LDFA: Latent Diffusion Face Anonymization for Self-driving Applications

Marvin Klemp[1] Kevin Rösch[1,2] Royden Wagner[1] Jannik Quehl[1] Martin Lauer[1]
[1]Karlsruhe Institute of Technology [2] FZI Research Center for Information Technology
{firstname.lastname}@kit.edu

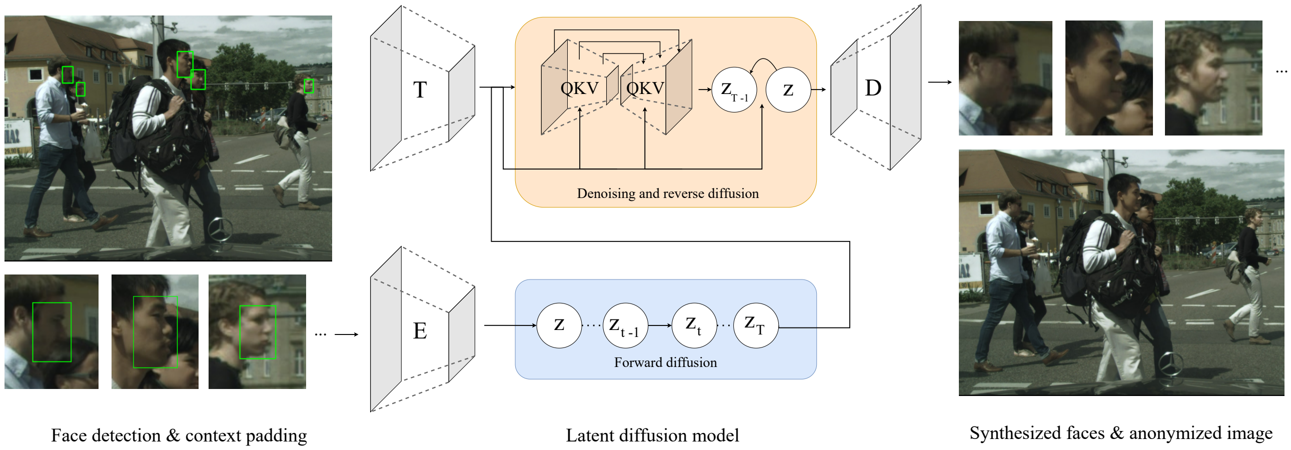
T
QKV
QKV
z_{T-1}
z
D
Denoising and reverse diffusion
...
E
z
z_{t-1}
z_t
z_T
Forward diffusion
Face detection & context padding
Latent diffusion model
Synthesized faces & anonymized image

Figure 1. **LDFA pipeline.** Faces are detected by a RetinaFace detector [4], afterwards, detections are padded to provide a latent diffusion model [28] with context to synthesize realistic faces. Finally, the synthesized faces are inserted in the input image to generate an anomymized image. Further details are in Section 3.

Abstract

In order to protect vulnerable road users (VRUs), such as pedestrians or cyclists, it is essential that intelligent transportation systems (ITS) accurately identify them. Therefore, datasets used to train perception models of ITS must contain a significant number of vulnerable road users. However, data protection regulations require that individuals are anonymized in such datasets. In this work, we introduce a novel deep learning-based pipeline for face anonymization in the context of ITS. In contrast to related methods, we do not use generative adversarial networks (GANs) but build upon recent advances in diffusion models. We propose a two-stage method, which contains a face detection model followed by a latent diffusion model to generate realistic face in-paintings. To demonstrate the versatility of anonymized images, we train segmentation methods on anonymized data and evaluate them on non-anonymized data. Our experiments reveal that our pipeline is better suited to anonymize data for segmentation than naive methods and performes comparably with recent GAN-based methods. Moreover, face detectors achieve higher mAP scores for faces anonymized by our method compared to naive or recent GAN-based methods.

Code is available at https://github.com/KIT-MRT/latent_diffusion_face_anonymization.

1. Introduction

In the context of intelligent transportation systems (ITS), pedestrians and cyclists are classified as vulnerable road users [21]. In case of an accident, they have little to no protection from the impact forces. Therefore, it is crucial that an ITS accurately and robustly detects such vulnerable road users (VRUs) in order to protect them. Hence, datasets used to train perception models of ITS are required to contain a significant number of pedestrians and cyclists. But VRUs should not only be protected physically, as of 25th May 2018, the general data protection regulation (GDPR) came to effect in Europe [6] to also protect their data privacy. The GDPR affects all processing of personal data including images of individuals and requires the consent of the individual for processing their data. This complicates the creation of datasets focused on perception in urban scenes, such as Cityscapes [3] or Mapillary Vistas [23]. Since obtaining the consent of all individuals from whom such a dataset contains recordings is impractical, the recordings must be anonymized. However, naive face anonymization techniques, such as blurring or cropping, can reduce the performance of deep learning models for perception [35]. For instance, perception models trained on datasets with blurring as face anonymization method learn a representation of humans with blurred faces. This limits their ability to generalize to real world scenarios without anonymization. [14] Consequently, anonymization techniques, which produce images that are as realistic as possible are superior [11]. For this purpose, deep learning methods are used that replace faces by artificially generated faces [10, 11, 22]. These methods require complex semantic reasoning to recognize faces and their poses and subsequently replace them with similar artificially generated faces.

In this work, we introduce a novel deep learning-based pipeline for face anonymization in the context of ITS. In contrast to related methods, we do not use generative adversarial networks (GANs), but build upon recent advances in diffusion models [28]. Our main contributions can be summarized as follows:

- We introduce a two-stage pipeline using a face detection model followed by a latent diffusion model to generate realistic face anonymization.
- We show that general diffusion models are equally fitted for face anonymization in comparison to recent, specialized face anonymization methods.
- We show that the mAP of a face detection network inferred on images, which are anonymized with LDFA, is higher compared to recent GAN-based methods.

2. Related Work

Anonymization for ITS. Zhou and Beyerer [35] train semantic segmentation models on anonymized versions of the Cityscapes dataset. Their experiments reveal that face anonymization using blurring or cropping degrades segmentation performance, while face in-painting using GANs does not effect the segmentation performance. Geyer et al. [7] and Wilson et al. [33] use naive anonymization techniques such as blurring or pixelization to provide privacy-preserving datasets for ITS development. These techniques are well suited to anonymize license plates [30] but degrade the segmentation performance when used to anonymize faces [35].

Recent GAN models for face anonymization. DEEPPRIVACY [11] and DEEPPRIVACY2 [10] are closely related methods that jointly perform face detection, artifical face generation, and face in-painting. DEEPPRIVACY detects faces using DSDF [16] and estimates face keypoints using a modified MASK R-CNN [8]. Artificial faces are generated by a conditional GAN model [12] conditioned on the surroundings of faces and estimated face keypoints. DeepPrivacy2 detects and segments faces or whole bodies using an ensemble of three detectors, DSDF, CSE [24], and MASK-RCNN. Artificial faces are generated by a style-based generator inspired by STYLEGAN2 [13].

CIAGAN [22] detects face keypoints using a histogram of oriented gradients (HOG) model. Face keypoints are further processed to generate an abstract face landmark image, which contains pose and expression. Face landmark images and face surroundings are used to guide a conditional GAN for face generation.

These GAN-based methods need to be trained or finetuned on face-specific datasets (e.g., CelebA [17]) to generate realistic artifical faces. In contrast, recent general pretrained diffusion models [19, 28] can generate realistic face in-paintings without further fine-tuning.

3. Method

3.1. Diffusion Models

Diffusion models (DMs) are generative models that recently showed particularly great success in the domain of image generation [26, 27, 29]. But because of how adaptable they are, they can also be used in other fields, like point cloud generation [20] or audio generation [15]. During training, a forward diffusion process based on a Markov Chain is used to incrementally add noise to the input data. Afterwards, a neural network is trained to reverse this process and reconstruct data from noisy representations z_T, see Figure 1. During inference, input data can be transformed into a noisy latent representation in the same way. The amount of added noise controls how much information from the original data is used to synthesize new data. Alternatively, random noise can be used as input and a transformer T to encode multimodal data (e.g., text, semantic maps, images) for conditioning. Conditioning signals are fed into the denoising U-Net using a cross-attention mechanism (QKV) to guide the reverse diffusion process.

3.2. Latent Diffusion Models

In contrast to regular DMs, which work directly on the pixel space, latent diffusion models (LDMs) [28] first transform the input image into a latent feature space and then perform the diffusion process within this lower dimensional space. This is realized by training an autoencoder, which reduces the dimensionality of the input data considerably to a more efficient representation in latent space. The autoencoder consists of an encoder E and a decoder D (cf. Figure 1). This allows for better scaling properties with respect to the spatial dimensionality as well as for more efficient training and evaluation of the resulting model [28].

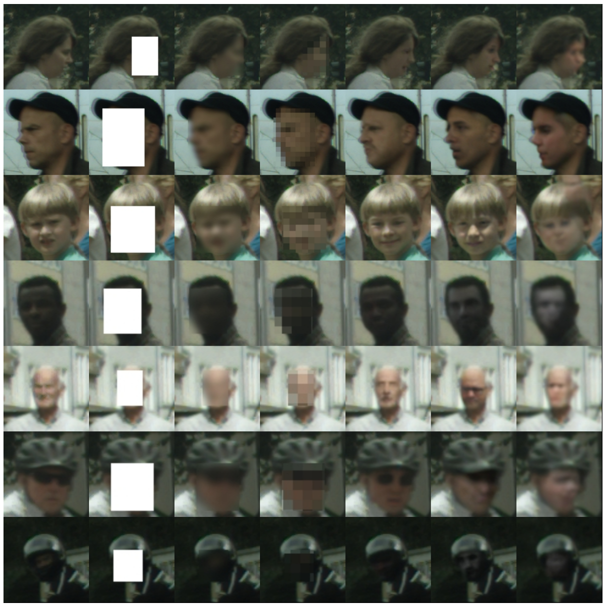

Figure 2. **Application of anonymization methods in different situations**. The applied methods (from left to right) are: None (original image), CROP, GAUSS, PIXEL, LDFA, DEEPPRIVACY1, DEEPPRIVACY2. We advise the reader to zoom in.

3.3. Face Anonymization using Latent Diffusion

Our pipeline starts by detecting the faces which are supposed to be anonymized. For this, we use RETINAFACE, which achieves state-of-the-art results for face detection on the WIDER FACE dataset [4, 32]. As DMs are able to generate arbitrary patches of an image, we favor high recall over high precision in detecting faces and use a low detection confidence threshold of 0.4. Once faces are detected, one image-to-image process is started for each detected bounding box. Each face is padded by 32 pixels before it is cut, see Figure 1. The padded area provides context for the scene during the reverse diffusion process. However, only the facial region without padding is used to replace the contents of the original image. We re-scale the image patch to 512^2 pixels and pass the resized image patch into the image-to-image LDM. Our proposed pipeline is visualized in Figure 1. We parameterize the LDM as follows: we use the *stable-diffusion-2-inpainting* weights [28], no prompt, a CFG-Scale of 1, the *k_euler_a* sampler and 50 inference steps. This process is repeated for each face individually.

4. Experiments

4.1. Anonymization Methods

In the following experiments, we compare various anonymization methods. We consider deep learning based methods DEEPPRIVACY, and DEEPPRIVACY2 which are state-of-the-art methods for face and full-body anonymization. DEEPPRIVACY is used in the default, and DEEPPRIVACY2 in the face-anonymization configuration. Additionally, we evaluate our LDM-based pipeline (LDFA). Next to deep learning based methods, we compare naive machine vision based anonymization methods. The GAUSS method applies a Gaussian filter ($\mu = 0, \sigma = 3$). CROP entirely crops the facial contents and replaces them by the maximal value for all channels. The PIXEL method performs pixelization as anonymization. In detail, face areas are split into 8x8 pixel patches and each patch is assigend the mean value of this patch. Figure 2 shows faces, synthesised by our method, naive, and comparable state-of-the-art methods in a variety of different scenarios which may occur in an autonomous dataset. For our experiments we used the Cityscapes dataset as a showcase [3].

4.2. Training Segmentation Models on Anonymized Images

In this experiment, we evaluate the impact of anonymization methods on semantic segmentation models. Therefore, we train several segmentation models on differently anonymized training sets and evaluate them on the same non-anonymized validation set. As dataset, we use the Cityscapes dataset, which contains recordings of urban driving in European cities. We use the official training split for training, and the validation split is used as a test set.

Experimental setup. As segmentation model, we use the MASK2FORMER model [2]. MASK2FORMER models are state-of-the-art segmentation models based on the vision transformer (ViT) [5] and detection transformer (DETR) [1] architectures. We use the semantic segmentation configuration with a ResNet-50 [9] backbone. All models are trained for 54 epochs with a batch size of 16. We chose ADAMW [18] as optimizer with an initial learning rate of $1 * 10^{-5}$. In addition, we compare our results with the results from a fully-convolutional model PSPNET [34] reported by Zhou and Beyerer [35].

Evaluation metrics. To evaluate the segmentation performance, we are using the intersection-over-union (IoU), the realtive IoU change w.r.t baseline models ($\Delta\text{IoU}^{\text{rel}}$), and the instance-level IoU (iIoU) metric. The $\Delta\text{IoU}^{\text{rel}}$ is computed as

$$\Delta\text{IoU}^{\text{rel}} = \frac{\text{IoU}_{\text{anon}} - \text{IoU}_{\text{base}}}{\text{IoU}_{\text{base}}}, \tag{1}$$

where IoU_{anon} is the IoU score achieved by a model trained on an anonymized train split and IoU_{base} is the IoU score achieved by the same model trained on the non-anonymized train split. The iIoU is computed with

$$\text{iIoU} = \frac{\text{iTP}}{\text{iTP} + \text{FP} + \text{iFN}}, \tag{2}$$

where FP are false positive pixels, iTP and iFN are instance-weighted true positive and false negative pixels. Instance

Table 1. Impacts of anonymization methods on semantic segmentation

Anon. method	Seg. model	$\text{IoU}_{\text{Person}}$	$\Delta\text{IoU}^{\text{rel}}_{\text{Person}}$	$\text{iIoU}_{\text{Person}}$	$\text{IoU}_{\text{Rider}}$	$\Delta\text{IoU}^{\text{rel}}_{\text{Rider}}$	$\text{iIoU}_{\text{Rider}}$	$\text{IoU}_{\text{Human}}$	$\text{iIoU}_{\text{Human}}$
Baselines									
-	Mask2Former	0.818	0.00%	0.660	0.624	0.00%	0.492	0.830	0.694
-	PSPNet	-	0.00%	-	-	0.00%	-	-	-
Naive									
Gauss	Mask2Former	0.813	-0.61%	0.653	0.609	-2.40%	0.479	0.827	0.686
Gauss	PSPNet	-	-0.43%	-	-	-0.82%	-	-	-
Crop	Mask2Former	0.812	-0.73%	0.639	0.608	-2.56%	0.490	0.828	0.677
Crop	PSPNet	-	-0.99%	-	-	-2.87%	-	-	-
Pixel	Mask2Former	0.816	-0.24%	0.643	0.614	-1.60%	0.495	0.829	0.691
Deep learning-based									
DeepPrivacy	Mask2Former	0.817	-0.12%	0.647	0.619	-0.80%	0.497	0.830	0.686
DeepPrivacy	PSPNet	-	+0.08%	-	-	+0.09%	-	-	-
DeepPrivacy2	Mask2Former	0.816	-0.24%	0.646	0.618	-0.96%	0.485	0.829	0.686
LDFA (ours)	Mask2Former	0.816	-0.24%	0.647	0.620	-0.64%	0.496	0.831	0.689

weights are determined by computing the contribution of each pixel by the ratio of the class' average instance size to the size of the respective ground truth instance [3].

Results. Overall, all models achieve higher IoU scores for the person class and the human category than for the rider class. The iIoU scores are lower for all classes; accordingly, large instances are segmented more accurately than small ones. Naive anonymization methods degrade the IoU score for the person class on average by 0.6% compared to the baselines, whereas deep learning-based methods only by 0.13% on average. For the rider class, the performance drops are more significant. Naive methods degrade the performance by 2.05% on average and deep learning-based methods by 0.58%. The cityscapes dataset contains an order of magnitude less pixels for the rider class than for the person class. Hence, it can be inferred that the choice of anonymization method is more important for underrepresented classes or smaller datasets. Gauss anonymization degrades segmentation with Mask2Former more than with PSPNet. Crop anonymization, on the other hand, degrades segmentation performance more with PSPNet than with Mask2Former. Overall, the deep learning-based methods tend to affect segmentation performance less and are therefore better suited to anonymize datasets in this context.

4.3. Face Detection on Anonymized Images

While training on synthetic data it is highly important that the artificially created data is close to real-world data. After applying an anonymization method, persons should not be recognizable compared to the original image, but their faces should still be detectable. In this experiment we investigate the impacts of face anonymization methods on a recent face detector.

Experimental setup. We run the RetinaFace face detection algorithm on the original Cityscapes train set images. Detected faces are considered as ground truth. In total, 3765 faces are detected in the train set. Faces can be categorized into a small (face $< 32^2$ pixels), medium (face $< 32^2$ pixels $\wedge$ face $> 96^2$ pixels) and large (face $\geq 96^2$ pixels) category. Since the cameras are mounted on a test vehicle, there is always a safety distance between camera and pedestrians. As a result, the train set is heavily skewed towards smaller people, and thus smaller faces. (3214 small, 543 medium, and 8 large faces). We apply each anonymization method to the original image and then repeat the face detection.

Evaluation metrics. We use the mAP as evaluation metric and provide mAP_{S}, mAP_{M}, and mAP_{L} for the small, medium and large face category. In our evaluation, we assume that a high mAP corresponds to the fact, that from an algorithmic viewpoint, a face is still detected as a face after anonymization. Furthermore, an essential aspect of anonymization is that all persons within an image have to be anonymized. Hence, we further evaluate the methods by the number of anonymized faces (NoA).

Table 2. Impacts of anonymization methods on face detection

Model	mAP	mAP_{S}	mAP_{M}	mAP_{L}	
Naive					
Gauss	0.3728	0.3041	0.6991	0.4673	
Crop	0.0012	0.0000	0.0034	0.0000	
Pixel	0.2153	0.1697	0.4221	0.2792	
Deep learning-based					NoA
DeepPrivacy	0.5661	0.5320	0.6568	**0.7943**	1878
DeepPrivacy2	0.5206	0.4738	0.6657	0.0874	**3765**
LDFA (ours)	**0.6754**	**0.6652**	**0.6930**	0.3168	**3765**

Results. Table 2 shows the evaluation results of this experiment. Naive methods heavily change facial contents without maintaining enough semantic properties for successfully detecting faces after anonymization, resulting in a low mAP. In contrast, deep learning based methods attempt to generate semantically accurate faces. The impact of that is repesented by the substantial improvement in mAP in all deep learning based methods compared to naive methods. Our method outperforms both DeepPrivacy, DeepPrivacy2 and all naive methods in mAP, mAP_S, and mAP_M. One assumption for this is that LDMs generally perform better for the generation of smaller faces compared to GANs. Furthermore, our method detects nearly twice as many faces shown by NoA compared to the default settings of DeepPrivacy. A high NoA may result wrongly detected faces. However, compared to GAN based methods, an LDM which is trained on a general dataset can infer an alternative version of a false positive. Whereas a GAN would infer a face within the falsely detected face. An example of this behaviour is shown in Figure 3.

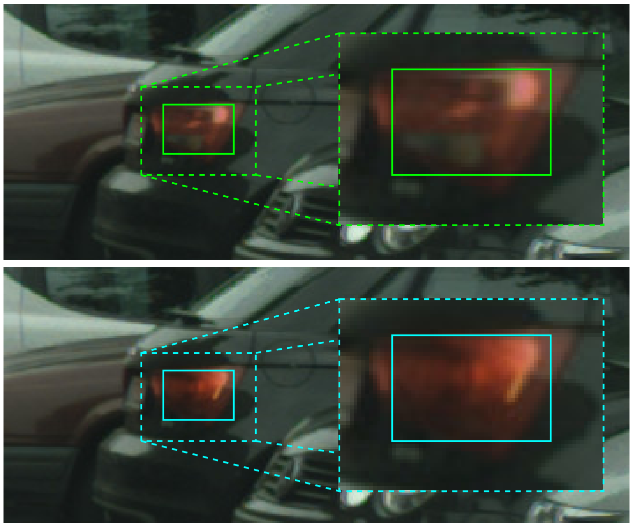

Figure 3. **An indicator which is falsely detected as face**. The solid line is the detected bounding box and the dotted line is a crop-out for visualization purpose. The Figure shows that our pipeline can accurately generate even non-facial contents without any specialized fine-tuning.

4.4. Embedding Distance as Measure for Anonymization Level

In this experiment we validate the face recognition ability after applying a face anonymization method. We use a recent face recognition framework to analyze face embeddings. These embeddings are high dimensional latent representations of faces generated by a convolutional neural network designed. Various distance metrics can be used between the face embeddings to determine similar faces. In contrast to face recognition, for anonymization, embeddings of a face and of the corresponding anonymized version should be dissimilar.

Experimental setup. We use the LightFace [32] framework for detecting and analyzing faces. As in previous experiments, RetinaFace is used for detection. Face embeddings are generated using a VGG-Face [25] model. Respectively, the generated embeddings are 2622 dimensional vectors.

Evaluation metrics. As a similarity measure, we use L2 normalized euclidean distance (L2), since it is considered as a robust metric in this context [31]. The L2 distance is calculated by first l2-normalizing both embedding vectors and then calculating the euclidean distance. Initially, faces are detected in the original images from the Cityscapes train set. We then compare these facial areas with those from the anonymized images and calculate the metrics over the entire dataset.

Results. Overall, the distances are relatively small, which is due to the fact that the majority of faces in this dataset is very small (face $< 32^2$ pixels). Figure 4 shows the distance for all metrics in a histogram. DeepPrivacy1 and our method have a comparable distribution over the L2 distance. But there is a big peak coming from images with where $L_2 = 0$. We assume this peak results from faces, which are not detcted by the DeepPrivacy pipeline and therefore not anonymized at all. The distribtion of DeepPrivacy2 is shifted to the right in comparison to our method. This may be caused by the poor quality of the in-paintings, where unrealistic faces with mismatched skin colors are created (cf. Figure 2). The pixelation method yields the highest distances, which makes sense, because the faces should not be recognizable at all. This means, there is a trade-off between realistic anonymization and making it difficult for CNNs to recognize anonymized faces.

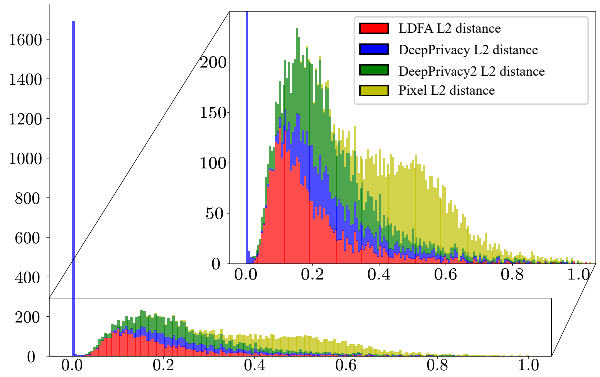

Figure 4. **Embedding distance as measure for anonymization level.** The histogram shows the L2 distance of each anonymization method over the dataset.

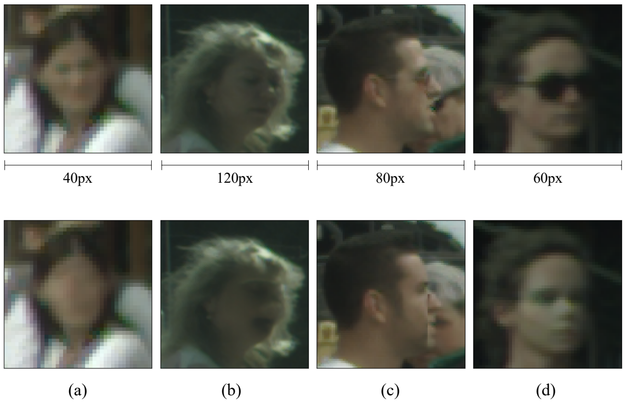

Figure 5. **Limitations of our pipeline**. The top row shows the input images, the bottom row the corresponding face in-painting generated by our method.

4.5. Limitations

As shown in Table 2, face detectors achieve higher mAP scores for faces anonymized by our pipeline compared to naive or recent GAN-based anonymization methods. Nevertheless, there is still a difference of 0.3246 in the mAP scores compared to the ground truth from non-anonymized faces. Figure 5 shows some limitations of our pipeline that can lead to this discrepancy. For small faces (area $< 32^2$ pixels), the LDM used in our pipeline may generate in-paintings that resemble blurry faces (a). In rare cases, the used LDM generates severely deformed faces (b). Accordingly, these faces can no longer be detected after anonymization. The subfigures (c) and (d) show limitations that do not affect detection, but appear unrealistic. Overlapping bounding boxes may lead to artifacts (c). The removal of sunglasses or glasses may generate in-paintings with mismatching skin colors (d).

5. Conclusion

We presented a two stage anonymization pipeline that integrates state-of-the-art diffusion models into an anonymization task. We clearly exhibited that our method is better suited for anonymization, because the recall of the first stage face recognition can be tuned. This leads to a higher NoA than comparable methods. In combination with the general LDM the impact of false positives on the overall anonymization task is negligible. To compare anonymization methods, we trained segmentation methods on anonymized data and evaluated them on non-anonymized data. Our experiment revealed that our method is better suited to anonymize data for segmentation than naive methods and performed comparably with recent GAN-based methods. Additionally, we showed that the mAP score for face recognition tasks improves drastically in comparison to other recent GAN-based methods. The success on face anonymization indicates that this method could be extended to full-body anonymization. This should reduce the recognition capability significantly, but the impact on segmentation tasks needs to be investigated.

Acknowledgment

This work was supported from the project KI Data Tooling (19A20001O) funded by the German Federal Ministry for Economic Affairs and Climate Action (BMWK). We also acknowledge funding by the Karlsruhe School of Optics and Photonics (KSOP).

References

[1] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-to-end object detection with transformers. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part I 16*, pages 213–229. Springer, 2020.

[2] Bowen Cheng, Ishan Misra, Alexander G Schwing, Alexander Kirillov, and Rohit Girdhar. Masked-attention mask transformer for universal image segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1290–1299, 2022.

[3] 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 *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3213–3223, 2016.

[4] Jiankang Deng, Jia Guo, Yuxiang Zhou, Jinke Yu, Irene Kotsia, and Stefanos Zafeiriou. Retinaface: Single-stage dense face localisation in the wild. *CoRR*, abs/1905.00641, 2019.

[5] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*, 2020.

[6] European Union and Proton AG. What is gdpr, the eu's new data protection law? https://gdpr.eu/what-is-gdpr, 2020. Accessed: 2023-01-17.

[7] Jakob Geyer, Yohannes Kassahun, Mentar Mahmudi, Xavier Ricou, Rupesh Durgesh, Andrew S Chung, Lorenz Hauswald, Viet Hoang Pham, Maximilian Mühlegg, Sebastian Dorn, et al. A2d2: Audi autonomous driving dataset. *arXiv preprint arXiv:2004.06320*, 2020.

[8] 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.

[9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[10] 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.

[11] Håkon Hukkelås, Rudolf Mester, and Frank Lindseth. Deepprivacy: A generative adversarial network for face anonymization. In *International symposium on visual computing*, pages 565–578. Springer, 2019.

[12] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017.

[13] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8110–8119, 2020.

[14] 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, 2021.

[15] Zhifeng Kong, Wei Ping, Jiaji Huang, Kexin Zhao, and Bryan Catanzaro. Diffwave: A versatile diffusion model for audio synthesis. *arXiv preprint arXiv:2009.09761*, 2020.

[16] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5060–5069, 2019.

[17] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep learning face attributes in the wild. In *Proceedings of the IEEE international conference on computer vision*, pages 3730–3738, 2015.

[18] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019.

[19] Andreas Lugmayr, Martin Danelljan, Andres Romero, Fisher Yu, Radu Timofte, and Luc Van Gool. Repaint: Inpainting using denoising diffusion probabilistic models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11461–11471, 2022.

[20] Shitong Luo and Wei Hu. Diffusion probabilistic models for 3d point cloud generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2837–2845, 2021.

[21] Patrick Mannion. Vulnerable road user detection: state-of-the-art and open challenges. *arXiv preprint arXiv:1902.03601*, 2019.

[22] Maxim Maximov, Ismail Elezi, and Laura Leal-Taixé. Ciagan: Conditional identity anonymization generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5447–5456, 2020.

[23] Gerhard Neuhold, Tobias Ollmann, Samuel Rota Bulo, and Peter Kontschieder. The mapillary vistas dataset for semantic understanding of street scenes. In *Proceedings of the IEEE international conference on computer vision*, pages 4990–4999, 2017.

[24] Natalia Neverova, David Novotny, Marc Szafraniec, Vasil Khalidov, Patrick Labatut, and Andrea Vedaldi. Continuous surface embeddings. *Advances in Neural Information Processing Systems*, 33:17258–17270, 2020.

[25] Omkar M Parkhi, Andrea Vedaldi, and Andrew Zisserman. Deep face recognition.

[26] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. Hierarchical text-conditional image generation with clip latents. *arXiv preprint arXiv:2204.06125*, 2022.

[27] Aditya Ramesh, Mikhail Pavlov, Gabriel Goh, Scott Gray, Chelsea Voss, Alec Radford, Mark Chen, and Ilya Sutskever. Zero-shot text-to-image generation. In *International Conference on Machine Learning*, pages 8821–8831. PMLR, 2021.

[28] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10684–10695, 2022.

[29] Chitwan Saharia, William Chan, Saurabh Saxena, Lala Li, Jay Whang, Emily Denton, Seyed Kamyar Seyed Ghasemipour, Burcu Karagol Ayan, S Sara Mahdavi, Rapha Gontijo Lopes, et al. Photorealistic text-to-image diffusion models with deep language understanding. *arXiv preprint arXiv:2205.11487*, 2022.

[30] Lukas Schnabel, Stephan Matzka, Martin Stellmacher, Michael Pätzold, and Elmar Matthes. Impact of anonymization on vehicle detector performance. In *2019 Second International Conference on Artificial Intelligence for Industries (AI4I)*, pages 30–34. IEEE, 2019.

[31] Sefik Ilkin Serengil and Alper Ozpinar. Lightface: A hybrid deep face recognition framework. In *2020 Innovations in Intelligent Systems and Applications Conference (ASYU)*, pages 23–27. IEEE, 2020.

[32] Sefik Ilkin Serengil and Alper Ozpinar. Hyperextended lightface: A facial attribute analysis framework. In *2021 International Conference on Engineering and Emerging Technologies (ICEET)*, pages 1–4. IEEE, 2021.

[33] Benjamin Wilson, William Qi, Tanmay Agarwal, John Lambert, Jagjeet Singh, Siddhesh Khandelwal, Bowen Pan, Ratnesh Kumar, Andrew Hartnett, Jhony Kaesemodel Pontes, et al. Argoverse 2: Next generation datasets for self-driving perception and forecasting. *arXiv preprint arXiv:2301.00493*, 2023.

[34] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid scene parsing network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2881–2890, 2017.

[35] Jingxing Zhou and Jürgen Beyerer. Impacts of data anonymization on semantic segmentation. In *2022 IEEE Intelligent Vehicles Symposium (IV)*, pages 997–1004. IEEE, 2022.