S3, S5, S13, S20]. Additionally, we applied random shifting, scaling, rotation, transpose, grid-distortion, and elastic-transform from Albumentation library [S1].

S3.2. Utility Metric

We apply the same training, validation, and testing sets as in the privacy metric training. The utility model was chosen as a YOLOv11 detector [S11], and trained following the same strategy presented by the authors, on the 11 anonymized versions of HR-VISPR. The dataset is augmented for training by the same techniques applied in the default implementation [S11]. For evaluation, the precision, F1-score, and AUC scores are computed for each anonymized version.

S3.3. Practicality Metric

Throughput Score. We measured throughput in a unified setting to eliminate variations due to image size and computational power reported in previous works. First, we selected 40 images from HR-VISPR containing multiple human instances. The images were then resized into 640×640 and processed by all anonymization methods, on a gpu-enabled RTX 4090 device, to compute the throughput according to Eq. 1. The processing time includes both the detection and anonymization time.

Robustness Score. We applied human detection on anonymized images, similarly to utility, but employing a pre-trained model [S11]. First, human objects were detected on the anonymized and original test sets of the HR-VISPR. Then, detections in the original and anonymized sets were matched based on IoU and SSIM scores. Since

Figure S1. Examples of the newly introduced attributes in HR-VISPR, reflecting the clothing type. Together with the previous clothing attributes in VISPR, such as sports, ethnic, and religious clothing, they form a comprehensive set of soft-biometric attributes in HR-VISPR.

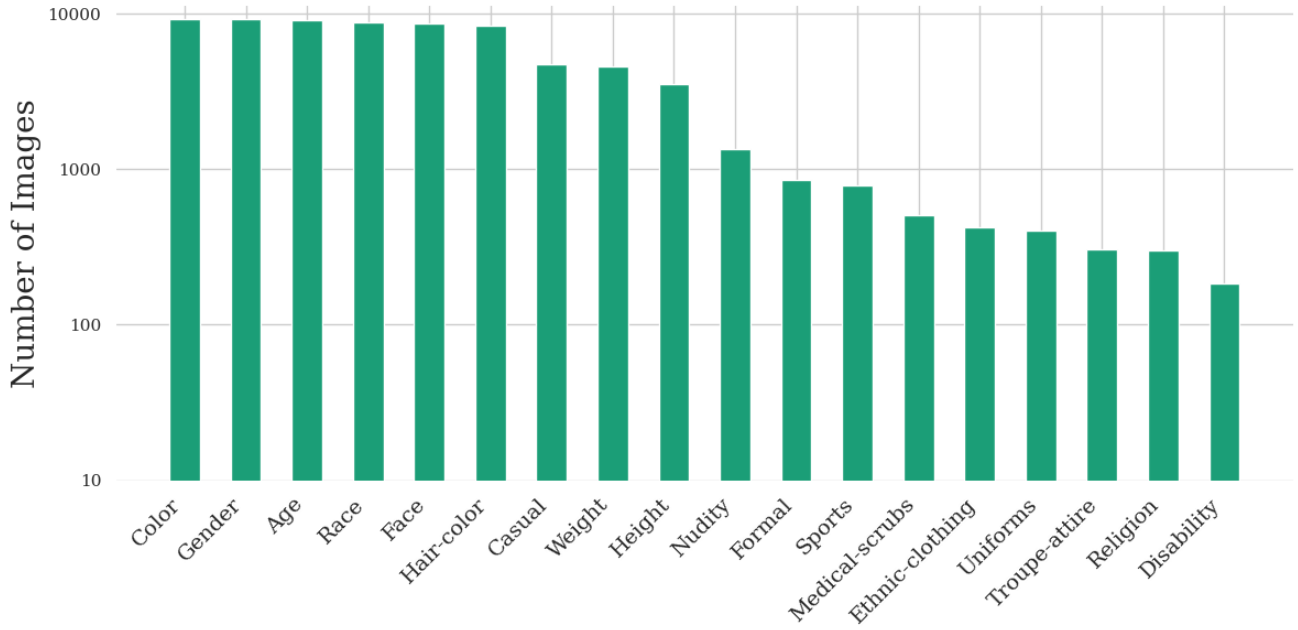


Figure S2. Label distribution in the HR-VISPR dataset. The vertical axis represents the number of images per attribute.

background details are visible within human-object bounding boxes, we applied a 0.99 threshold to prevent matches driven by background similarity alone. Instances with similarity scores higher than the threshold are summed up, representing the method's robustness score.

Intelligibility Score. We computed the CMMD metric [S10] between the original and anonymized HR-VISPR test sets. Since CMMD quantifies discrepancy, lower values indicate a higher similarity to the original data, thereby, higher intelligibility.

The three scores are combined by a weighted sum of normalized values, as detailed in Section 3.2.3. To ensure comparability, the inverted robustness and intelligibility scores were normalized (min-max scaling) before integration into the practicality score. The weights were adjusted between 0.1 and 0.8, to show the contribution of one factor at a time, as explained in the next subsection.

S4. Additional Results

S4.1. Privacy Evaluation

We analyze the privacy scores in conjunction with robustness in Fig. S4, showing how joint analysis of both scores can interpret ambiguities implied by the privacy metric. Ideally, low privacy and high robustness scores reflect effective anonymization, e.g. maintaining low false negative rate and

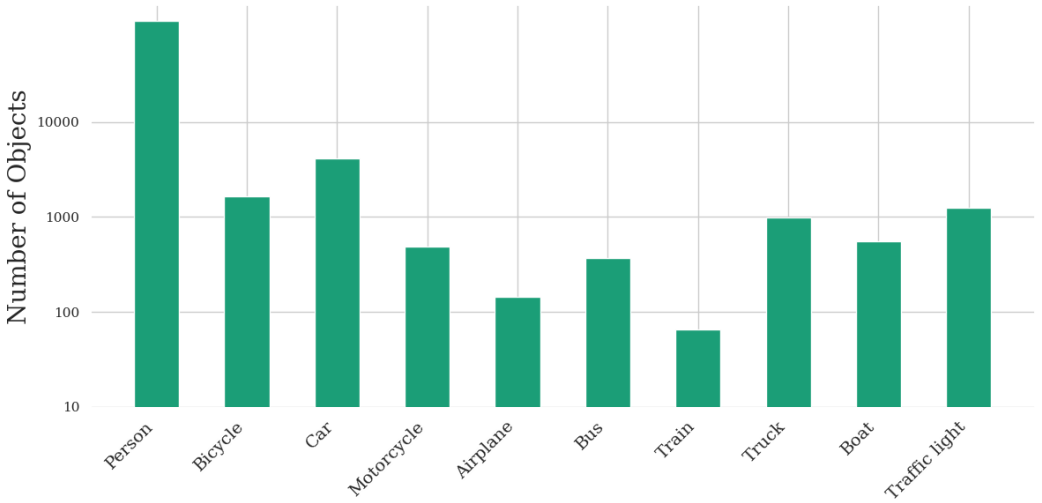


Figure S3. Object label distribution in the HR-VISPR dataset. The vertical axis represents the number of object instances per object class.

Table S1. Implementation details and notation of the anonymization methods selected for the trade-off analysis.

Methods	Papers	Notation	Implementation Details
Human Blurring	[S18, S22]	HB	Segmentation [S7], followed by blurring ($k = 101$)
Human Pixelation	[S2]	HP	Segmentation [S7], followed by pixelation ($k = 20$)
Human Embossing	[S2]	HE	Segmentation [S7], followed by embossing ($k = 3$)
Human Masking	[S21]	HM	Segmentation [S7], followed by blackening
Human Encryption	[S16]	HEN	Segmentation [S7], followed by AES pixel Encryption
Human 2D Avatars	[S2]	H2D	DensePose [S6] with customized avatar
Human 3D Avatar	[S14, S15]	H3D	ROMP [S17]
Human Synthesis	[S9]	HS	DeepPrivacy2 [S9]
Low Resolution + Super Resolution	[S8]	LR+SR	Downscale (30×30) [S19], upscale by SR (×8) [S12], followed by SR (×3) [S4]
SPAct	[S3]	SPct	Public Implementation
TeD-SPAD	[S5]	TSD	Public Implementation

high dissimilarity of anonymized and original sensitive objects. In the light of this, the analysis of HS method reveals its effectiveness since it exhibits high robustness despite showing the lowest protection according to the privacy score only. The high robustness compensates for the privacy metric limitation in distinguishing between generated and original human identities. While the robustness of HS is relatively lower than that of the other methods, this is primarily due to the preservation of gender and clothing style details in some cases, which consequently leads to high similarity scores, thereby lowering the robustness. See the illustration in Fig. S5.

Conversely, H3D shows extremely low robustness due to missing human detections and poor alignment of the avatar-to-human segments in many cases, as seen in Fig. S6.

S4.2. Utility Evaluation

Fig. S7 summarizes the utility results, presenting the average Area Under PR Curves (AUC) across classes for each method. Clearly, the HS and HE show the highest utility performance, followed by HM, H3D, H2D, HP, and HB. Although this analysis highlights utility differences across anonymization methods, identifying the exact causes of utility model failures remains challenging. Certain anonymization techniques introduce additional noise due to poor human segmentation, missing detections, and occlusions involving humans and other objects. Additionally, the introduction of unique features, such as blurred regions, avatars, embossing effects (highlight and shadow), or encrypted pixels, may introduce bias to the models. For these reasons, it can be difficult to draw generalized con-

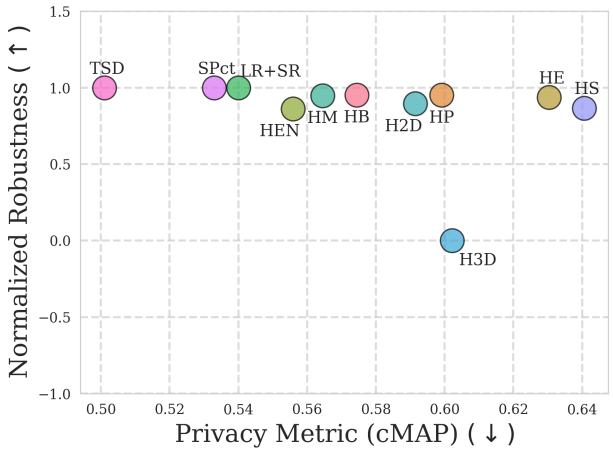


Figure S4. Privacy metric and normalized robustness. Ideal anonymization methods show low privacy and high robustness scores.

Figure S5. Examples of HS anonymization on HR-VISPR samples, generated by HS (DeepPrivacy2 [S9]). Top row shows original, and bottom row shows anonymized images. Despite generating new identities, the method retains gender and clothing attributes, thereby increasing the similarity and lowering the robustness score for these cases.

Figure S6. Visualizations of H3D limitations [S17], which affect its robustness score.

clusions to other tasks and contexts. Nevertheless, applying the same evaluation framework and procedure to these tasks offers a structured approach to extracting the same trade-off insights.

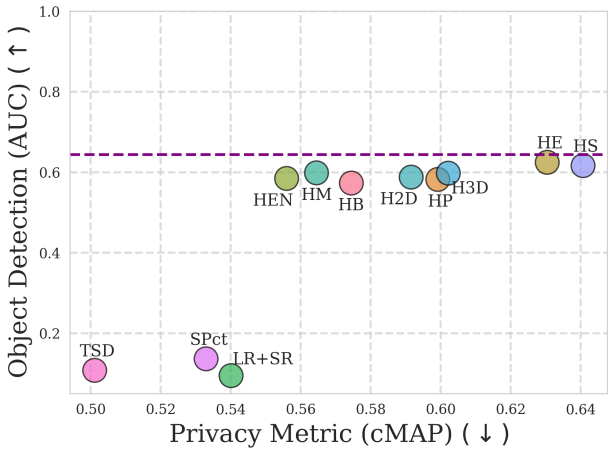


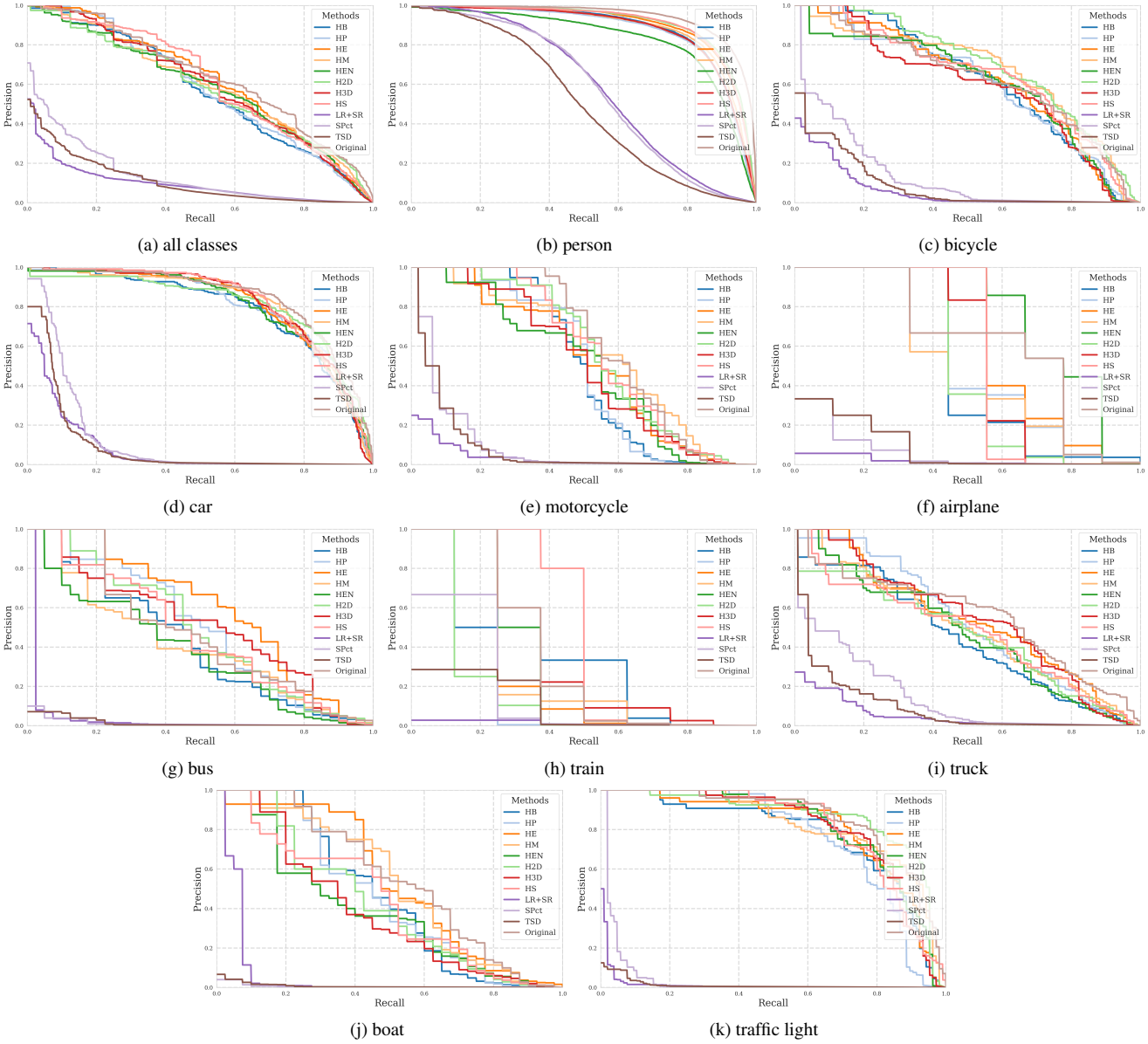
Figure S7. The privacy-utility trade-off evaluation for the anonymization methods, with utility represented as the average AUC of all HR-VISRP object.

To further demonstrate the effect of anonymization methods on utility, we analyze the Precision-Recall (PR) curves for the different classes under each anonymization method, as shown in Fig. S8. The analysis reveals that the approaches which show similar anonymization effects, according to the human perception, may differ in their influence on utility. For instance, although both HM and HEN fully remove human figures, their utility performance varies significantly. HEN consistently underperforms HM (person, bicycle, car, airplane, bus, truck, and boat), suggesting that the noisy masks in HEN disrupt the utility more than the single-color masks HM. In contrast, HP and HB tend to align in most cases despite their different anonymization effects, except that HP shows better performance on few (non-human) classes, such as bus and truck. Despite revealing fine human features in the highlight and shadow effects, HE does not contribute to a higher utility compared to HM, HEN, HP, and HB, except for the person class, suggesting a reinforcement of model bias towards human detection. H2D and H3D reveal human parts in virtual avatars, and are mostly correlated in performance, except for few objects, such as person, motorcycle, and bicycle, where H2D is superior to H3D. This is likely due to the misalignment between human figures and avatars in H3D, which introduces significant noise into the data.

References

[S1] Alexander Buslaev, Vladimir I Iglovikov, Eugene Khvedchenya, Alex Parinov, Mikhail Druzhinin, and Alexandr A Kalinin. Albumentations: fast and flexible image augmentations. *Information*, 11(2):125, 2020. S1

[S2] Pau Climent-Pérez and Francisco Florez-Revuelta. Protection of visual privacy in videos acquired with rgb cam-



(a) all classes (b) person (c) bicycle

(d) car (e) motorcycle (f) airplane

(g) bus (h) train (i) truck

(j) boat (k) traffic light

Figure S8. Precision-Recall curves for all utility classes of HR-VISPR.

eras for active and assisted living applications. *Multimedia Tools and Applications*, 80(15):23649–23664, 2021. S3

[S3] Ishan Rajendrakumar Dave, Chen Chen, and Mubarak Shah. Spact: Self-supervised privacy preservation for action recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20164–20173, 2022. S1, S3

[S4] Chao Dong, Chen Change Loy, and Xiaoou Tang. Accelerating the super-resolution convolutional neural network. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14*, pages 391–407. Springer, 2016. S3

[S5] Joseph Fioresi, Ishan Rajendrakumar Dave, and Mubarak Shah. Ted-spad: Temporal distinctiveness for self-supervised privacy-preservation for video anomaly detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 13598–13609, 2023. S1, S3

[S6] Rıza Alp Güler, Natalia Neverova, and Iasonas Kokkinos. Densepose: Dense human pose estimation in the wild. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7297–7306, 2018. S3

[S7] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017. S3

[S8] Mingzheng Hou, Song Liu, Jiliu Zhou, Yi Zhang, and Ziliang Feng. Extreme low-resolution activity recognition using a super-resolution-oriented generative adversarial network. *Micromachines*, 12(6):670, 2021. S3

[S9] Håkon Hukkelås and Frank Lindseth. Deepprivacy2: Towards realistic full-body anonymization. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1329–1338, 2023. S3, S4

[S10] Sadeep Jayasumana, Srikumar Ramalingam, Andreas Veit, Daniel Glasner, Ayan Chakrabarti, and Sanjiv Kumar. Rethinking fid: Towards a better evaluation metric for image generation. *arXiv preprint arXiv:2401.09603*, 2023. S2

[S11] Glenn Jocher, Ayush Chaurasia, and Jing Qiu. Ultralytics YOLO, 2023. S1

[S12] Wei-Sheng Lai, Jia-Bin Huang, Narendra Ahuja, and Ming-Hsuan Yang. Fast and accurate image super-resolution with deep laplacian pyramid networks. *IEEE transactions on pattern analysis and machine intelligence*, 41(11):2599–2613, 2018. S3

[S13] Tribhuvanesh Orekondy, Bernt Schiele, and Mario Fritz. Towards a visual privacy advisor: Understanding and predicting privacy risks in images. In *Proceedings of the IEEE international conference on computer vision*, pages 3686–3695, 2017. S1

[S14] Hosnieh Sattar, Katharina Krombholz, Gerard Pons-Moll, and Mario Fritz. Body shape privacy in images: understanding privacy and preventing automatic shape extraction. In *Computer Vision–ECCV 2020 Workshops: Glasgow, UK, August 23–28, 2020, Proceedings, Part V 16*, pages 411–428. Springer, 2020. S3

[S15] Jifan Shen and Yuling Sun. Privacy-preserved video monitoring method with 3d human pose estimation. In *2023 26th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, pages 1502–1507. IEEE, 2023. S3

[S16] Amna Shifa, Muhammad Babar Imtiaz, Mamoona Naveed Asghar, and Martin Fleury. Skin detection and lightweight encryption for privacy protection in real-time surveillance applications. *Image and Vision Computing*, 94:103859, 2020. S3

[S17] Yu Sun, Qian Bao, Wu Liu, Yili Fu, Michael J Black, and Tao Mei. Monocular, one-stage, regression of multiple 3d people. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 11179–11188, 2021. S3, S4

[S18] ChiatPin Tay, Vigneshwaran Subbaraju, and Thivya Kandappu. Privobfnet: A weakly supervised semantic segmentation model for data protection. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 2421–2431, 2024. S3

[S19] Yuntao Wang, Zirui Cheng, Xin Yi, Yan Kong, Xueyang Wang, Xuhai Xu, Yukang Yan, Chun Yu, Shwetak Patel, and Yuanchun Shi. Modeling the trade-off of privacy preservation and activity recognition on low-resolution images. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–15, 2023. S3

[S20] Zhenyu Wu, Haotao Wang, Zhaowen Wang, Hailin Jin, and Zhangyang Wang. 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. S1

[S21] Jiawei Yan, Federico Angelini, and Syed Mohsen Naqvi. Image segmentation based privacy-preserving human action recognition for anomaly detection. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 8931–8935. IEEE, 2020. S3

[S22] Zhixiang Zhang, Thomas Cilloni, Charles Walter, and Charles Fleming. Multi-scale, class-generic, privacy-preserving video. *Electronics*, 10(10):1172, 2021. S3