Does Image Anonymization Impact Computer Vision Training?

Håkon Hukkelås Frank Lindseth
Deparment of Computer Science, Norwegian University of Science and Technology
hakon.hukkkelas@ntnu.no

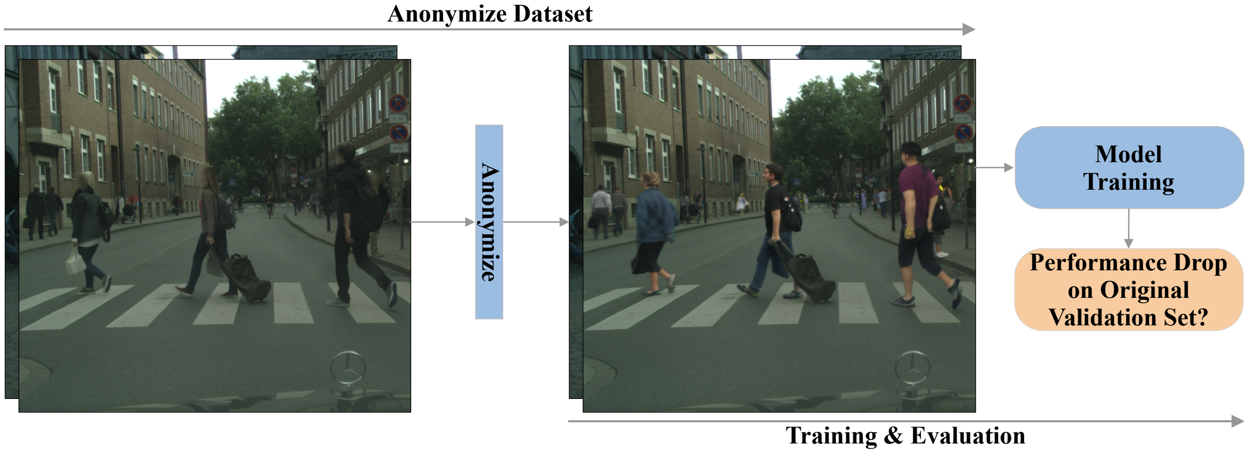

Figure 1. To assess the impact of anonymization, we first anonymize common computer vision datasets, then train various models using the anonymized data, and finally evaluate the models on the original validation datasets. The figure depicts our Cityscapes [8] full-body anonymization experiment. Note that the leftmost image is anonymized with face blurring, following Cityscapes [8] terms of use.

Abstract

Image anonymization is widely adapted in practice to comply with privacy regulations in many regions. However, anonymization often degrades the quality of the data, reducing its utility for computer vision development. In this paper, we investigate the impact of image anonymization for training computer vision models on key computer vision tasks (detection, instance segmentation, and pose estimation). Specifically, we benchmark the recognition drop on common detection datasets, where we evaluate both traditional and realistic anonymization for faces and full bodies. Our comprehensive experiments reflect that traditional image anonymization substantially impacts final model performance, particularly when anonymizing the full body. Furthermore, we find that realistic anonymization can mitigate this decrease in performance, where our experiments reflect a minimal performance drop for face anonymization. Our study demonstrates that realistic anonymization can enable privacy-preserving computer vision development with minimal performance degradation across a range of important computer vision benchmarks.

1. Introduction

Collecting and storing large amounts of visual data is a fundamental task in developing robust and efficient computer vision algorithms. However, this raises concerns regarding the individual's right to privacy, as visual data is rich in privacy-sensitive information, *e.g.* persons, license plates, and street signs. Recent privacy legislation (*e.g.* GDPR [11] in the European Union) requires anonymization when collecting visual data or consent from individuals, which is often infeasible. This can be viewed as a barrier to research and development, particularly for the data-dependent field of Autonomous Vehicle (AV) research. To compensate for these restrictions, practitioners have adopted traditional image anonymization (*e.g.* blurring) for collecting AV datasets [6, 15] and street view images [12].

Traditional image anonymization can protect privacy, but it severely distorts the visual data, potentially reducing its utility for computer vision development. Despite this, face obfuscation (*e.g.* blurring) is the standard method employed to anonymize public autonomous vehicle datasets [6, 15], and its impact on final model performance is currently unclear. Previous work analyzed the impact of face anonymization for classification [59], semantic segmentation [15, 63], object detection [10], action recognition [54], and face detection [30]. In summary, their findings reveal that face anonymization can impact visual recognition related to the human class, and it can severely hurt tasks where the human is in focus [30, 54].

Our literature review, detailed in Sec. 2, resulted in two unanswered questions, which we address in this study.

First, *is realistic anonymization more effective to preserve image utility compared to traditional methods?* Realistic anonymization replaces privacy-sensitive information with synthesized content from generative models, which are found to better preserve utility compared to traditional methods [24, 52]. Previous work has found realistic anonymization to improve utility preservation for semantic segmentation [30, 63]. Our work builds upon this by investigating different objectives and datasets.

Secondly, *to what extent does full-body anonymization impact the training of computer vision models?* The human body is recognizable from many cues outside the face (*e.g*. gait, clothes, ear, body shape), often requiring full-body anonymization to protect privacy. A few studies explore the impact of full-body anonymization [23, 25], where they find it to improve over traditional methods. However, they rely on automatic detection methods, which opens the question if the performance degradation is due to detection errors or the anonymization model. Furthermore, their model requires dense pose estimation [18, 43], which limits its anonymization to individuals close to the camera due to limited long-range detection recall of dense pose models.

In this paper, we focus on key computer vision tasks related to autonomous vehicles, namely instance segmentation and human pose estimation. We evaluate the full-body and face anonymization models built in DeepPrivacy2 [23] and compare realistic anonymization to traditional methods. See https://github.com/hukkelas/deep_privacy2/blob/master/docs/anonymizing_datasets.md to reproduce our experiments.

2. Related Work

Image Anonymization The goal of image anonymization is to remove any privacy-sensitive information contained in the image. Traditional anonymization is widely adopted in practice, where methods anonymize the image via obfuscation (*e.g*. blurring, masking), encryption [20], or k-means [17, 28, 44]. Often, these methods are sufficient to protect privacy; however, they degrade the quality of the data reducing its utility for downstream tasks.

Recent work has introduced *realistic image anonymization*, where anonymization is done by replacing persons with synthesized identities from a generative model. The majority of previous work focuses on face anonymization, where current methods anonymize by *inpainting* a masked out region [24, 38, 52, 53], or *transforming* [7, 13, 50] the original identity to remove privacy-sensitive information. Transformative models often maintain higher utility (*e.g.* preserving facial expression) but offer no formal guarantee of removing the original identity from the image, making them vulnerable to adversarial attacks. A few methods explore anonymizing the full-body [4, 23, 25, 38], where the current state-of-the-art [23, 26] can generate convincing full-bodies given sparse keypoints [26] or dense pose annotations [23]. Finally, some methods insert adversarial perturbation in the image, which is invisible to the human eye but able to fool face recognition systems [46].

Privacy Guarantees of Anonymization Most current anonymization systems offer no formal guarantee of anonymization, and the identity can often be recognized from other cues in the image. Image blurring is discussed numerous times in the literature [3, 16, 35, 36, 42, 44], where the identity is often recognizable due to limited blurring. Furthermore, the identity is recognizable even though the face is anonymized through other identifying attributes of the human body [32, 39, 56], such as gait [27], clothing [14], and body appearance [45, 62]. This makes full-body anonymization more effective than face anonymization in terms of privacy. Finally, most anonymization systems rely on automatic detection, which is far from perfect and vulnerable to adversarial attacks [31].

Public Anonymized Datasets The prominent computer vision datasets employ no form of anonymization, where only a few datasets are anonymized. NuScenes [6] contains images from vehicles driving in Singapore and Boston, where faces and license plates are anonymized via blurring. A2D2 [15] includes data from southern Germany, where license plates and heads are blurred to comply with German privacy regulations. AViD [48] is a video dataset for action recognition with blurred heads. P3M [33] is a portrait matting dataset where every face is blurred. [55] propose a dataset containing street view scenes where cars and pedestrians are removed via image inpainting.

Visual Recognition on Anonymized Data There exists a limited set of studies exploring the effect that anonymization has on training computer vision models. For ImageNet [9] training, face obfuscation (blurring) has little effect on top-5 accuracy and no impact on feature transferability to scene recognition, object localization, and face attribute classification. Nevertheless, anonymization slightly degrades accuracy in classes appearing together with faces (*e.g.* facial masks). For autonomous vehicle datasets, traditional face anonymization can degrade instance segmentation on Cityscapes [8, 63], whereas realistic

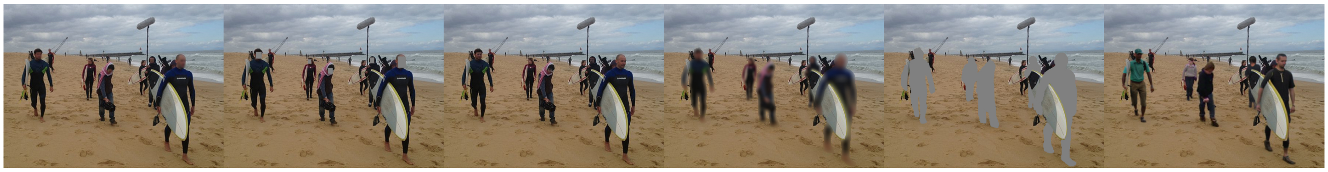
(a) Face - Gaussian (b) Face - Maskout (c) Face - Realistic (d) Body - Gaussian (e) Body - Mask out (f) Body - Realistic

Figure 2. The different anonymization methods evaluated in this paper. Image from COCO train2017 [37], image id=000000097507.

face anonymization has no noticeable negative impact. Furthermore, they find that larger backbones and multi-scale features are more robust to image anonymization [63]. Dvoracek *et al*. [10] finds little impact of face anonymization on object detection on the same dataset. Geyer [15] finds that face anonymization has little effect on semantic segmentation on the A2D2 dataset. For face detection, realistic anonymization performs substantially better than traditional methods for training face detectors [30]. For action recognition, face obfuscation significantly degrades performance [54], where the authors propose a teacher-student self-distillation framework to mitigate the degradation.

Finally, we note that some studies focus on the human perspective and investigate the effect of different anonymization techniques on the users' perceived experience [19, 35].

3. Anonymization Method

In this paper, we explore three different anonymization techniques for full-body and face anonymization; blurring, mask-out, and realistic anonymization (see Fig. 2). Given the image I and a mask M indicating the region to be anonymized, the goal of each method is to remove any privacy-sensitive information within M. In this section, we first define M for face and full-body anonymization (Section 3.1), then introduce the anonymization methods in Section 3.2 and Section 3.3.

3.1. Anonymization Region

To define the anonymization region, we employ the pre-defined instance segmentation annotations for the person/pedestrian class, as every dataset in this paper includes such annotations. Note that we do not anonymize annotations marked as ”crowd” or ”ignored” in the datasets, nor classes that often contain a person (*e.g*. bicycle, motorcycle), as the realistic anonymization techniques require distinct instance-wise annotations. Given the two aforementioned filtering criteria, it is important to note that we are not able to anonymize all individuals in the dataset. An alternative option is to obtain instance-wise annotations by manual annotation or automatic detection. However, we decided against this approach, as the former is too time-consuming, and the latter may introduce detection errors, making it unclear if performance degradation is due to detection errors or poor anonymization.

Face Region As none of the benchmark datasets include annotated faces, we define the face anonymization region following a standard face detection dataset, WIDER-Face [60]. Specifically, the region is the minimal bounding box containing the forehead, chin, and cheek. We annotate each dataset with a pre-trained face detector (DSFD [34]), where we filter the detections by matching them with annotated instance segmentations. We match boxes to segmentations via Intersection over Union (IoU), where we select the match with the highest IoU and bounding box score. Any matches with an IoU $< 1\%$ are removed.

Full-Body Anonymization Since all benchmark datasets include annotated instance segmentations, we use these to define the full-body anonymization region. To compensate for annotations where the segmentations don't fully encompass the body (often segmentation does not include bordering pixels), we slightly dilate the segmentation following [23].

3.2. Traditional Anonymization

We evaluate two commonly used obfuscation techniques for traditional anonymization, namely blurring and masking out. Note that we employ the same method for both face and full-body anonymization.

Mask-Out Mask-out defines the anonymized image as $I_{new} = I \odot (1 - M) + M \odot 127$, where $\odot$ is element-wise multiplication.

Gaussian Blur Gaussian blur defines the anonymized image as $I_{new} = I \odot (1 - M) + M \odot I_{blur}$. Here, I_{blur} is the blurred image with a Gaussian filter ($\sigma = 7$, k-size= $3 \cdot \sigma$).

3.3. Realistic Anonymization

For realistic anonymization, we employ pre-trained models from DeepPrivacy2 [23]. Note that DeepPrivacy2 anonymizes by inpainting (illustrated in Fig. 3), such that it never observes the masked region in I. Thus, it provides similar privacy protection as mask-out anonymization.

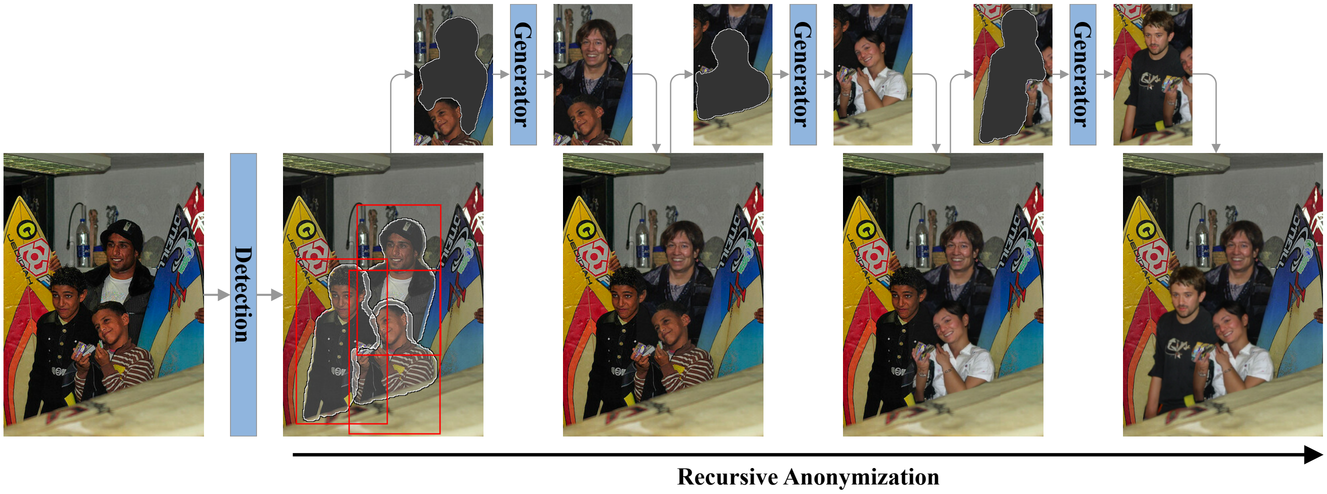

Figure 3. DeepPrivacy2 [23] anonymizes one instance at a time, then paste each synthesized individual into the original image. For our experiments, detection is not performed, as segmentation masks are defined from pre-defined annotations (see Sec. 3.1). Note that the generator relies on keypoint annotations, which are not depicted here.

Face Anonymization For face anonymization, we employ the face anonymization model in DeepPrivacy2 [23], which is a U-Net GAN trained on FDF [24] that synthesizes faces at 128×128 resolution. This model does not rely on keypoint annotations, which enables it to anonymize all faces detected.

Full-Body Anonymization For full-body anonymization, we employ a U-Net GAN [26] relying on keypoint annotations following the COCO format [37]. This model is trained on the FDH dataset [23], and the model is integrated into the DeepPrivacy2 framework [23]. For datasets without keypoint annotations, we use a top-down pose estimation network (ViTPose [58]) which estimates the pose given the image and the minimal bounding box encompassing the instance segmentation. All keypoints with a confidence $\geq 30\%$ are assumed to be visible.

3.4. Global Context for Full-Body Synthesis

In our preliminary experiments, we observed that the full-body generative model often generated human bodies that fit the local context of the generative model but did not align with the global context. We believe this is not a limitation of the generative model itself but a limitation to the crop-based anonymization method used by DeepPrivacy2 (see Fig. 3). In this paper, we explore two solutions to this issue; ad-hoc histogram equalization and histogram matching via latent optimization illustrated in Fig. 4

Histogram Matching (HM) A naive approach for matching the generated body to the global context is naive histogram equalization. Specifically, we match the synthesized (cropped) image to the original (cropped) image by using skimage match_histogram. This adjusts the synthesized image such that each color channel (RGB) matches the cumulative histogram of the original image. To reduce bordering effects when pasting the equalized image into the original image, we smoothly transition the border by slightly blurring the mask with a gaussian filter. That is, given the cropped image x, the corresponding mask M_c, and the synthesized image y, the new image is given by; $y_{new} = x \odot (1 - M_c^{blurred}) + y \odot M_c^{blurred}$, where $M_c^{blurred}$ is M_c blurred with a gaussian filter with size=$[19, 19]$ and $\sigma = 9$. We note that this is far from an optimal solution, where naive histogram matching can introduce severe visual artifacts Fig. 5.

Histogram Matching via Latent Optimization (HM-LO) An alternative approach to post-processing the output is a search in the latent space of the generator. Conceptually, if the exact environmental context (*e.g.* scene lightning) is not given by the cropped image, it should be possible to adjust such factors through the latent space of the generator. Therefore, we suggest utilizing gradient descent to modify the latent vector of the generator, aligning the histogram of the generated image with that of the original image

Given the cropped image x and the mask M_c, the generated image is $y = G(x \odot M_C, \omega)$, where ω is the latent space of the generator, following StyleGAN [29]. Given x, we adjust a sampled ω via gradient descent such that y matches the histogram of x in the S and V channel of the HSV transform of x and y. Specifically, we optimize;

$$\mathcal{L}(x_{hsv}, y_{hsv}) = \mathbb{W}(P_S(x_{hsv}), P_S(y_{hsv})) + \mathbb{W}(P_V(x_{hsv}), P_V(y_{hsv})), \tag{1}$$

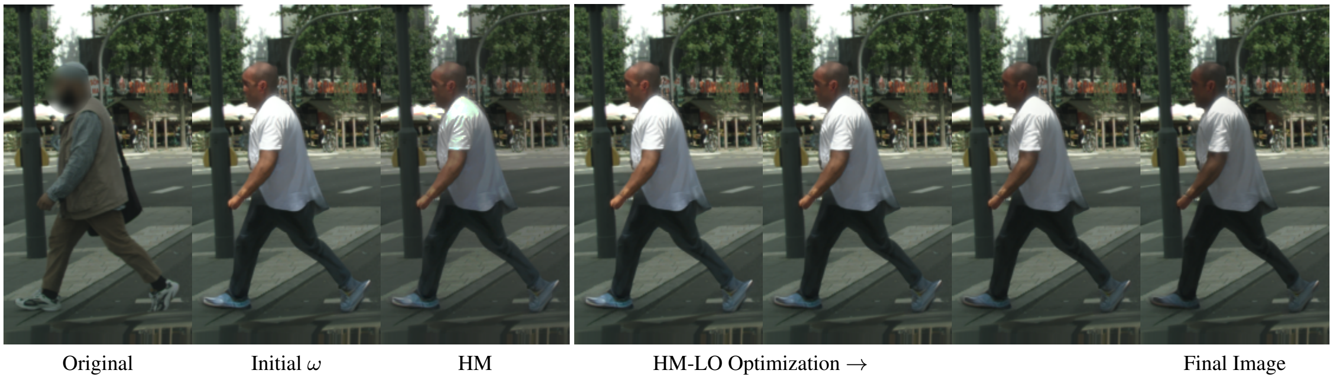

Figure 4. The initial synthesized identity (”initial ω”) may not align with the global context of the image, making the synthesized identity ”stick out” compared to the original identity. We explore two options to address this issue: naive histogram matching (**HM**), and Histogram matching via latent optimization (**HM-LO**), which iteratively adjusts the initial ω to better fit the histogram of the original image (in HSV)

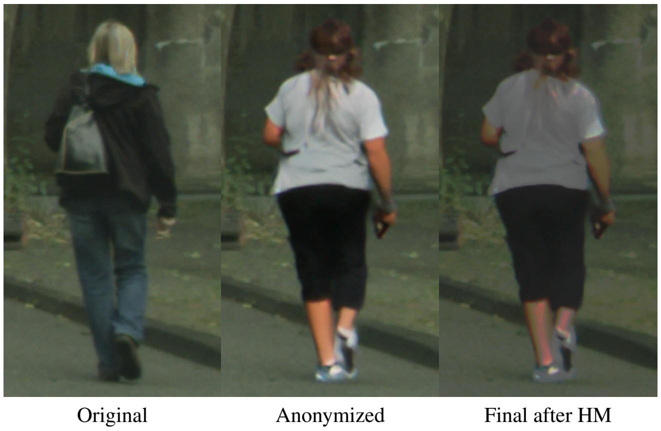

Figure 5. Naive histogram matching can introduce visual artifacts.

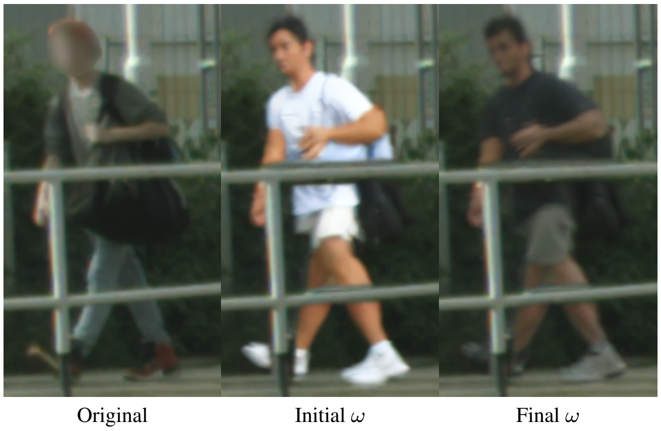

Figure 6. Histogram Matching via Latent Optimization can induce significant semantic changes to the synthesized identity, due to directly optimizing ω to match the HSV histogram (S/V channels).

where $\mathbb{W}$ is the Wasserstein-1 distance, and P_V, P_S is the histogram of the S and V color channel in the HSV transformed image of x and y. Then, we perform gradient descent on ω for 100 steps or until $\mathcal{L}(x_{hsv}, y_{hsv}) < 0.02$.

Often, HM-LO induces slight adjustments to the generated image such that it better matches the context of the image (Fig. 4). However, we note that HM-LO can induce significant semantic changes if the original sampled colors deviate from the original identity (Fig. 6).

4. Experiments

In this section, we report results for training on anonymized data. We train each model on the anonymized dataset and report standard evaluation metrics on the original validation set. To reduce randomness, we report the average and standard error over three independent training runs using seeds 0, 1, and 2. All experiments are done with Pytorch 1.12 [47] on a single NVIDIA A100-40GB. Random qualitative examples from our experiments are given in Appendix B.

4.1. Experimental Details

COCO Pose Estimation We train a Keypoint R-50 FPN R-CNN using detectron2 [57] on the COCO2017 dataset [37]. The training dataset contains 118,287 images with 149,813 person instances (after filtration following Sec. 3), and we evaluate on the original validation dataset (5K images). Out of 149,813 instances, 95,295 are detected by the face detector. Detectron2 is run with commit: 58e472e076

Cityscapes Instance Segmentation We train Mask R-CNN [21] R-50 FPN using detectron2 [57] on the Cityscapes dataset [8]. The training dataset contains 2,975 images with 17,919 person instances (after filtration following Sec. 3), and we evaluate on the original validation dataset (500 images). Out of 17,919 instances, 4,456 were

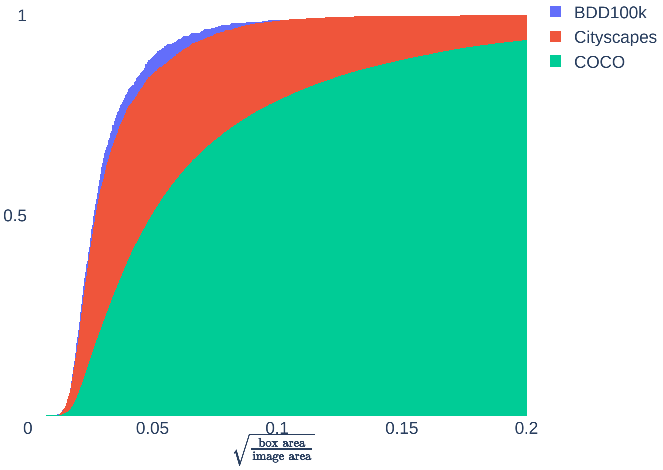

Figure 7. Cumulative histogram of average bounding box length normalized to image size.

Table 1. Instance segmentation AP on the Cityscapes [8] validation set with a Mask R-CNN [21] R-50 FPN. **HM**=Histogram matching (Sec. 3.4). **HM-LO**=Histogram matching via Latent Optimization (Sec. 3.4).

	Anonymization Method	**AP** ↑	**AP50** ↑	**AP$_{\text{person}}$**
	Original	36.7 ± 0.1 (Δ)	62.8 ± 0.2	35.0 ± 0.2 (Δ)
Face	Blur	36.4 ± 0.2 (-0.3)	62.5 ± 0.2 (-0.3)	34.9 ± 0.1 (-0.1)
	Mask-out	**36.7** ± 0.2 (0.0)	**63.1** ± 0.2 (0.3)	34.9 ± 0.1 (-0.1)
	Realistic	36.6 ± 0.1 (-0.1)	62.8 ± 0.3 (0.0)	**35.0** ± 0.1 (0.0)
Body	Blur	31.4 ± 0.2 (-5.3)	54.5 ± 0.4 (-8.3)	2.1 ± 0.1 (-32.9)
	Mask-out	31.2 ± 0.1 (-5.5)	53.2 ± 0.1 (-9.6)	0.7 ± 0.1 (-34.3)
	Realistic	34.6 ± 0.1 (-2.1)	59.0 ± 0.3 (-3.8)	20.3 ± 0.2 (-14.7)
	Realistic + HM	34.3 ± 0.2 (-2.4)	58.9 ± 0.2 (-3.9)	21.3 ± 0.3 (-13.7)
	Realistic + HM-LO	**34.8** ± 0.2 (-1.9)	**60.0** ± 0.3 (-2.8)	**21.5** ± 0.1 (-13.5)

detected by the face detector. Interestingly, this is a noticeably smaller percentage than for the COCO dataset, which we speculate is due to the dataset distribution (persons in COCO often face the camera, while they often do not in Cityscapes).

BDD100K Instance Segmentation We train Mask R-CNN [21] R-50 FPN using MMDetection [40] on the BDD100K dataset [61]. The training dataset contains 7K images with 9,954 person instances (after filtration following Sec. 3), and we evaluate on the original validation dataset (1K images). Out of 9,954 instances, 687 were detected by the face detector. MMdetection is run with commit: b95583270c.

4.2. Effect of Face Anonymization

We start our analysis by focusing on face anonymization. On Cityscapes and BDD100k (Tab. 1, 2), we observe no significant performance difference from any type of face anonymization. We note that realistic anonymization slightly outperforms mask-out anonymization for both datasets. In Figure 7, we find that the majority of boxes in BDD100K/Cityscapes cover less than 1% of the image area. Thus, it is not surprising that face anonymization has little impact on these datasets.

Table 2. Instance segmentation AP on the BDD100K [61] validation set with a Mask R-CNN [21] R-50 FPN.

	Anonymization Method	**AP** ↑	**AP50** ↑	**AP$_{\text{person}}$**
	Original	20.2 ± 0.2 (Δ)	34.9 ± 0.4 (Δ)	32.0 ± 0.0 (Δ)
Face	Blur	20.5 ± 0.1 (0.3)	35.9 ± 0.1 (1.0)	31.7 ± 0.1 (-0.3)
	Mask-out	20.3 ± 0.1 (0.1)	35.3 ± 0.3 (0.4)	31.4 ± 0.1 (-0.6)
	Realistic	**20.6** ± 0.1 (0.4)	**35.8** ± 0.3 (0.9)	**31.6** ± 0.2 (-0.4)
Body	Blur	15.4 ± 0.1 (-4.8)	26.3 ± 0.2 (-8.6)	0.5 ± 0.0 (-31.5)
	Mask-out	15.3 ± 0.0 (-4.9)	25.5 ± 0.1 (-9.4)	0.0 ± 0.0 (-32.0)
	Realistic	**17.0** ± 0.1 (-3.2)	**28.9** ± 0.4 (-6.0)	**12.8** ± 0.1 (-19.2)

Table 3. Keypoint (Kp.) AP on the COCO [37] validation set with a Keypoint R-50 FPN R-CNN [21].

	Anonymization Method	**Box AP** ↑	**Kp. AP** ↑
	Original	55.7 ± 0.0 (Δ)	65.2 ± 0.0 (Δ)
Face	Blur	50.3 ± 0.2 (-5.4)	53.5 ± 0.2 (-11.7)
	Mask-out	49.9 ± 0.2 (-5.8)	52.0 ± 0.3 (-13.2)
	Realistic	54.3 ± 0.1 (-1.4)	60.6 ± 0.1 (-4.6)
	Realistic + HR Faces	**54.4** ± 0.0 (-1.3)	**60.8** ± 0.2 (-4.4)
Body	Blur	17.8 ± 0.0 (-37.9)	4.4 ± 0.1 (-60.8)
	Mask-out	17.4 ± 0.1 (-38.3)	2.0 ± 0.1 (-63.2)
	Realistic	**24.0** ± 0.1 (-31.7)	**15.6** ± 0.1 (-49.6)

For COCO pose estimation (Tab. 3), face anonymization severely impacts performance, where both mask-out and blurring degrade keypoint AP by $> 10\%$. This performance drop is significant for bounding box AP as well, reflecting that the performance difference is not due to the inability to predict keypoints in the facial region. Likely, this is due to learning that blurring/masking artifacts correlate to the human body. Furthermore, we hypothesize that the major performance drop compared to Cityscapes and BDD100k is due to dataset distribution and not the task at hand. To validate this, we train an instance segmentation model on the anonymized COCO datasets and observe a similar performance drop [1].

Refining COCO Faces Although realistic anonymization significantly improves over traditional methods, there remains a considerable degradation between it and the original COCO dataset. We hypothesize that this degradation results from the following factors; limited synthesis quality, facial keypoint mismatch, and low-resolution synthesis. As the generative model is not conditioned on facial keypoints,

[1] For mask-out, we observe a 6.7% performance drop for Box AP for COCO instance segmentation, compared to a 10.4% drop for Box AP for Keypoint R-CNN in Tab. 3. See Appendix A.2 for more details.

the synthesized identity will likely not match the annotated keypoints. There exists keypoint guided anonymization models [24, 38, 52], which we leave for further work to investigate. Furthermore, the generative model synthesizes faces at 128×128 resolution, introducing upsampling artifacts for any face above. In total, we found 14,688 faces with an area larger than 128^2. To remove these upsampling artifacts, we employ a higher resolution (256×256) face synthesis model from DeepPrivacy2 [23] to anonymize any face larger than 128×128. This slightly improved downstream use (marked *Realistic + HR Faces* in Tab. 3), supporting our hypothesis that upsampling artifacts can degrade image utility for COCO keypoint detection training.

4.3. Effect of Full-Body Anonymization

For full-body anonymization, we observe a substantial decline in performance for both traditional and realistic anonymization methods (Tab. 1, 2, 3). Traditional anonymization leads to a complete degradation in performance, whereas realistic anonymization improves this significantly. Interestingly, the performance of realistic full-body anonymization on BDD100K [61] is noticeably worse than for Cityscapes [8], which we discuss further below.

Clearly, realistic full-body anonymization significantly degrades the performance compared to the original dataset, which we attribute to the following three issues: keypoint detection errors, synthesis limitations, and global context mismatch. Synthesizing realistic human bodies is difficult, and current models may introduce severe visual artifacts for many contexts. Furthermore, current methods rely on a crop-based anonymization method (discussed in Section 3.4), which can result in synthesized identities that do not fit the global context of the image. Section 3.4 introduced naive histogram matching and HM-LO to mitigate this issue, which we find to significantly improve results on the Cityscapes dataset (Table 1).

BDD100k *vs.* Cityscapes The decline in performance is significantly more prominent for BDD100k than Cityscapes, despite both datasets being collected for the same purpose. We suspect this discrepancy stems from two sources; keypoint annotations and dataset resolution. First, ViTPose [58] detects keypoints for 95.8% of the instances in the Cityscapes dataset, whereas it only detects for 85.5% in the BDD100k dataset. Secondly, the BDD100k images are of lower resolution (720p) than Cityscapes (2048×1024). This results in 36% of the instance crops having an area $< 32^2$, compared to 24% for Cityscapes. While lower-resolution bodies are easier to synthesize in theory, the employed generative model operates at the resolution 288×160, and major deviations from this resolution can induce visual artifacts. For example, if we do not anonymize any detections $< 32^2$, BDD100k $\text{AP}_{\text{person}}$ is increased from 12.8% to 19.9%. In contrast, this increases $\text{AP}_{\text{person}}$ from 20.3% to 23.4% for Cityscapes.

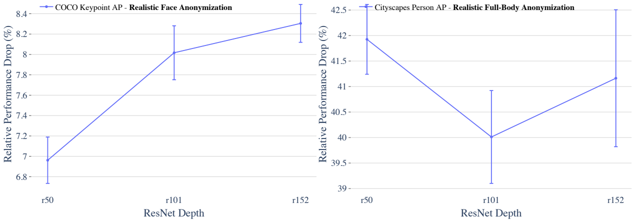

Figure 8. The relative performance drop of realistic anonymization (face or body) for different ResNet depths.

Table 4. Instance segmentation AP on the Cityscapes [8] validation set with full-body anonymization using different latent sampling strategies. Results from Mask R-CNN [21] R-50 FPN.

Anonymization Method Original	**AP ↑** 36.7 ± 0.1 (Δ)	**AP50 ↑** 62.8 ± 0.2 (Δ)	$\textbf{AP}_{\textbf{person}}$ 35.0 ± 0.2 (Δ)
No Truncation	34.0 ± 0.2 (-2.7)	57.7 ± 0.5 (-5.1)	18.6 ± 0.2 (-16.4)
Unimodal Truncation	33.9 ± 0.2 (-2.8)	58.1 ± 0.3 (-4.7)	19.7 ± 0.5 (-15.3)
Multi-modal Truncation (**Default**)	**34.6** ± 0.1 (-2.1)	**59.0** ± 0.3 (-3.8)	**20.3** ± 0.2 (-14.7)

4.4. Ablations

Do Larger Models Generalize Better? Zhou *et al.* [63] observes that deeper models are less impacted by realistic image anonymization. In our experiments, we observed the reverse to be true. We train a ResNet-50, 101, and 152 and compare the relative performance drop of realistic anonymization compared to the original dataset. We investigate this for realistic face anonymization on COCO and full-body anonymization for Cityscapes. Figure 8 reflects that larger models perform worse for both the COCO, whereas it is not clear for the Cityscapes dataset.

Diversity vs. Quality Trade-off GANs can trade off the diversity of samples with quality through the truncation trick [5]. Specifically, by interpolating the input latent variable $z \sim \mathcal{N}(0, 1)$ towards the mode of $\mathcal{N}(0, 1)$, generated diversity is traded off for improved quality. This leaves the question, what is best for anonymization purposes? Limited diversity might result in a detector primarily being able to detect a small diversity of the population, whereas limited quality might reduce transferability to real-world data.

We explore the use of the truncation trick for anonymization purposes, where we investigate the use of no truncation, multi-modal truncation [41] [2], and standard truncation [5]. Note that in all other experiments, multi-modal truncation is used for full-body anonymization, while we use no truncation for face anonymization.

[2] Multi-modal truncation [41] approximates multiple modes of the latent distribution, enabling sampling high-quality images while minimizing the loss of diversity. We estimate 512 cluster centers following [23].

Table 4 reflects that both standard and multi-modal truncation performs substantially better than no truncation for AP_{person}. Furthermore, we observe that multi-modal truncation further improves over standard truncation.

Does Anonymization Impact Other Classes? For many tasks, person detection is not the intended task of the anonymized data (*e.g.* road damage detection [1]). Thus, we investigate the impact of anonymization where person detection is not part of the task. To answer this, we re-train the instance segmentation for the Cityscapes dataset and exclude the ”person” class from the segmentation task.

Our experiment (see Appendix A.3) reflects that full-body anonymization does not impact the detection of the following classes: bus, car, motorcycle, train, or truck. However, we do notice a performance drop for detecting ”rider” and ”bicycle”. We believe this is due to detection overlaps.

5. Conclusion

In this work, we investigated the impact of anonymization for training computer vision models, with a focus on autonomous vehicle datasets. Our experiments reflect that face anonymization (obfuscation and realistic) has little to no impact for instance segmentation on the BDD100K [61] and Cityscapes [8] datasets. In contrast, face obfuscation severely degrades the performance of keypoint detection models on the COCO [37] dataset, as faces are more prevalent in comparison to the BDD100k and Cityscapes datasets. We find that realistic face anonymization can significantly reduce this performance drop. Furthermore, we find that full-body obfuscation severely impairs performance on all datasets, where realistic full-body anonymization can notably alleviate this issue. In summary, our findings reflect that realistic anonymization is a superior option compared to traditional methods. However, they are not a complete substitute for real data, especially for full-body anonymization, as current generative models can often produce unnatural humans that do not fit the given context.

Societal Impact Computer vision models are becoming increasingly adopted for solving challenging tasks everywhere in our society, from manufacturing to driving our cars. These models require task-specific training data to specialize for the task at hand. Collecting such data is troublesome due to privacy legislation, especially for autonomous vehicles which operate in environments where individuals appear everywhere. Our findings indicate that realistic anonymization can effectively substitute the original data, encouraging companies to protect individuals’ privacy without compromising model performance. Our main societal concern is that we do not advocate that the anonymization methods studied in this paper give any sort of privacy guarantee. The detailed discussion in Section 2 clarifies that face anonymization and image blurring are questionable with respect to privacy. Furthermore, anonymized bodies could still be identified, *e.g.* from gait recognition [27].

5.1. Limitations and Further Work

Limitations The primary limitation of our study is the reliance on automatic annotations, where we use DSFD [34] for face detections, and ViTPose [58] for keypoint annotations. While the performance of these methods is impressive, they introduce ambiguity in our results, questioning if the current performance degradation is due to annotation errors or synthesis limitations. Furthermore, due to the filtering criteria for full-body anonymization and automatic annotation of faces, we are not able to anonymize all individuals in the images. Finally, it is also worth mentioning that our analysis is restricted to ResNet [22] and R-CNN [49] based models and that other architectures (*e.g.* YOLO [2]) may respond differently to anonymization artifacts.

Further Work Our explorative analysis of current realistic anonymization techniques highlights several areas of improvement and limitations. To the best of our knowledge, all current anonymization techniques rely on a crop-based anonymization method to improve synthesis quality. However, this can result in a mismatch between the synthesized identity and the global image. For example, the synthesized identity may not align with the global context of the image despite fitting the local crop given to the generative model. To mitigate this, we show that histogram equalization can reduce the impact of this, but we note that histogram equalization is far from the optimal solution. Furthermore, our experiments reflect that there are major practical difficulties remaining in effectively employing generative models for anonymization. For example, current anonymization techniques operate at a fixed synthesis resolution, where large deviations from the operating resolution (*e.g.* bodies smaller than 32^2) result in unnatural images, which impacts performance. Finally, we note that there are several intriguing and unexplored challenges to handle for synthesizing human figures for anonymization in autonomous vehicles. *E.g.* handling multi-view consistency, temporal consistency, or ensuring that the synthesized demography matches the demography of the original data.

Acknowledgement The computations were performed on resources provided by the NTNU IDUN/EPIC computing cluster [51]. Furthermore, we thank Rudolf Mester for his general support and thoughtful discussion.

References

[1] Deeksha M Arya, Hiroya Maeda, S Ghosh, Durga Toshniwal, Yoshihide Sekimoto Indian Institute of Technology Roorkee, India, T U O Tokyo, Japan., UrbanX Technologies, Inc., and Tokyo. RDD2022: A multi-national image dataset for automatic Road Damage Detection. *ArXiv*, abs/2209.0:null, sep 2022. 8

[2] Alexey Bochkovskiy, Chien-Yao Wang, and Hong-Yuan Mark Liao. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv preprint arXiv:2004.10934*, 2020. 8

[3] Michael Boyle, Christopher Edwards, and Saul Greenberg. The effects of filtered video on awareness and privacy. In *Proceedings of the 2000 ACM conference on Computer supported cooperative work*, pages 1–10. ACM, 2000. 2

[4] Karla Brkic, Ivan Sikiric, Tomislav Hrkac, and Zoran Kalafatic. I Know That Person: Generative Full Body and Face De-identification of People in Images. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, volume 2017-July, pages 1319–1328. IEEE, jul 2017. 2

[5] Andrew Brock, Jeff Donahue, and Karen Simonyan. Large Scale GAN Training for High Fidelity Natural Image Synthesis. In *International Conference on Learning Representations*, 2019. 7

[6] Holger Caesar, Varun Bankiti, Alex H. Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. nuScenes: A Multimodal Dataset for Autonomous Driving. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11618–11628. IEEE, jun 2020. 1, 2

[7] Umur A. Ciftci, Gokturk Yuksek, and Ilke Demir. My Face My Choice: Privacy Enhancing Deepfakes for Social Media Anonymization. In *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2023. 2

[8] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The Cityscapes Dataset for Semantic Urban Scene Understanding. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3213–3223. IEEE, jun 2016. 1, 2, 5, 6, 7, 8

[9] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255. IEEE, jun 2009. 2

[10] Petr Dvořáček and Petr Hurtik. What Is the Cost of Privacy? In *Communications in Computer and Information Science*, volume 1602 CCIS, pages 696–706. 2022. 2, 3

[11] European Commission. 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, and repealing Directive 95/46/EC (General Da, 2016. 1

[12] Andrea Frome, German Cheung, Ahmad Abdulkader, Marco Zennaro, Bo Wu, Alessandro Bissacco, Hartwig Adam, Hartmut Neven, and Luc Vincent. Large-scale privacy protection in Google Street View. In *2009 IEEE 12th International Conference on Computer Vision*, pages 2373–2380. IEEE, sep 2009. 1

[13] Oran Gafni, Lior Wolf, and Yaniv Taigman. Live Face De-Identification in Video. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 9377–9386. IEEE, oct 2019. 2

[14] Andrew C. Gallagher and Tsuhan Chen. Clothing cosegmentation for recognizing people. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, jun 2008. 2

[15] Jakob Geyer, Yohannes Kassahun, Mentar Mahmudi, Xavier Ricou, Rupesh Durgesh, Andrew S. Chung, Lorenz Hauswald, Viet Hoang Pham, Maximilian Mühlegg, Sebastian Dorn, Tiffany Fernandez, Martin Jänicke, Sudesh Mirashi, Chiragkumar Savani, Martin Sturm, Oleksandr Vorobiov, Martin Oelker, Sebastian Garreis, and Peter Schuberth. A2D2: Audi Autonomous Driving Dataset. 2020. 1, 2, 3

[16] Ralph Gross, Latanya Sweeney, Jeffrey Cohn, Fernando de la Torre, and Simon Baker. Face De-identification. In *Protecting Privacy in Video Surveillance*, pages 129–146. Springer London, London, 2009. 2

[17] Ralph Gross, Latanya Sweeney, F. de la Torre, and Simon Baker. Model-Based Face De-Identification. In *2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06)*, pages 161–161. IEEE, 2006. 2

[18] Riza Alp Guler, Natalia Neverova, and Iasonas Kokkinos. DensePose: Dense Human Pose Estimation in the Wild. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7297–7306. IEEE, jun 2018. 2

[19] Rakibul Hasan, Eman Hassan, Yifang Li, Kelly Caine, David J. Crandall, Roberto Hoyle, and Apu Kapadia. Viewer Experience of Obscuring Scene Elements in Photos to Enhance Privacy. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, volume 2018-April, pages 1–13, New York, NY, USA, apr 2018. ACM. 3

[20] Jianping He, Bin Liu, Deguang Kong, Xuan Bao, Na Wang, Hongxia Jin, and George Kesidis. PUPPIES: Transformation-Supported Personalized Privacy Preserving Partial Image Sharing. In *2016 46th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*, pages 359–370. IEEE, jun 2016. 2

[21] Kaiming He, Georgia Gkioxari, Piotr Dollar, and Ross Girshick. Mask R-CNN. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 2980–2988. IEEE, oct 2017. 5, 6, 7

[22] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778. IEEE, jun 2016. 8

[23] Hakon Hukkelas and Frank Lindseth. DeepPrivacy2: Towards Realistic Full-Body Anonymization. In *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 1329–1338. IEEE, jan 2023. 2, 3, 4, 7

[24] Håkon Hukkelås, Rudolf Mester, and Frank Lindseth. DeepPrivacy: A Generative Adversarial Network for Face Anonymization. In George Bebis, Richard Boyle, Bahram

Parvin, Darko Koracin, Daniela Ushizima, Sek Chai, Shinjiro Sueda, Xin Lin, Aidong Lu, Daniel Thalmann, Chaoli Wang, and Panpan Xu, editors, *Advances in Visual Computing*, pages 565–578. Springer International Publishing, Cham, 2019. 2, 4, 7

[25] Hakon Hukkelas, Morten Smebye, Rudolf Mester, and Frank Lindseth. Realistic Full-Body Anonymization with Surface-Guided GANs. In *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 1430–1440. IEEE, jan 2023. 2

[26] Håkon Hukkelås and Frank Lindseth. Synthesizing anyone, anywhere, in any pose. *arXiv preprint arXiv:2304.03164*, 2023. 2, 4

[27] Arun Jain, Anil and Flynn, Patrick and Ross. *Handbook of Biometrics*. Springer US, Boston, MA, 2008. 2, 8

[28] Amin Jourabloo, Xi Yin, and Xiaoming Liu. Attribute preserved face de-identification. In *2015 International Conference on Biometrics (ICB)*, pages 278–285. IEEE, may 2015. 2

[29] Tero Karras, Samuli Laine, and Timo Aila. A Style-Based Generator Architecture for Generative Adversarial Networks. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4396–4405. IEEE, jun 2019. 4

[30] Sander R. Klomp, Matthew Van Rijn, Rob G.J. Wijnhoven, Cees G.M. Snoek, and Peter H.N. De With. Safe Fakes: Evaluating Face Anonymizers for Face Detectors. In *2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021)*, pages 1–8. IEEE, dec 2021. 2, 3

[31] Alexey Kurakin, Ian J. Goodfellow, and Samy Bengio. Adversarial Examples in the Physical World. In *Artificial Intelligence Safety and Security*, pages 99–112. jul 2018. 2

[32] Karen Lander, Vicki Bruce, and Harry Hill. Evaluating the effectiveness of pixelation and blurring on masking the identity of familiar faces. *Applied Cognitive Psychology*, 15(1):101–116, jan 2001. 2

[33] Jizhizi Li, Sihan Ma, Jing Zhang, and Dacheng Tao. Privacy-Preserving Portrait Matting. In *Proceedings of the 29th ACM International Conference on Multimedia*, pages 3501–3509, New York, NY, USA, oct 2021. ACM. 2

[34] Jian Li, Yabiao Wang, Changan Wang, Ying Tai, Jianjun Qian, Jian Yang, Chengjie Wang, Jilin Li, and Feiyue Huang. DSFD: Dual Shot Face Detector. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5055–5064. IEEE, jun 2019. 3, 8

[35] Yifang Li, Nishant Vishwamitra, Bart P. Knijnenburg, Hongxin Hu, and Kelly Caine. Blur vs. Block: Investigating the Effectiveness of Privacy-Enhancing Obfuscation for Images. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, volume 2017-July, pages 1343–1351. IEEE, jul 2017. 2, 3

[36] Yifang Li, Nishant Vishwamitra, Bart P. Knijnenburg, Hongxin Hu, and Kelly Caine. Effectiveness and Users' Experience of Obfuscation as a Privacy-Enhancing Technology for Sharing Photos. *Proceedings of the ACM on Human-Computer Interaction*, 1(CSCW):1–24, dec 2017. 2

[37] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft COCO: Common Objects in Context. In *European conference on computer vision*, volume 8693 LNCS, pages 740–755. Springer, Cham, 2014. 3, 4, 5, 6, 8

[38] Maxim Maximov, Ismail Elezi, and Laura Leal-Taixe. CIAGAN: Conditional Identity Anonymization Generative Adversarial Networks. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5446–5455. IEEE, jun 2020. 2, 7

[39] Richard McPherson, Reza Shokri, and Vitaly Shmatikov. Defeating Image Obfuscation with Deep Learning. sep 2016. 2

[40] MMDetection Contributors. OpenMMLab Detection Toolbox and Benchmark, 2018. 6

[41] Ron Mokady, Omer Tov, Michal Yarom, Oran Lang, Inbar Mosseri, Tali Dekel, Daniel Cohen-Or, and Michal Irani. Self-Distilled StyleGAN: Towards Generation from Internet Photos. In *ACM SIGGRAPH 2022 Conference Proceedings*, pages 1–9. ACM, aug 2022. 7

[42] Carman Neustaedter, Saul Greenberg, and Michael Boyle. Blur filtration fails to preserve privacy for home-based video conferencing. *ACM Transactions on Computer-Human Interaction*, 13(1):1–36, mar 2006. 2

[43] Natalia Neverova, David Novotny, Vasil Khalidov, Marc Szafraniec, Patrick Labatut, and Andrea Vedaldi. Continuous Surface Embeddings. In *Advances in Neural Information Processing Systems*, volume 33, pages 17258—-17270. Curran Associates, Inc., nov 2020. 2

[44] E.M. Newton, Latanya Sweeney, and Bradley Malin. Preserving privacy by de-identifying face images. *IEEE Transactions on Knowledge and Data Engineering*, 17(2):232–243, feb 2005. 2

[45] Seong Joon Oh, Rodrigo Benenson, Mario Fritz, and Bernt Schiele. Faceless Person Recognition: Privacy Implications in Social Media. In *Computer Vision – ECCV 2016*, pages 19–35. Springer Verlag, 2016. 2

[46] Seong Joon Oh, Mario Fritz, and Bernt Schiele. Adversarial Image Perturbation for Privacy Protection A Game Theory Perspective. In *2017 IEEE International Conference on Computer Vision (ICCV)*, volume 2017-Octob, pages 1491–1500. IEEE, oct 2017. 2

[47] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, and Others. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019. 5

[48] A. J. Piergiovanni and Michael S. Ryoo. AViD dataset: Anonymized videos from diverse countries. In *Advances in Neural Information Processing Systems*, volume 2020-Decem, 2020. 2

[49] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*, pages 91–99, 2015. 8

[50] Zhongzheng Ren, Yong Jae Lee, and Michael S Ryoo. Learning to Anonymize Faces for Privacy Preserving Action

Detection. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 639–655. 2018. 2

[51] Magnus Själander, Magnus Jahre, Gunnar Tufte, and Nico Reissmann. EPIC: An energy-efficient, high-performance GPGPU computing research infrastructure, 2019. 8

[52] Qianru Sun, Liqian Ma, Seong Joon Oh, Luc Van Gool, Bernt Schiele, and Mario Fritz. Natural and Effective Obfuscation by Head Inpainting. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5050–5059. IEEE, jun 2018. 2, 7

[53] Qianru Sun, Ayush Tewari, Weipeng Xu, Mario Fritz, Christian Theobalt, and Bernt Schiele. A Hybrid Model for Identity Obfuscation by Face Replacement. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 570–586. 2018. 2

[54] Matteo Tomei, Lorenzo Baraldi, Simone Bronzin, and Rita Cucchiara. Estimating (and fixing) the Effect of Face Obfuscation in Video Recognition. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 3257–3263. IEEE, jun 2021. 2, 3

[55] Ries Uittenbogaard, Clint Sebastian, Julien Vijverberg, Bas Boom, Dariu M. Gavrila, and Peter H.N. de With. Privacy Protection in Street-View Panoramas Using Depth and Multi-View Imagery. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2019-June, pages 10573–10582. IEEE, jun 2019. 2

[56] Michael J. Wilber, Vitaly Shmatikov, and Serge Belongie. Can we still avoid automatic face detection? In *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1–9. IEEE, mar 2016. 2

[57] Yuxin Wu, Alexander Kirillov, Francisco Massa, Wan-Yen Lo, and Ross Girshick. Detectron2, 2019. 5

[58] Yufei Xu, Jing Zhang, Qiming Zhang, and Dacheng Tao. ViTPose: Simple Vision Transformer Baselines for Human Pose Estimation. 4, 7, 8

[59] Kaiyu Yang, Jacqueline Yau, Li Fei-Fei, Jia Deng, and Olga Russakovsky. A Study of Face Obfuscation in ImageNet. In *International Conference on Machine Learning*, pages 25313—-25330, mar 2022. 2

[60] Shuo Yang, Ping Luo, Chen Change Loy, and Xiaoou Tang. WIDER FACE: A Face Detection Benchmark. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2016-Decem, pages 5525–5533. IEEE, jun 2016. 3

[61] Fisher Yu, Haofeng Chen, Xin Wang, Wenqi Xian, Yingying Chen, Fangchen Liu, Vashisht Madhavan, and Trevor Darrell. BDD100K: A Diverse Driving Dataset for Heterogeneous Multitask Learning. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2633–2642. IEEE, jun 2020. 6, 7, 8

[62] Ning Zhang, Manohar Paluri, Yaniv Taigman, Rob Fergus, and Lubomir Bourdev. Beyond frontal faces: Improving Person Recognition using multiple cues. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 07-12-June, pages 4804–4813. IEEE, jun 2015. 2

[63] Jingxing Zhou and Jurgen Beyerer. Impacts of Data Anonymization on Semantic Segmentation. In *2022 IEEE Intelligent Vehicles Symposium (IV)*, volume 2022-June, pages 997–1004. IEEE, jun 2022. 2, 3, 7