UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE INFORMATICA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

PEDRO KEMPTER BRANT

Title: Gaze Preservation on Artificially Generated Faces for Privacy Compliance

Recife

2025

PEDRO KEMPTER BRANT

Title: Gaze Preservation on Artificially Generated Faces for Privacy Compliance

Trabalho apresentado ao Programa de Pós-graduação em Ciência da Computação do Centro de Informática da Universidade Federal de Pernambuco, como requisito parcial para obtenção do grau de Mestre Profissional em Ciência da Computação.

Área de Concentração: Computação Inteligente

Orientador (a): Verônica Teichrieb

Coorientador (a):Lucas Silva Figueiredo

Recife

2025

.Catalogação de Publicação na Fonte. UFPE - Biblioteca Central

Brant, Pedro Kempter.
   Gaze Preservation on Artificially Generated Faces for Privacy Compliance / Pedro Kempter Brant. - Recife, 2025.
   72f.: il.

   Dissertação (Mestrado) - Universidade Federal de Pernambuco, Centro de Informática, Programa de Pós-Graduação em Ciência da Computação, 2025.
   Orientação: Veronica Teichrieb.
   Coorientação: Lucas Silva Figueiredo.
   Inclui referências.

   1. Anonimização facial; 2. Privacidade; 3. GANs; 4. Estimativa de olhar. I. Teichrieb, Veronica. II. Figueiredo, Lucas Silva. III. Título.

UFPE-Biblioteca Central

Pedro Kempter Brant

“Gaze Preservation on Artificially Generated Faces for Privacy Compliance”

Dissertação de mestrado apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Mestre em Ciência da Computação. Área de Concentração: Inteligência Computacional

Aprovada em: 18/12/2024.

_____

Orientadora: Profa. Veronica Teichrieb

BANCA EXAMINADORA

_____

Prof. Dr. Divanilson Rodrigo de Sousa Campelo
Centro de Informática / UFPE

_____

Prof. Dr. George Augusto Valença dos Santos
Departamento de Computação / UFRPE

_____

Prof. Dr. Lucas Silva Figueiredo
Departamento de Computação / UFRPE
(**coorientador**)

Essa dissertação é o resultado de todo o esforço e de tudo que precisei superar para concluir o mestrado. Agradeço à minha orientadora, Verônica Teichrieb, e ao meu coorientador, Lucas Silva Figueiredo, pelo apoio e pelos conselhos ao longo dessa trajetória. Dedico também esta conquista àqueles que me apoiaram e acreditaram em mim, mesmo quando nem eu mesmo acreditava. Meus sinceros agradecimentos à minha namorada, Mariana; aos meus familiares, especialmente aos meus pais, Maurício e Dominique; à minha irmã, Giovanna; à minha avó, Madalena; e à minha gata, Hase. Em memória dos meus avós, Pedro, Carin e Moacyr, e da minha cadela, Marie. Também estendo meus agradecimentos aos amigos e colegas de escola, graduação e pós-graduação. Por fim, agradeço ao Voxar Labs por sempre agir pensando no melhor para mim.

ACKNOWLEDGEMENTS

The authors would like to thank Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) (process 88887.713334/2022-00) for partially funding this research.

RESUMO

Muitos geradores de face são baseados em técnicas como GANs (*Generative Adversarial Networks*) ou modelos de difusão, que podem criar artificialmente rostos humanos realistas e de aparência natural. Estas técnicas têm um grande potencial para a privacidade, pois podem substituir os identificadores biométricos de alguém. No entanto, ao substituir o rosto, grande parte da utilidade da imagem, como estimar a direção do olhar, também pode ser perdida. Embora a estimativa do olhar seja crucial em algumas aplicações, como no monitoramento do motorista ou de pedestres em cenários automotivos, a literatura sobre geradores de faces não possui métricas ou *benchmarks* relacionados a esse assunto. Desenvolvemos o MetaGaze, um conjunto de dados anotado com quase 70.000 imagens de 30 rostos sintéticos de modelos de pessoas disponíveis no *MetaHuman*, na plataforma *Unreal*. Analisamos duas técnicas populares de geração de rosto, DeepPrivacy2 e GANonymization, usando nosso conjunto de dados, MetaGaze, e um conjunto de dados de direção de olhar no contexto veicular, DMD, junto com um estimador de olhar , L2CS, para medir a preservação do olhar. Aplicamos duas estratégias para melhorar a preservação do olhar: modificamos a entrada condicional da técnica base e fizemos *fine-tuning* no modelo do GANonimização, adicionando nosso conjunto de dados ao treinamento para aumentar a diversidade de ângulos de olhar disponíveis no conjunto de treinamento. Nossos experimentos demonstraram que o *fine-tuning* com MetaGaze reduziu o erro absoluto médio na preservação do olhar de 10.8° graus para 7.9° graus em *pitch* e de 6.4° graus para 5.9° graus em *yaw* em comparação com o modelo original de GANonymização. Além disso, indicamos que os cenários mais desafiadores para a preservação do olhar são ângulos de câmera acima de 10° graus, direções do olhar acima de 30° graus, FOV de 60° graus e olhos semicerrados. O conjunto de dados, MetaGaze, está disponível de forma pública em <www.zenodo.org/records/13345194>.

Palavras-chaves: Anonimização facial. Privacidade. GANs. Estimativa de olhar.

ABSTRACT

Many face generators are based on techniques such as GANs (Generative Adversarial Networks) or diffusion models, which can artificially create realistic and natural-looking human faces. These techniques have great potential for privacy, as they can replace someone's biometrical identifiers. However, by substituting the face, most of the image utility, like estimating the gaze direction, could also be lost. Even though gaze estimation is crucial in some applications, like monitoring the driver or pedestrians in automotive scenarios, the literature on face generators does not have metrics or benchmarks related to this matter. We developed MetaGaze, an annotated dataset with almost 70,000 images from 30 synthetic faces from premade person models, available on MetaHuman inside the Unreal engine. We analyzed two popular face generator techniques, DeepPrivacy2 and GANonymization, using our MetaGaze dataset and a vehicular gaze dataset, DMD, along with a gaze estimator, L2CS, to measure gaze preservation. We applied two strategies to improve gaze preservation: modified the conditional input from the base technique and fine-tuned the GANonymization model, adding our dataset to the training to enhance the diversity of gaze angles available on the training set. Our experiments demonstrated that fine-tuning with MetaGaze reduced the mean absolute error in gaze preservation from 10.8° degrees to 7.9° degrees in pitch and 6.4° degrees to 5.9° degrees in yaw compared to the original GANonymization model. Besides, we indicate that the most challenging scenarios for gaze preservation are camera angles above 10° degrees, gaze directions above 30° degrees, FOV of 60° degrees, and eyes semi-closed. The dataset, MetaGaze, is publicly available at <www.zenodo.org/records/13345194>.

Keywords: Face anonymization. Privacy. GANs. Gaze estimation.

LIST OF FIGURES

LIST OF TABLES

CONTENTS

1 INTRODUCTION

In this Chapter, we start by motivating the dissertation (Section 1.1), explaining the current anonymization solutions (Section 1.2) to solve our selected problem, "Does training face generators with synthetic datasets improve their utility for preserving gaze direction during face anonymization?", our contributions (Section 1.4), some basic GANs 1.3 and how the dissertation is structured (Section 1.5). We are particularly interested in the automotive scenario, but the dataset we developed and the models trained on this work can be applied to any application concerned with gaze direction preservation in anonymization.

1.1 MOTIVATION

Computer vision approaches toward privacy have become an increasing topic of interest over the last few years (XIANG, 2022a). There are several concerns regarding the misleading usage of captured images that would benefit from vision-based solutions. Cases include leaked sensitive images from home security cameras of wrongfully accused and convicted people due to the poor accuracy of some face detection and recognition algorithms, which, in addition, more than often present a racial bias (YANG et al., 2022).

In particular, considering the autonomous vehicle sector, there are leakage cases where Tesla workers share sensitive images recorded by customers' cars (KOLODNY, 2023). In Germany, Volkswagen was fined $1.1M euros due to violations of the GDPR regarding unauthorized camera data collection (Compliance Week, 2023). A report from Mozilla Foundation - Privacy not Included 2023 - reviewed over 25 brands of cars and states that they do not present valuable privacy strategies. The report concludes that all brands "collect too much personal data" and "84% share or sell your data" (Mozilla Foundation, 2023).

By applying computer vision techniques, users can be protected from malicious usage of captured images. A set of techniques anonymizes users by applying noise and blur to sensitive parts of the image (ADEBOYE et al., 2022; BRKIć; HRKAć; KALAFATIć, 2017; CONINCK et al., 2024; BERA; KHANDEPARKAR, 2023). However, the utility of those images is deprecated, and information about what behavior occurred in those scenes is lost in the process. For example, data such as whether the users were happy or afraid, attentive or distracted, and so on can be lost.

As an alternative, several methods utilize Generative Adversarial Networks (GANs) to anonymize users. These techniques replace representative portions of the image with generated places, cars, or users, effectively hiding users' identities while preserving the observable behaviors. This approach allows for further image analysis and maintains the utility of the images, offering an optimistic outlook on the future of image anonymization.

In this work, we are particularly interested in the autonomous vehicular sector and how to effectively prevent those leakages and privacy violations. We focus on a study that aims to maintain the utility of anonymized faces by preserving the direction of drivers' gaze. The direction of a driver's gaze is a crucial factor in the automotive context, as it can indicate where their attention is focused on the road, rearview mirrors, infotainment systems, or passengers. This research underscores the significance of understanding and preserving the direction of drivers' gaze.

This way, we leverage features such as the direction of the eyes in our study and different degrees of eyelid opening, camera positions, and camera opening angles, representing ordinary and fisheye cameras (commonly used in the vehicle interior (GEYER et al., 2020; BAEK et al., 2021)).

For this purpose, we created an annotated synthetic dataset with proportional distributions of gaze directions and eyelid openness to support the evaluation of state-of-the-art gaze estimation techniques applied on top of anonymized faces. Our dataset is entitled MetaGaze. Furthermore, we evaluated state-of-the-art anonymization techniques on the proposed dataset to assess the resulting performance regarding its gaze preservation capabilities. Finally, we propose fine-tuning procedures and two tailoring methods for the input mesh to attain better accuracy regarding the estimated gaze of the anonymized face.

1.2 ANONYMIZATION TECHNIQUES

There are different ways to achieve user anonymity. We delimit four core strategies, elicit the pros and cons, and use the associated threat model to determine which approach would best preserve the gaze direction from the non-anonymized faces in our context. Figure 1 shows the steps where Anti-Facial Recognition technology (AFR) can occur. Alternative sensors avoid the first step by avoiding collecting all the information, while GANs, Noise, and Blur techniques are used in step 2 as they anonymize the image collected in step 1. Lastly, Differential Privacy is used to avoid step 5, which aims to protect the dataset from query attacks. Categories 3

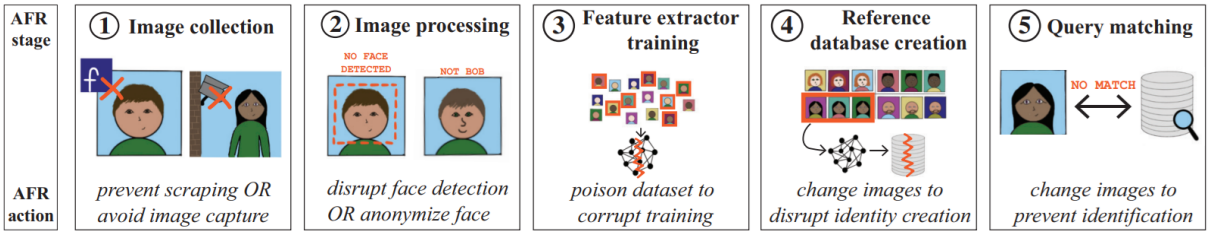

Figure 1 – Anti-Facial Recognition technology (AFR) focuses on disturbing the face recognizer capabilities. To this end, the disturbance can occur in any of the five face recognition steps. From the image collection or processing (1 and 2) to poisoning the dataset already gathered (3 and 4) (WENGER et al., 2023) or even in database query base attacks (5). Image reused from "Sok: Anti-facial recognition technology" (2023).

and 4 were not considered in this dissertation as they assume access to the training dataset of the target model.

Considering the vehicular context, we grouped the anonymization solutions into four categories of solutions, their respective threat models, and pros and cons:

1.2.1 Alternative Sensors

The use of alternative sensors to detect pedestrians by LiDAR (point cloud) (OHNO et al., 2024), thermal sensors (MY et al., 2019), or even by its skeleton (NEFF et al., 2020) are the possible solutions. The concept is not to capture information that could be useful for identification models, such as facial traits; instead, only information that can distinguish people from objects is used.

Threat Model Three scenarios are considered: first, unintended access of images; second, authorized access; third, someone physically accessing the edge device and collecting identifiable material from recorded data.

Pros and Cons Even though being a relatively simple method, it discards potentially useful information from the source image/video, and there is also no guarantee that the re-ID is entirely avoidable. The attacker can use gait, posture, or height techniques to differentiate and identify an individual.

1.2.2 GANs for Image Modification

This group of networks receives the untreated photo or video as input containing locations (XIONG et al., 2019), cars or people (LI et al., 2023) or faces (CAI et al., 2024) and outputs a

new image that can no longer be used by another application to retrieve the original subject. Our work is based on GANs and uses (HELLMANN et al., 2024) as a base technique, so we are inherently included in this strategy.

Threat Model Any external party potentially receiving or intercepting the image/video after transformation is considered untrusted. The adversary is assumed to have access to the training set and the model's architecture but not the random number generator (RNG).

Pros and Cons It is considered the best solution for edge applications as it provides complete anonymization while potentially preserving another attribute. However, it is the most computationally expensive solution and thus is unsuitable for running on some edge hardware.

We have selected GANs as a technique to solve privacy issues mainly because they are suitably effective in removing identifiers while maintaining image utility. We also consider their versatility and opportunities for deeper exploration in the image generator field of research.

1.2.3 Noise and Blur

Simple image processing techniques that can protect information like location (ADEBOYE et al., 2022), people (BRKIć; HRKAć; KALAFATIć, 2017; CONINCK et al., 2024) or vehicles and other subjects (BERA; KHANDEPARKAR, 2023). More advanced techniques can use style transfer to preserve some useful data.

Threat Model As in "GANs for Image Modification," any external party potentially receiving or intercepting the image/video after transformation is considered untrusted.

Pros and Cons This is the opposite of GANs because image processing usually has a low computational cost. Still, the process can potentially remove valuable data from the input image. Besides, in some cases, the anonymization can be reversible (the adversary can restore the original image).

1.2.4 Federated Learning, Differential Privacy, and Cryptography

The collected data should be transmitted to a server for processing, which can leak sensitive data during transmission or even on the server. Federated learning acts distributing the image processing to do part of it locally rather than on a server. This idea prevents the risk of centralized data breaches. Cryptography applications are based on mathematical techniques for guaranteeing data confidentiality, integrity, and authentication (ZHOU et al., 2022b; BAI

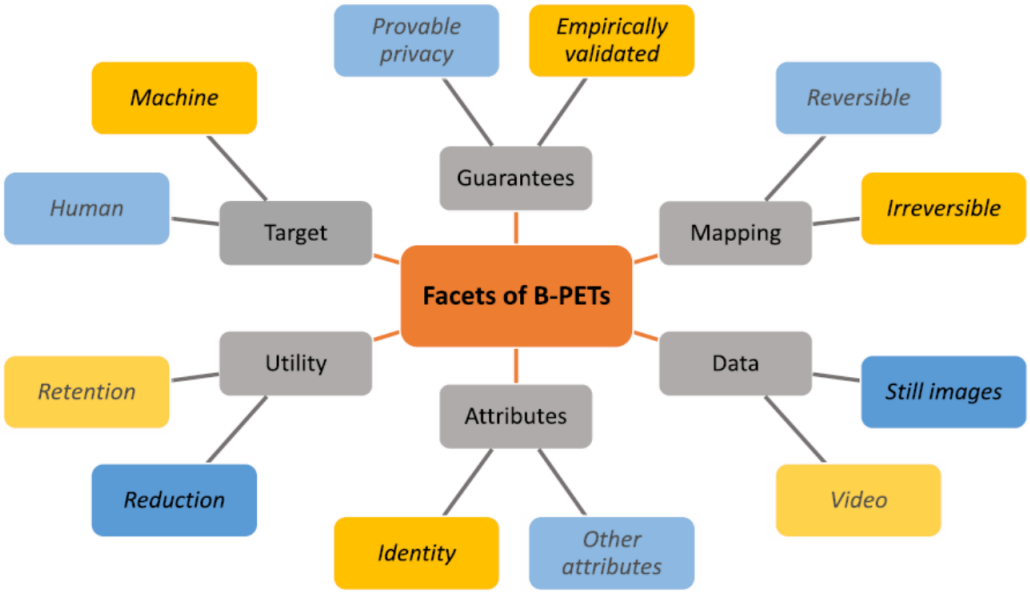

Figure 2 – B–PETs (Biometric Privacy Enhancing Techniques) can be classified using this taxonomy. This taxonomy is useful to delimit our techniques' targets (highlighted) and limitations. Image reused from "Privacy–Enhancing Face Biometrics: A Comprehensive Survey" (2021).

et al., 2023; BAI; FU; YANG, 2022). In addition, Differential Privacy can be defined as "the increased risk to one's privacy incurred by participating in a database" (DWORK, 2006), which means that the very fact that someone's data is in a dataset makes them vulnerable to data leaked. These techniques can also be applied independently from the previous methods (as a complementary or standalone privacy method).

Threat Model It considers honest-but-curious applications, which operate legitimately and can access public data but attempt to extract as much sensitive information as possible. These applications may eavesdrop on transmissions and potentially gather sensitive information from edge applications.

Pros and Cons Cryptography or differential privacy techniques guarantee the non-reversibility of the data mathematically; however, those two techniques, along with the federated learning process, are usually difficult to implement on a full application pipeline. The computational cost and its complexity can be too high for some applications.

Considering Figure 2, the use of GANs has the following features: we have empirically validated guarantee, as opposed to **Federated Learning, Differential Privacy, and Cryptography** that is mathematically proved. Our mapping is irreversible, which is a good privacy trait. In our case, we work with still image data. The attribute we want to protect is the identity. Concerning utility, we want to reduce instead of retention, as in **Noise and Blur**.

Finally, our main target is to avoid machine face recognition. However, face swaps also avoid human identification.

1.3 BASIC GANS CONCEPTS

This work uses GANs to generate images; this section will explain the main concepts to understand this process and some peculiarities of the chosen architecture.

GAN architecture: Figure 3 exemplifies how a general GAN works. Two networks work as "adversaries. The first is the generator, which is responsible for taking input, like random noise, and then trying to replicate a given distribution from another set of images (for example, faces or cats). The discriminator will receive an image from the real dataset or the generator. It works as a classifier to determine which images are fake and real (this step can be clearer, as seen in Figure 4).

U-Net: Figure 5 shows the original U-Net architecture. Its name comes from the "u" like form of the network. The network was designed for biomedical image segmentation (RONNEBERGER; FISCHER; BROX, 2015) but was also adopted as a generator in some GANs architectures, as in pix2pix (ISOLA et al., 2017). The U-Net works as follows: the image from the input is reduced in dimension by convolutional and max pooling layers until it becomes a feature vector. Upsampling layers and skip connections (copies of some high-frequency details from the image to make the image reconstruction fast) are used to reconstruct an image from the feature vector as output.

PatchGAN: Figure 6 shows the PatchGAN (DEMIR; UNAL, 2018) differential as a discriminator in GANs, especially in pix2pix, where it was first applied. The main difference from a traditional discriminator, as in Figure 4, is that instead of the output from the image is 0 or 1 for how close to the real dataset it looks, the output is a matrix of 0s and 1s. Each value in this matrix represents the evaluation of a patch from the image (preserving its spatial consistency). This change makes it easier to determine which parts or patches of the images are more "real." It is beneficial in pix2pix architecture once the spatiality of the image is essential, and the generator can focus its adjustments on patches that are not so real.

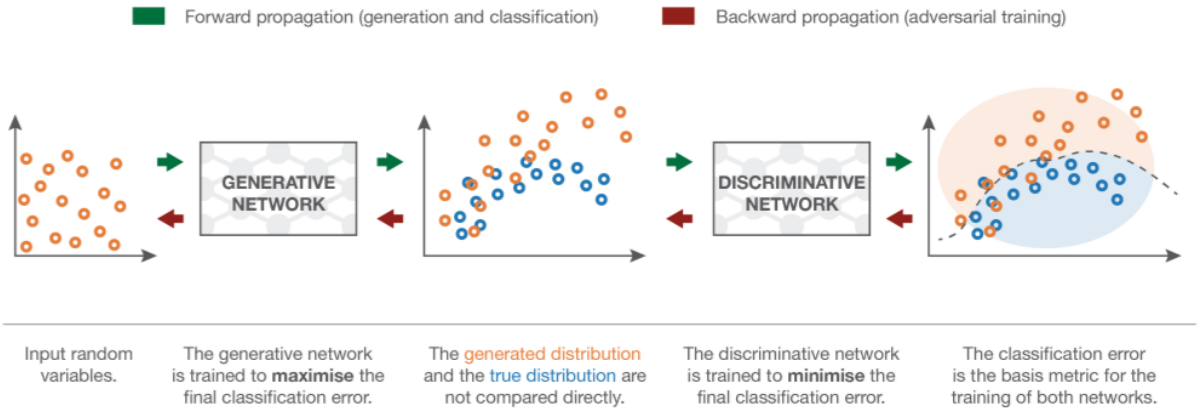

Figure 3 – Illustration from the two adversarial networks in GANs: generator and discriminator. Image from medium.com/towards-data-science/understanding-generative-adversarial-networks-gans-cd6e4651a29.

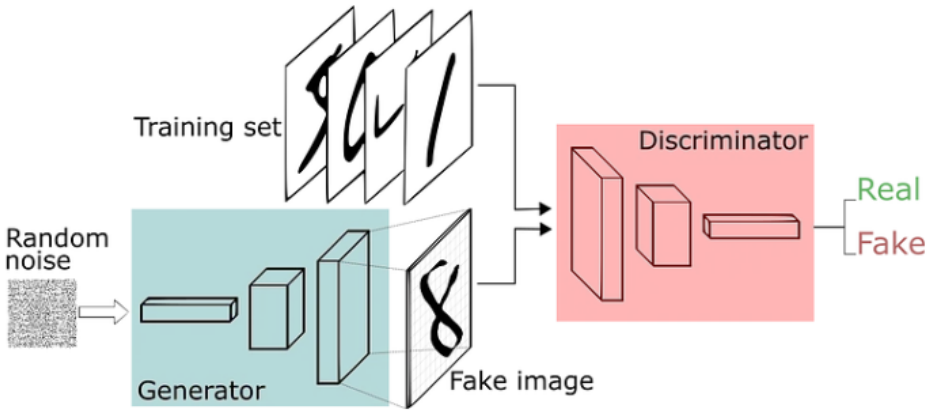

Figure 4 – Fluxogram of a general GAN improving the generation of a number handwrite. Image from "A Short Introduction to Generative Adversarial Networks" 2017

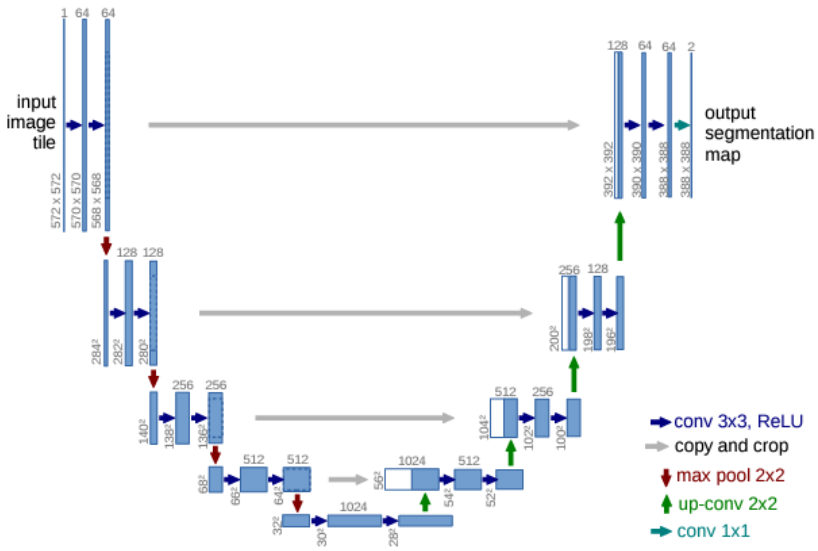

Figure 5 – U-net architecture. Image from "U-Net: Convolutional Networks for Biomedical Image Segmentation" 2015

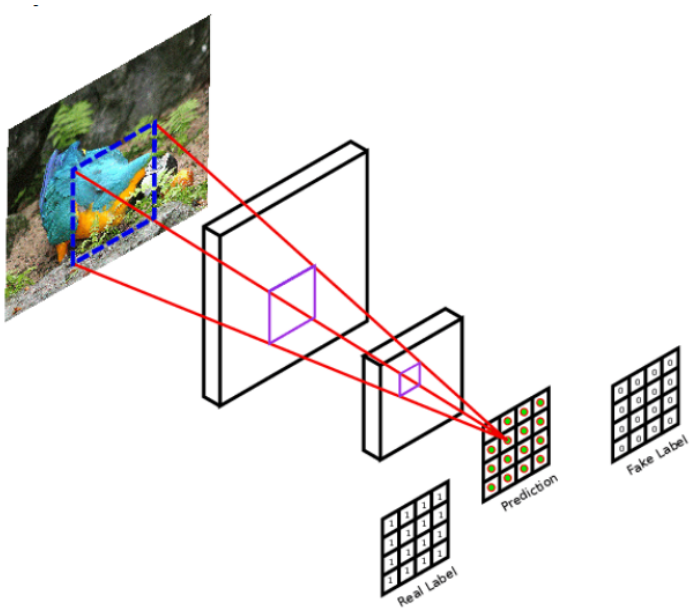

Figure 6 – Patchgan discriminator. Image from "Patch-Based Image Inpainting with Generative Adversarial Networks"

1.4 CONTRIBUTIONS

To answer our research question, "Does training face generators with synthetic datasets improve their utility for preserving gaze direction during face anonymization?", our main contributions are as follows:

- We design a annotated synthetic dataset to validate gaze preservation on anonymized faces. With almost 70000 faces, we achieve high variability by combining 30 characters that contain 3 Fields of Views (FOVs), 25 camera angle variations, 15 gaze directions, and 5 eyelid openness variations. The dataset is available at <www.zenodo.org/records/13345194>.
- We compare the main face anonymization techniques (DeepPrivacy2 and GANonymization) regarding our gaze preservation benchmark and Driver Monitoring Dataset (DMD);
- We leverage the results on the benchmark by training the GANonymization technique along with fine-tuning or input modifying strategies;
- We point out the scenarios attributes where the evaluated methods have more difficulty preserving gaze.

1.5 DISSERTATION STRUCTURE

After the introduction, this dissertation is arranged into five more chapters to facilitate the reader's flow. The chapters are as follows: On Related Work (Section 2), we gathered relevant papers on Generative Face Anonymizers (Section 2.1), solutions for Privacy in Automotive Camera Systems (Section 2.2). In addition, we split data set references into three categories: Gaze Datasets (Section 2.3), Face Generation Training Sets (Section 2.5) and Synthetic Datasets (Section 2.4). For Preliminary Works (Section 3), we start by detailing Problem Statement (Section 3.1), where we developed Priface (Section 3.2), a pipeline to enhance and measure face quality. To this end, we also display our Experiments (Section 3.3) and Findings (Section 3.4). Lastly, Exploring GANonymization (Section 3.5) explains why we shifted the original research problem and technology. In Methodology (Section 4), we describe our proposed dataset MetaGaze (Section 4.1). We also detail the base technique, GANonymization (Section 4.2), and our two improvement approaches: Fine-Tuning Variations (Section 4.3) and

Mesh Variations (Section 4.4). In addition, in the Experiments (Section 4.5), we detail the choices on datasets, methods, metrics, and implementation details. In Results (Section 5), we have done a Quantitative Analysis Section 5.1), along with Further Quantitative Analysis 5.2 to dig deeper into the results by comparing the variations by each of the dataset attributes. We also did a Qualitative Analysis (Section 5.3) to evaluate the outputs visually. Finally, the main points are summarized in Section 5.4. Our Conclusion (Chapter 6) summarizes the dissertation, along with a list of limitations and future work identified.

2 RELATED WORK

In this Chapter, we gathered relevant papers on Generative Face Anonymizers (Section 2.1), solutions for Privacy in Automotive Camera Systems (Section 2.2). Besides, we collected datasets that are divided into three categories: Gaze Datasets (Section 2.3), Synthetic Datasets (Section 2.4), and Face Generation Training Sets (Section 2.5).

2.1 GENERATIVE FACE ANONYMIZATION

Regarding architecture and capacity, GANs have demonstrated impressive capabilities in generating realistic, high-resolution images (GOODFELLOW et al., 2014a; KARRAS; LAINE; AILA, 2019). Diverse methods (REN; LEE; RYOO, 2018; LI; LIN, 2019; WU et al., 2019; WEN et al., 2022; SUN et al., 2018) have utilized GANs for face anonymization, leveraging their ability to learn and replicate the distribution of training data. The use of GAN architectures such as DCGANs (SUN et al., 2018), StarGAN (LI; LIN, 2019), and StyleGAN (SHAMSHAD; NASEER; NANDAKUMAR, 2023; JIANG et al., 2023; ZHAO et al., 2020), as well as StyleGAN2 (KARRAS et al., 2020; KARRAS et al., 2020; SKOROKHODOV; TULYAKOV; ELHOSEINY, 2022) and StyleGAN3 (KARRAS et al., 2021; BODDETI; SREEKUMAR; ROSS, 2023; QIU et al., 2022; FARD et al., 2023; HELLMANN et al., 2024), has significantly advanced face anonymization techniques. These developments have enhanced visual quality, preserving the realism of generated images while effectively obfuscating identity information. However, as the technologies advance in resolution and generation capacity, so is the network size. Therefore, depending on the application limitation, like onboarding processing, the engineer responsible should select techniques less powerful but more practical should be used instead.

Another reference to face generation is the techniques based on diffusion models (HO; JAIN; ABBEEL, 2020). Those models are newer and more robust than GANs based once they can generate more variation between images and with better resolution (HE et al., 2024; MORVAY et al., 2023; KLEMP et al., 2023a). However, diffusion models require immense computational resources and time for training and inference compared to GANs; therefore, we did not investigate further techniques.

2.2 PRIVACY IN AUTOMOTIVE CAMERA SYSTEMS

There are also works focused on anonymization for the vehicular field of application. For instance, the LFDA method (KLEMP et al., 2023b) is applied to anonymize pedestrians' faces using a latent diffusion method. On INSPIRE (LI et al., 2023), the authors propose a technique to synthesize humans and cars, generating full-body synthetic versions of pedestrians. Another framework introduces a privacy-preserving pedestrian analysis in which a wireframe representation takes place over the original user body in the image, preserving the body shape and posture (but not the gaze particularly) (KUNCHALA; BOUROCHE; SCHOEN-PHELAN, 2023). DeepPrivacy2's authors have conducted a study focused on pedestrian datasets to understand if there is any impact regarding training models on anonymized data for the instance segmentation and key points detection tasks (HUKKELÅS; LINDSETH, 2023b). The authors concluded that the performance drop (preserving behavioral characteristics) is significantly reduced if the anonymization is realistic. Despite the results and findings in these works towards optimizing the utility of the generated images, gaze preservation has not been tackled in any of the cited papers.

2.3 GAZE DATASETS

In the automotive scenario, it is crucial to check if the pedestrians (external camera) and the driver (internal camera) are aware of each other, which can be checked by gaze estimation. For external datasets, Gaze360 (KELLNHOFER et al., 2019) covers a wide range of head poses, distances, illumination, occlusion, and other conditions in the wild. P-DESTRE (KUMAR et al., 2020) is a pioneer pedestrian re-id and tracking dataset that collects images using Unmanned Aerial Vehicles (UAVs). For internal environments, many datasets used videos (real and simulated) that were recorded with drivers (GHOSH et al., 2021; ORTEGA et al., 2020; VORA; RANGESH; TRIVEDI, 2018).

Synthetic datasets related to gaze have been developed as well. Some datasets use synthetic eye images to evaluate deep networks for driver gaze estimation (FAHMY et al., 2021). The aim is to discover underrepresented scenarios and retrain the dataset to fix and prevent critical failure cases. In FEXGAN-META (SIDDIQUI, 2022), a synthetic facial expression dataset using MetaHuman (Epic Games, 2023a) was developed because of the scenario's lack of good-quality images. Synthetic data is more reliable and controllable in generating high-quality data that

can evenly represent even classes often underrepresented in real data. However, the datasets available did not explicitly consider variations on FOV (Field of view) in the camera and did not have a good dataset distribution concerning eyelid closeness.

2.4 SYNTHETIC DATASETS

Many deep learning models are performing fine-tuned on synthetic datasets or even doing synthetic training when the model is training exclusively on synthetic data (JOSHI et al., 2024). Some results show that it can be even better than using real data (TREMBLAY et al., 2018). The main advantages come from the performance boost, built by a dataset that has complete controllability and scalability of the datasets along with mitigating privacy concerns as there are no real people in the images.

Synbody (YANG et al., 2023) is a synthetic dataset with 1.2M images of people in a wide range of poses and backgrounds; their annotations can be used even by Neural Radiance Fields (NeRF) models. EmoFace (LIU et al., 2024) is an Audio-driven emotional 3D face animation dataset that can control the rigs for emotion and lipsync in MetaHuman to generate animations. FEXGAN-META (SIDDIQUI, 2022) is a facial expression dataset with 162K images that also uses MetaHuman developed to expand the quantity of good quality facial expression images. Another synthetic dataset that uses MetaHuman (HERASHCHENKO; FARKAŠ, 2023) has 57k images and was used to extend the training set of a gaze estimator and managed to reduce the error in the estimation. In Heatmapbased Unsupervised Debugging of DNNs (HUDD) (FAHMY et al., 2021), using UnityEyes simulator, the authors were able to not only fine-tune the gaze estimator model but also identify the group of images that does not have a good performance and thus, leverage the security and accuracy of the whole model. Those datasets did not explicitly take into account the FOV camera variations, which could be fundamental in certain scenarios, like onboarding car cameras;

2.5 FACE GENERATION TRAINING SETS

Often, GANs and training datasets are developed together. GANonymization (HELLMANN et al., 2024) and CIAGAN (MAXIMOV; ELEZI; LEAL-TAIXÉ, 2020) used CelebA (LIU et al., 2015) for training their models. CelebaHQ (KARRAS et al., 2017) was a similar dataset with higher resolution that was released with Progressive GAN, as it generated high-resolution images.

CelebV-HQ (ZHU et al., 2022) is similar, but for videos. Flickr-Faces-HQ (FFHQ) dataset also has HD quality and was developed along with StyleGAN (KARRAS; LAINE; AILA, 2019). Multi-purpose ExtremePose-Face-HQ dataset (EFHQ) (DAO et al., 2024) was an alternative created to make the models more robust to a wider range of face poses, as extreme poses are not enough represented in their predecessor. DeepPrivacy2 (HUKKELÅS; LINDSETH, 2023a) also developed Flickr Diverse Humans(FDH) and Flickr Diverse Faces 256 (FDF256) datasets because the previous datasets did not cover enough poses for faces or bodies in their training. Those training sets tackle specific gaps from the predecessors or some limitations from their network architecture. However, many of them are not labeled, which may make it difficult to analyze where the models trained are not satisfactory. These datasets also did not consider gaze direction explicitly as an attribute that should be well distributed in their data, which may cause class representation problems (for example, faces looking up not being well generated) and bias in training.

3 PRELIMINARY WORKS

During the development of this master's, we conducted a set of preliminary works that led to our early findings regarding core gaps in image anonymization. We found that no studies have specifically investigated Face Quality Assessment in the context of privacy-preserving images. Therefore, in this Chapter, we detail our first approach to setting the research problem in Problem Statement (Section 3.1), where we developed Priface (Section 3.2), a pipeline to enhance and assess face image quality. To this end, we also displayed our Experiments (Section 3.3) and Findings (Section 3.4). We exhibited the PRIFACE paper on the ICCV 2023 (International Convention on Computer Vision) as an extended abstract at the LXCV (LatinX in Computer Vision) workshop. Lastly, Exploring GANonymization (Section 3.5) explains why we shifted the original research problem and technology.

3.1 BROADER PROBLEM STATEMENT

One of the biggest concerns about privacy violations is the deployment of camera-based monitoring systems in industries and smart cities, coupled with advancements in computer vision (DUFRESNE-CAMARO; CHEVALIER; AHMED, 2020; XIANG, 2022b). Acquired images contain sensitive data that can reveal people's identities and behaviors, leading to issues like undesired advertising, behavior monitoring, and biased arrests. New data protection laws impose obligations on data controllers and empower individuals to control their data (LYNSKEY, 2015; PURTOVA, 2018). Storing anonymized images is a solution to comply with these laws, but traditional privacy techniques like blurring hinder image analysis. Effective protection methods could balance privacy and utility, like fabricating replacement identities using "deep fake" techniques or Generative Adversarial Networks (GANs) for anonymization. However, even though current GANs can generate credible faces in high-resolution frontal face datasets (KARRAS; LAINE; AILA, 2019; LIU et al., 2015), in the early stages of our research we observed a set of gaps for in the wild applications' image quality and, especially, in how to evaluate its quality effectively. Hence, we dedicated our preliminary efforts to delve deeper into two topics to understand how to measure and enhance generated face image quality.

Face Image Restoration. Aiming to improve the quality of the generated anonymous faces, we investigated restoration techniques. A series of face image restoration methods have

been proposed to address various types of facial image degradation (HONG; RYU, 2020; YU; PORIKLI, 2016; ZHANG et al., 2021; ZHU et al., 2016), including low-resolution, noise, and blur (CHEN et al., 2021). However, when dealing with real-world images, they often exhibit poor performance. Several blind face restoration (BFR) approaches have been developed to restore faces without prior knowledge about the type of degradation. (YANG et al., 2021a). Another strategy involves using Transformers (ZHAO et al., 2023; LI et al., 2022; WANG et al., 2022), first capturing the superpixel-wise global dependency and then transferring it into each pixel. Recently, GAN inversion approaches can invert the latent space representation of StyleGAN (WANG et al., 2022; YANG; QUAN; ZHANG, 2021; PAN et al., 2020) and generate high-quality facial images by learning to map a degraded image to its corresponding latent code. Their effectiveness has been demonstrated in various studies (CHEN et al., 2021; YANG et al., 2021a; POIRIER-GINTER; LALONDE, 2023).

Image Quality Assessment. Another related topic of interest is to automatically assess the quality of the generated image. Properly defining if the new anonymous face is well formed is crucial to attribute a utility score to each technique. Image Quality Assessment (IQA) can be categorized into two classes: (1) full-reference IQA (FR-IQA) evaluates the statistical or perceptual similarity between restored and reference images, often relying on handcrafted features and statistical analysis such as mean squared error (MSE), peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), or perceptual quality index (PQI) (SHAO et al., 2015; SHEIKH; BOVIK; VECIANA, 2005; WANG et al., 2004). (2) No-Reference IQA (NR-IQA) uses Convolutional Neural Networks (CNNs) and other deep architectures to capture complex image features and incorporate perceptual aspects of image quality. These models are typically trained on extensive datasets that may include human ratings (LIU; WEIJER; BAGDANOV, 2017; ZHANG et al., 2023). Although previous works have explored the utility of face images (FU et al., 2022), no studies have specifically investigated Face Quality Assessment in the context of privacy-preserving images.

In sum, in our initial endeavors, we explore the effects of applying face enhancement methods and evaluating the generated faces using facial metrics. The main contributions of this work are as follows: (1) A plug-and-play pipeline resulting in an anonymization method to improve the overall quality of GANs-generated images. (2) One of the first evaluations of enhancement face methods and face quality assessment on artificially generated faces (evaluated in the wild scenario on the dataset WIDER (YANG et al., 2016)).

3.2 PRIFACE

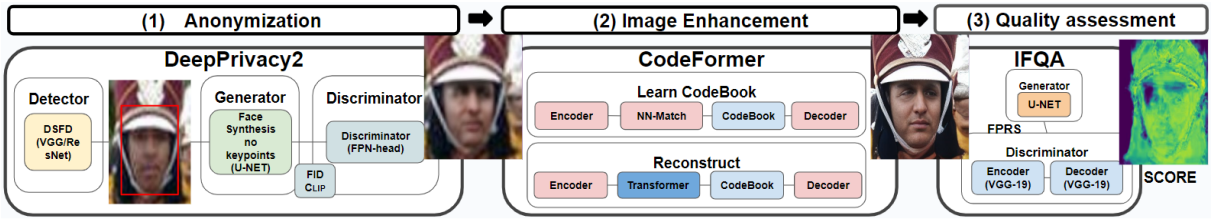

Figure 7 – A detailed view of the PRIFACE pipeline. The first step is *Anonymization*, in which we use DeepPrivacy2 (HUKKELÅS; LINDSETH, 2023a) to generate an anonymized face. This face is then fed to the *Image Enhancement* step, where we employ CodeFormer (ZHOU et al., 2022a), an Encoder-Decoder model that improves the overall quality of the image. Finally, we assess the image quality on *Quality Assessment* by applying the IFQA as a metric that will also output a heatmap highlighting more realistic regions .

As we show in Figure 7, given an input image, we (1) apply DeepPrivacy2 (HUKKELÅS; LINDSETH, 2023a) in the *Anonymization* step, generating a new set of data (which can be one or multiple images) with anonymized faces. We then use this output to the (2) *Image Enhancement* module to improve the quality of the image. Finally, we (3) assess the image quality in the *Quality Assessment* stage.

Face Anonymization Module. We chose DeepPrivacy2 (HUKKELÅS; LINDSETH, 2023a) to integrate our pipeline for two reasons: (1) instead of facial landmarks (often tricky to get on 'in the wild' scenarios), face surroundings and sparse pose information condition the generator; (2) it is capable of full-body anonymization, facilitating a forthcoming expansion on future works. DeepPrivacy2 generator architecture is based on U-Net (RONNEBERGER; FISCHER; BROX, 2015), and the face module is capable of receiving and generating faces up to 256×256 of resolutions (previously 128×128 on DeepPrivacy (HUKKELÅS; MESTER; LINDSETH, 2019)). ProGAN (KARRAS et al., 2017) inspired the growing GAN and upsample a generated face from 4×4 until the final resolution. DeepPrivacy2 has made significant additions, for instance, it concatenates each decoder layer with pose and surrounding information and transforms the previous network on conditional GANs. The network was trained on the FDF dataset (HUKKELÅS; LINDSETH, 2023a) proposed by the same authors. This dataset has 1.5M faces with challenging positions, illuminations, and occlusion factors, expanding the generated face range of action in diverse conditions.

Face Enhancement Module. CodeFormer (ZHOU et al., 2022a) was developed to solve the blind face restoration problem as an encoder-decoder method. They use a Transform-based prediction network capable of identifying facial regions. They restored face regions from

different degradation levels with a learned expressive codebook space. The applications go from noisy images to uncanny or low-resolution generated faces. CodeFormer has a hyperparameter 'w' in the range 0-1 that can prioritize either fidelity or quality. As the enhancement is applied to the generated faces, we have no objective of preserving any specific face trace. So, 'w' was set to 0 on all experiments, prioritizing quality over fidelity.

Face Quality Assessment Module. Interpretable face quality assessment (IFQA) (JO et al., 2023) is a face-centric metric (considers the face's primary regions, such as the eyes, nose, and mouth.) for assessing the perceptual quality of computationally generated faces. The authors show in their experiments that IFQA is highly correlated with human judgment regarding face naturality and quality.

The IFQA framework is based on adversarial learning. The generator consists of an encoder-decoder architecture. The discriminator outputs per-pixel score using a U-Net architecture (SCHöNFELD; SCHIELE; KHOREVA, 2020). This structure allows us to classify each pixel as real or fake. An image-level score or quality score (QS) (see Equation 3.1) can be obtained by aggregating pixel-level scores as follows:

$$QS = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} D_{i,j}^{U}(I), \tag{3.1}$$

where I is an input image, H and W its height and width, and $D_{i,j}^{U}(i)$ is a U-Net-based pixel-level discriminator.

3.3 EXPERIMENTS

Dataset. The WIDER FACE dataset (YANG et al., 2016) has images containing faces with small scale, oblique poses, and occlusion. These challenges represent potential real-world application scenarios. We evaluate our pipeline on the validation subset (10% randomly selected of the dataset), also used by DeepPrivacy (HUKKELÅS; MESTER; LINDSETH, 2019) to measure the impact of the anonymization. The full dataset contains 32203 images and 393703 labeled faces. Even with approximately 3200 images, the WIDER FACE validation subset is comparable with other face detection complete datasets like MALF (YANG et al., 2015) or FDDB (JAIN; LEARNED-MILLER, 2010) with 5250 and 2845 images, respectively.

Configurations and comparisons. We applied our pipeline on the WIDER FACE dataset to generate five dataset variations as outputs that will be referenced as: *Original*: WIDER FACE

set without any modifications, used as baseline; *DP2*: "Original" anonymized with DeepPrivacy2; *PRIFACE*: our proposed pipeline, "DP2" enhanced with CodeFormer; *CF*: "Original" enhanced with CodeFormer; *CF+DP2*: inverted pipeline, "Original" enhanced with CodeFormer and anonymized with DeepPrivacy2.

Validation Metrics. We evaluate the generated dataset variations using two metrics: (1) Fréchet Inception Distance (FID), a metric from the latest layer of an inception neural network that compares the distribution from the generated images with the real set used as ground truth. FID is used for GANs, including as the loss function of DeepPrivacy2. (2) Interpretable Face Quality Assessment (IFQA) was chosen as our Face Quality Assessment Model (see chapter 4) and is used to compare the impact of each module on the dataset. Both IFQA and CodeFormer studies compare themselves with comparative quality metrics, namely PSNR (HUYNH-THU; GHANBARI, 2008), SSIM (WANG et al., 2004), and LPIPS (ZHANG et al., 2018). In our analysis, we specifically use IFQA to evaluate the impact of CodeFormer, given that this metric is closer to human judgment by the authors(JO et al., 2023).

Implementation Details. We used the pre-trained available models and implementations for DeepPrivacy2[1],CodeFormer[2], IFQA [3] and FID (SEITZER, 2020) score for PyTorch[4]. We conducted our experiments on a desktop computer running Ubuntu 20.04 LTS with an Intel Xeon E-2226gp with 32 GB of RAM and a Quadro p1000 with 4GB of VRAM.

3.4 FINDINGS

Quantitative Evaluation. Table 1 displays the FID metric between each dataset and its configurations. FID can be understood as the distance between the features from a set (generated images) and the features from a second set (real); therefore, the smaller the distance, the better (at the beginning of the training, the FID is about 500 and should be minimized). The FID from *Original* and *DP2* is 8.68. This value means that even with the anonymized faces, the dataset still represents similar objects, i.e., faces. *DeepPrivacy2* registered an FID of 1.84 when comparing a validation set of 50,000 images from the FDF dataset and their own anonymized faces output. The second FID is 126.91, between *Original* and *PRIFACE*. The result expresses that the *PRIFACE* generates images more distinct from the *Original* than

[1] https://github.com/hukkelas/deep_privacy2
[2] https://github.com/sczhou/CodeFormer
[3] https://github.com/VCLLab/IFQA
[4] https://github.com/mseitzer/pytorch-fid

Configuration	Original (↓)	DP2 (↓)	PRIFACE (↓)
Original	0.00	–	–
DP2	8.68	0.00	–
PRIFACE	$\mathbf{126.91}$	$\mathbf{123.27}$	$\mathbf{0.00}$

Table 1 – FID scores by each pair of datasets. The distance between a dataset and itself should be zero as they are equal.

Configuration	Mean (↑)	Std Deviation
Original	0.096	0.157
DP2	0.096	0.159
PRIFACE	0.437	0.157
CF	0.419	0.159
CF+DP2	0.094	0.153

Table 2 – Mean and standard deviation of the IFQA score for each of the configurations.

just the *DP2* configuration. We obtained a similar FID of 123.27 between *DP2* and *PRIFACE*. A similar score with the last case, *Original* and *PRIFACE*, suggests that the CodeFormer impacted original and generated faces on a similar intensity.

In Table 2, we applied the IFQA score to quantify quality and resolution. Its values should be closer to one, as the image gets better resolution and seems more realistic and closer to zero otherwise. *PRIFACE* and *CF* got the best IFQA average scores of 0.437 and 0.419, respectively. *DP2* and *CF+DP2* both had low scores (close to 0.0095), similar to the *Original*. Those three modules do not end with our Enhancer module. The low IFQA score is probably due to the WIDER FACE dataset having 50% of its faces with a resolution of less than 50 pixels high. The standard deviation on the images is high (0.15) due to the nature of the WIDER dataset (YANG et al., 2016): the variety of challenging scenarios such as face resolution.

We present distributions for each configuration (Table 2) in Figure 8. The first row represents the proposed pipeline: the *Original* distribution has two-thirds of its images on a quality below 0.1 in IFQA score. Followed by a similar distribution of the artificial faces generated on *(*DP2), both IFQA mean scores 0.096 (Table 2). In the last stage, *(*PRIFACE) IFQA distribution is more evenly dispersed around 0.5 and 0.6. The quality mean improved by 353% (0.0966 to 0.4379 from Table 2) by enhancing the faces with CodeFormer. **The main contribution of this preliminary work is the use of the PRIFACE pipeline to align the protection offered by DeepPrivacy2 with the quality improvement on artificial faces**

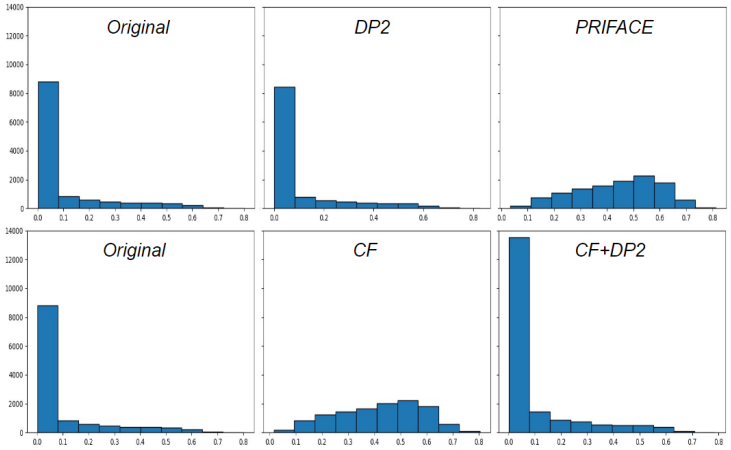

Figure 8 – Impact study on the distribution of IFQA score over the WIDER FACE dataset variations. The first row shows modules in the proposed order: *Original*, *CF*, and *PRIFACE* (CF+DP2). The second line shows applying the enhanced module before the anonymization module.

from CodeFormer.

The second line studies applying the modules in the opposite order: *Original*, *CF*, and *CF+DP2*. In *CF*, the mean IFQA score improved by 335% (from 0.0963 to 0.4193 from Table 2). However, in *CF+DP2*, the distribution produced an analogous to *Original* (0.094 in Table 2). The trade-off was that the enhanced input of *CF* allowed DeepPrivacy2 to anonymize almost 61% more faces (the distribution has more elements than the others). The detector was able to find more faces in each image. So, even if the generated quality were the same, more people were protected. **Higher input quality results in higher face detection rate**, reducing one of the main limitations of the anonymization methods: only faces detected by the models can be anonymized.

Qualitative Evaluation. We sort sample results of *PRIFACE* from worse to best final IFQA score (third column) in Figure 9. This Figure illustrates images from a lower to a higher spectrum score of IFQA. As discussed in quantity results, there is no real gain between the first and second images on the anonymization module. However, a notable IFQA score improvement occurs on the face enhancer module (right column).

In the first line, the Original face (left) has low resolution and a lateral pose. It initially scored 0.0065; after the anonymization (middle image), the score had no significant change, going to 0.0071. On the enhancer module, the resolution improved, but the face did not look natural. Thus, the final score only increased to 0.1432. The second line had a frontal face with 0.0097 of IFQA; the anonymized face(middle) got 0.0077. The anonymization module did not preserve attributes like gender, ethnicity, or age. Those results suggest that our models could be biased toward specific patterns, and further studies could be done to this extent. In the

Figure 9 – Qualitative analysis for a sample of images from the WIDER FACE dataset through PRIFACE. From left to right, we have the original image from the dataset, the results from anonymization using DeepPrivacy2, and our improved result with PRIFACE. Each face image is coupled with a heatmap, in which the higher scores (in yellow) are regions that the metric classifies as "more natural," while the lower scores (in blue) are low-quality fake components. The IFQA score is also placed under each pair of images.

last line, the IFQA scores during each module were 0.0041, 0.005 and 0.6253. This evolution represents the most common growth of the PRIFACE IFQA score distribution (see the two most common columns IFQA are between 0.5 and 0.7 Figure 8).

3.5 FURTHER VISUAL QUALITY QUALITATIVE ANALYSIS

After the development of PRIFACE, as previously explained, we perceived several quality-related patterns in the generated images. Pursuing this direction, we updated the baseline technique for anonymization, using the GANonymization (HELLMANN et al., 2024) mode, proceeding with the training by using the specifications detailed by the authors. We trained the model on the CelebA dataset for 25 epochs (same as the authors). Then, we conducted an in-depth visual analysis of the generated images to better identify the core issues regarding the visual quality. Considering the results, we have grouped them into six groups (Figure 10), namely:

(A) Dark Eye Region;

(B) Misaligned Irises;

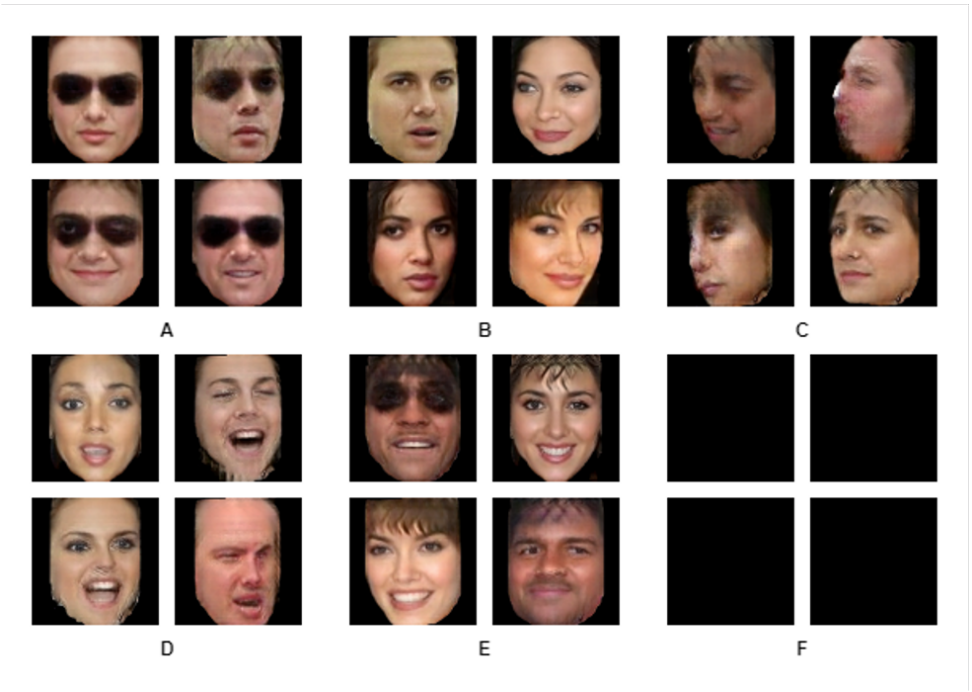

Figure 10 – Initial quality issues in our GANonymization training using the CelebA dataset. Each group represents an identified issue:(A) Dark Eye Region, (B) Misaligned Irises, (C, D) Low-quality Faces, (E) Residual Artifacts in the hair, and (F) Faces not identified by Mediapipe.

(C) Strange looking Faces for the unusual head pose inputs;

(D) Low-quality Faces for uncommon facial expression inputs;

(E) Residual Artifacts in the hair;

(F) Faces not identified by Mediapipe.

Considering those 6 groups of issues, we initiated an investigation for solutions within the scope of this dissertation. We first tackle group (A). We noticed that the issue relies on the training set, where some faces have sunglasses, and thus, the models associated the region with dark colors. We solved the problem by filtering the images in the training set with the glasses label. Issue (E) is likely related to the network architecture, as the skip connections from the U-Net carry some of the high-frequency details from the images, so we decided to focus on other issues. We also skip issue (F) related to face identification, which is one limitation in many face recognition or face anonymization pipelines.

We suspect that problems (C) and (D) are caused by the training set's lack of diversity in head poses and facial expressions. As pointed out by ISOLA et al., the pix2pix does generate a subset of the training set and, thus, does not have that much variety. To overcome this variety limitation, we explored a solution in this dissertation: complementing the training with another

dataset. Finally, we considered problem (B) vital according to our vehicular problem scenario, and it was the issue to which we dedicated following efforts of this master's scope, therefore investigating "Does training face generators with synthetic datasets improve their utility for preserving gaze direction during face anonymization?"

4 METHODOLOGY

Focusing on answering the question, "Does training face generators with synthetic datasets improve their utility for preserving gaze direction during face anonymization?" we expand our investigation strategy towards improvements using a base anonymization technique. This Chapter explains our proposed annotated dataset MetaGaze (Section 4.1). We detail the base technique, GANonymization (Section 4.2), and our improvement approaches Fine-Tuning Variations (Section 4.3) and Mesh Variations (Section 4.4).

4.1 METAGAZE

Study	Participants	Head Poses	On-screen Gaze Targets	Eyelid Openness	FOVs	Images
UTA Heracleia dataset (2012)	20	1	16	1	1	videos
BioID database (2013)	103	1	12	1	1	1,236
Ulm Gaze Dataset (2007)	20	19	2-9	1	1	2,220
Gaze locking (2013)	56	5	21	1	1	5,880
Eyediap (2014)	16	continuous	continuous	1	1	videos
Multi-view Gaze Dataset (2014)	50	8 + synthesised	160	1	1	64,000
MPIIGaze (2017)	15	continuous	continuous	1	1	213,659
MetaGaze (Ours)	30	25*	15	**5**	**3**	71,610

Table 3 – Comparison of Gaze Tracking Datasets. *Instead of head poses, we explored camera angles. Based on the MPIIGaze table (2017).

In addition to modifying the GANonymization technique, we address the need for a specialized annotated dataset to assess gaze preservation in face anonymization techniques quantitatively. To this end, we first thoroughly investigated existing gaze datasets to identify gaps and unmet needs, particularly in explicit variations in the field of view (FOV) and eyelid openness. This analysis revealed that no current dataset adequately addresses these dimensions, underscoring the necessity for creating the MetaGaze dataset. Table 3 compares our proposed MetaGaze dataset and other existing gaze datasets. The unique contribution of MetaGaze lies in its explicit consideration of five eyelid openness variations and three FOV variations, as detailed in Table 4. Additionally, MetaGaze offers greater diversity in head poses and on-screen gaze targets compared to earlier datasets such as "UTA Heracleia," "BioID," "Ulm Gaze," and "Gaze Locking" (MCMURROUGH et al., 2012; VILLANUEVA et al., 2013; WEIDENBACHER et al., 2007; SMITH et al., 2013). By leveraging synthetic data, MetaGaze allows complete control over scene conditions and facilitates rapid expansion by adding new attributes, such as light direction or facial expressions. This adaptability further strengthens its utility as a training set and

benchmark for gaze preservation in anonymized faces. Excluding video datasets such as "UTA Heracleia" and "Eyediap" (MCMURROUGH et al., 2012; MORA; MONAY; ODOBEZ, 2014), as well as the "Multi-view Gaze Dataset" (SUGANO; MATSUSHITA; SATO, 2014), MetaGaze includes a substantial volume of images (69,840), enabling its effective use in fine-tuning training, a topic discussed further in Section 4.5.

Attribute	Variations	Values
Eyelid Openness	5	Wide Open, Open, Semi-closed, Half-closed, Closed
FOV	3	60°, 90°, 120°
Gaze (pitch)	5	-60°, -30°, 0°, 30°, 60°,
Gaze (yaw)	3	-30°, 0°, 30°
Camera (pitch)	5	-20°, -10°, 0°, 10°, 20°,
Camera (yaw)	5	-20°, -10°, 0°, 10°, 20°,
Resolution	1	512x512px
Subjects	30	-
Images	69,840	-

Table 4 – MetaGaze attribute variations, including Eyelid Openness, FOV, Gaze, and Camera angle variations. Other specifications, like resolution, total number of images, and subjects, are also displayed.

We introduce MetaGaze, an annotated synthetic dataset with almost 70,000 images, each with a resolution of 512x512 px. Its face images cover a variety of gaze angles and camera conditions. It uses 30 prefab human models available on the Meta Humans (Epic Games, 2023a), inside the Unreal engine (Epic Games, 2023b). A sum of all the attribute configurations applied on the Humans prefab models to record the images from the MetaGaze dataset can be seen in Table 4. We have five Eyelid Openness variations, from wide open to closed. Regarding the camera, we have used 25 combinations of pitch and yaw, along with 3 FOVs configs. Finally, we have 15 gaze direction combinations, 3 in yaw and 5 in pitch.

Some of the variations presented in Table 4 can be seen in Figure 11. In 'Gaze,' we display all the 15 gaze directions available in the dataset, corresponding to the combinations in Gaze (pitch) and Gaze (yaw) from Table 4. In 'Camera,' we show 10 out of 25 positions, specifically 20° and 20° degrees in yaw combined with -20°, -10°, 0°, 10°, and 20° in pitch. Eyelid Openness shows the five variations present in our dataset. For last, in 2 FOVs 60°, and 120° degrees are in the image. All the other images have FOV of 90° degrees, wide-open eyes, and gaze direction of 0° pitch and 0° yaw for comparison purposes. The variation covers different camera specs available in vehicles or other environments as in A2D2, Audi dataset (GEYER et al., 2020) or in other autonomous driver sources (BAEK et al., 2021).

Figure 12 shows how to adjust the face parameters of the human model of MetaHuman

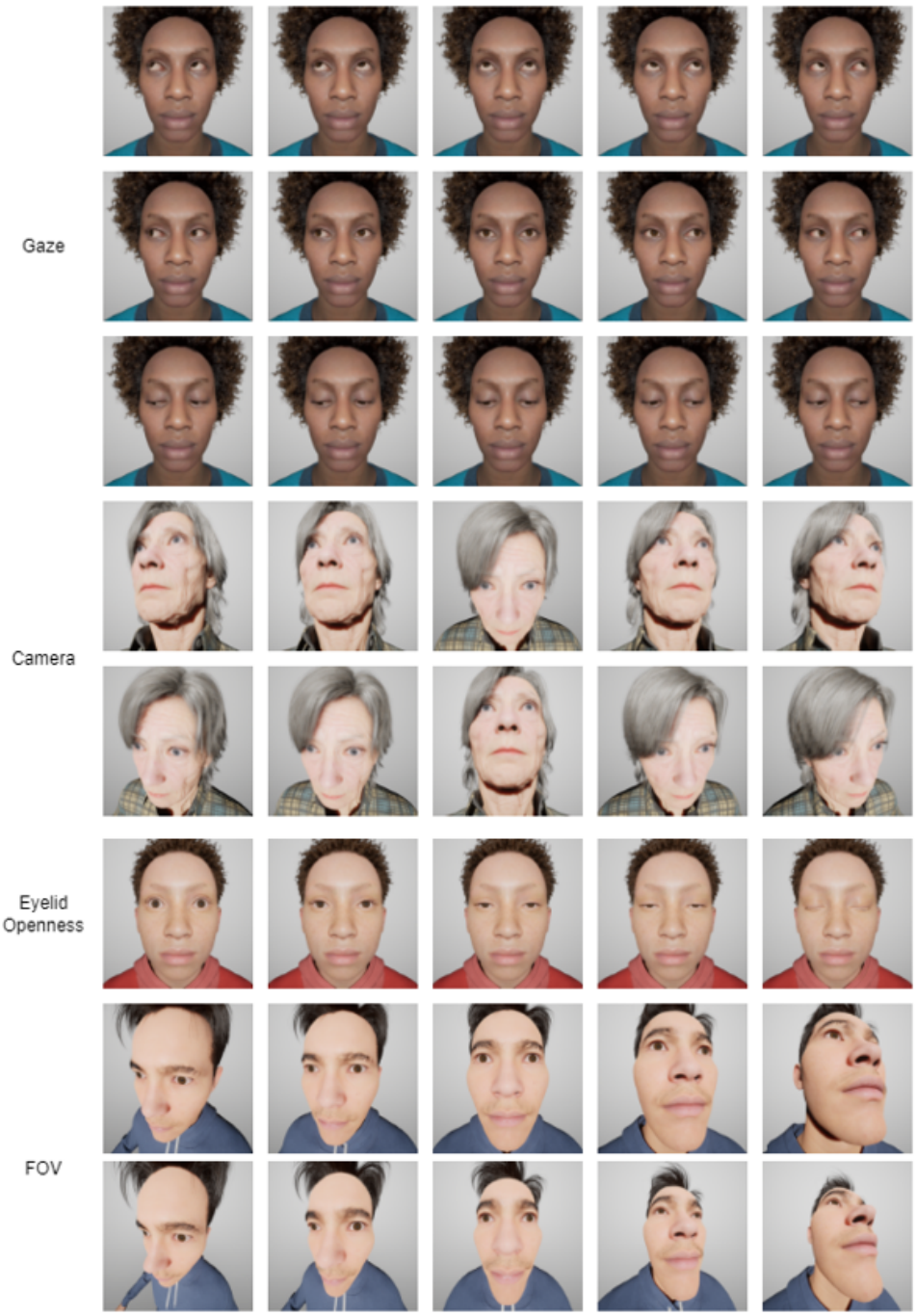

Figure 11 – Samples from MetaGaze show diversification in Human Models, Gaze, Camera, Eyelid Openness and FOV.

in the Unreal Engine. The attributes of gaze direction (pitch, yaw) and eyelid openness are directed, controlled on the control rig, and annotated. The camera parameters, such as distance and angle from the face and FOV, are also annotated, directed, and controlled on Unreal.

The camera was positioned 30 units from the center of the head. For the iris or gaze variations, we considered five configs on MetaHumans "CTRL_C_eye" (-1.0, -0.5, 0, 0.5, 1.0) in the pitch and 3 in the yaw (-1.0, 0, 1.0), totalizing 15 gaze configurations. We also got 5

Figure 12 – Interface of the Unreal engine with the MetaHuman model. In the right, the control rig capable of adjusting the face parameters of the 3D human model

eyelid openness values(-1, 0, 0.3, 0.5, and 1) by modifying "CTRL_C_eye_blink" on Unreal. -1 being wide open and 1 entirely closed. A total of 2328 images per human model. We split the dataset using the proportion 80% 10% 10%, so 24 synthetic human models were used for the training set and 3 synthetic human models for each validation and test set.

4.2 GANONYMIZATION

Reference	Application	Input	Model	Data Privacy
Ours	Gaze Estimation	Face Images	pix2pix	Identity
GANonymization (2024)	Expression Recognition	Face Images	pix2pix	Identity
Deepprivacy (2019) DeepPrivacy2 (2023a)	Face Image Synthesis	Face Images Full Body Images	Stylegan2	Identity
Disguise (2024)	Expression Recognition Gaze Estimation	Face Images	VED	Identity
CIAGAN (2020)	Face Image Synthesis	Face Images Full Body Images Videos	CGAN	Identity
PPRL-VGAN (2018)	Expression Recognition	Face Images	VGAN	Identity
TIP-IM (2021b)	Face Image Synthesis	Face Images	TIP-IM	Identity
AD-GAN (2019)	3D Face Image Synthesis	Face Images	AD-GAN	Identity
CAE (2018)	Face Recognition	Face Images	ACGAN	Gender, Age or Race
PrivacyNet (2020) Privacy-Protective-GAN (2019)	Face Recognition	Face Images	SAN PP-GAN	Soft-Biometric Attributes
Privacy Preserving Action Detection (2018)	Action Detection	Video	DCGAN	Face

Table 5 – Literature sum in private Face-related applications. Adapted from "Generative adversarial networks: A survey toward private and secure applications" (2021)

We compare the current face anonymization techniques to determine the best base technique to improve the gaze preservation trait. Table 5 gathered the comparison and was adapted from a survey of privacy applications (CAI et al., 2021). We have chosen to enhance the

anonymization technique (HELLMANN et al., 2024), as it is a pix2pix architecture, a relatively simple architecture that generates its output based on the input image (see Section 1.3 for more details). Those two traits gave us conditions to modify its properties better to achieve our goal. However, unlike his work, which evaluates facial emotion expression preservation, our goal application is gaze estimation, and we want to preserve the values between the input face image and the anonymized face output.

The work entitled "Disguise without disruption" (CAI et al., 2024) shares the goal of gaze preservation on anonymized faces; the authors propose a Variational Encoder-Decoder (VED) architecture that aims to transform the ID of the desired face. Instead of directly using the face mesh as input (as in GANonymization and Ours), his idea is to change only the identity and not the direction of the head and gaze. We did not directly compare them in the experiments because the authors did not provide the disguise's code in their paper. We also argue that their experiments in gaze validation using CelebA (LIU et al., 2015) and LFW (HUANG et al., 2008) datasets were not ideal, as neither dataset was designed for gaze evaluation as they do not have gaze labels and a proper distribution of gaze direction.

Other techniques do not have specific utility preservation and focus on face image synthesis as in DeepPrivacy2 and CIAGAN (YANG et al., 2021b; HUKKELÅS; LINDSETH, 2023a; MAXIMOV; ELEZI; LEAL-TAIXÉ, 2020). Most techniques use face images as input, but DeepPrivacy2, CIAGAN, and "Privacy Preserving Action Detection" (HUKKELÅS; LINDSETH, 2023a; MAXIMOV; ELEZI; LEAL-TAIXÉ, 2020; REN; LEE; RYOO, 2018). Diverse variants of GANs and a VED (Variational Encoder-Decoder) (CAI et al., 2024) are used as models. All but the last three aim to protect people's identity.

We explored different approaches to improve the GANonymization (HELLMANN et al., 2024) technique regarding its original results and also compared it to other impactful face anonymizer, DeepPrivacy2 (HUKKELÅS; LINDSETH, 2023a).

The GANonymization technique was selected, as previously mentioned, for two main reasons: first, it is a cGAN (conditional Generative Adversarial Network) (MIRZA; OSINDERO, 2014; GOODFELLOW et al., 2014b), which means an external parameter can guide the network to generate the face (see Figure 13). In this case, an image based on the mesh created by Media Pipe Face Mesh (KARTYNNIK et al., 2019) retains the geometrical traits of the face without carrying many identifiable traces. Second, the GANonymization is based on pix2pix (ISOLA et al., 2017) architecture. This relatively small network enables us to train many model configurations in a reasonable time (around one or two days in an RTX3090).

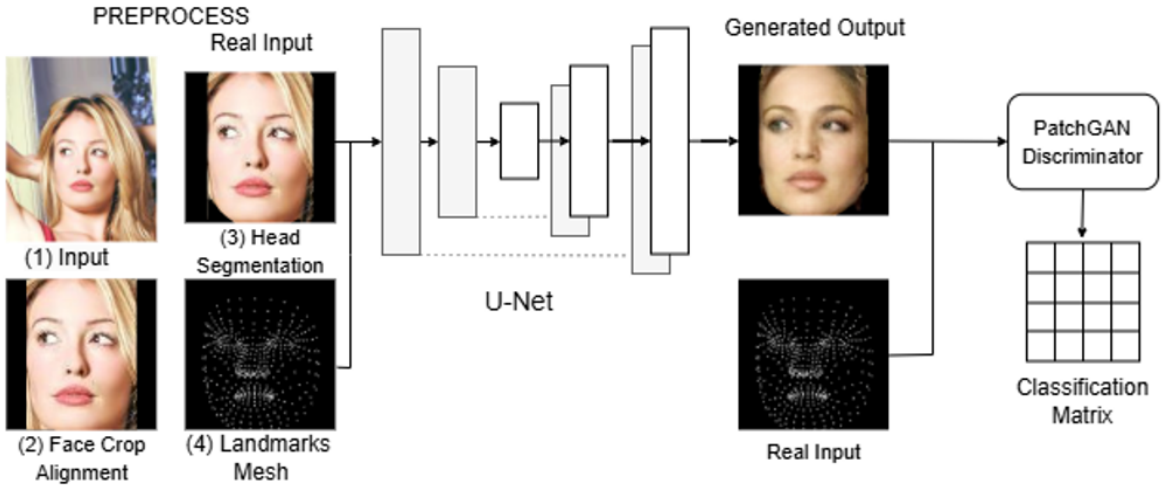

Figure 13 – GANonymization architecture, inspired by pix2pix. From left to right, the first step is to preprocess the data (1) input for training; it is done in three stages: (2) face crop and padding, (3) head segmentation, and (4) mesh estimation (an image of the landmarks). The second step is to input the head segmented along with the mesh to a generator based on U-Net, which will output a face conditioned by the input. Based on PatchGAN, the discriminator will receive the tuple input and output and will return a matrix of values from 0 to 1, determining whether each patch region looks like the real data.

Two main factors of pix2pix are responsible for its time efficiency: first, the U-Net generator is an encoder-decoder with skip connections, and these connections enable some high-frequency features from the input image to be used in the reconstruction, accelerating the loss function convergence. Second, the PatchGAN discriminator outputs a matrix where each image region (patch) receives a score from 0 to 1. 0 is fake (generated on the generator), and 1 is real data. It helps determine which regions need more adjustments in the backpropagation (see Section 1.3 for more details). Our efforts to improve the gaze preservation utility were based on two directions: fine-tuning and mesh input modification. In all variations, we have maintained the hyperparameters of the authors: learning rate of 0.0002 and decay of first-order momentum of gradient (adam) b1 of 0.5 and b2 of 0.999.

4.3 FINE-TUNING VARIATIONS

A representative training set is crucial for any model training, especially on a pix2pix (ISOLA et al., 2017) network that usually suffers from overfitting and needs dropout or other strategies to avoid it. Using fine-tuning instead of just merging the two datasets is specially indicated where the domain of the original training task and the final are similar, which is the case of generating human faces (LI et al., 2020). Besides, CelebA and MetaGaze have different image

resolutions, which would not be ideal for training together. We have trained four models from the GANonymization (pix2pix-based) versions. First, we call GANonymization as we reproduce the parameters and use the CelebA dataset like the GANonymization paper. The second version is named MetaGaze_model, and we evaluate how well our dataset performs as a training set once MetaGaze covers a larger range of gaze (see 16). We also have applied fine-tuning on those two models for more than 25 epochs to create the MetaGaze+CelebA and CelebA+MetaGaze configurations. We opt not to freeze any layer for the fine-tuning, as it would be out of scope to determine the best configuration. We have lowered the learning rate by 4 times, from 0.0002 to 0.00005. Usually, a reduction factor of 10 is used, but since the tasks are pretty much the same and the learning rate change is related to the similarity (LI et al., 2020), we have opted to use a factor of 4 instead (LI et al., 2020).

4.4 MESH VARIATIONS

As Figure 13 describes, the conditional GAN receives a non-random input (image containing the landmarks mesh) to guide the generator. We have modified the preprocessing step by modifying the input mesh image. In Figure 14, the first on the left corresponds to the original Media Pipe Face Mesh image used by GANonymization. The one in the middle was an effort to highlight the iris region to the network during training; this variation was called Iris. We also modified the mesh to be formed by a Tesselation instead of just the dots; the intention was to accentuate the geometrical aspects of the face to be "learned" by the network.

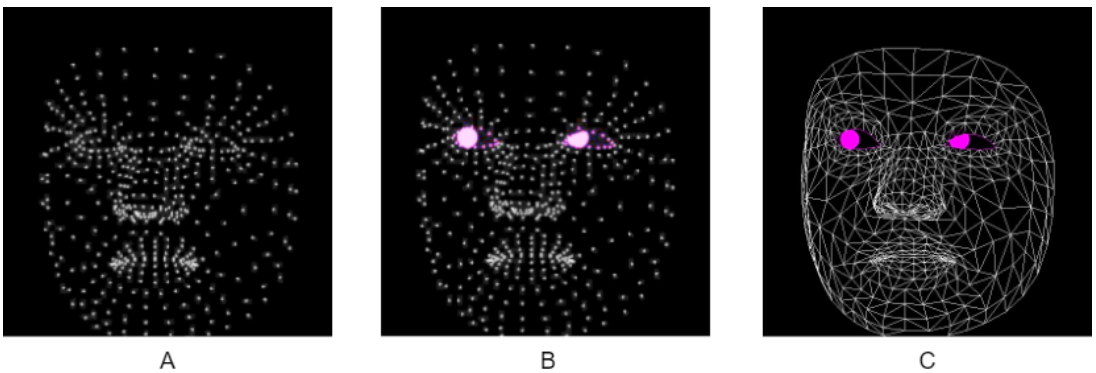

Figure 14 – Images A, B, and C are three Media Pipe Face Mesh variations. Being A, the original, B, our Iris variation, and C, Iris+Tesselation.

4.5 EXPERIMENTS

In this Section, we detail the choices of metrics, datasets, methods, and implementation details of our experiments. We develop experiments to compare our variations to the baselines on MetaGaze and DMD datasets. First, we compare the gaze direction estimation by L2CS of the generated faces using the annotated ground truth of our MetaGaze to reference the gaze direction angles. Next, we did the same but compared the L2CS estimator before anonymization and after, both on MetaGaze and DMD datasets. This was because DMD has no annotated values, and we applied the L2CS estimator on both initial and anonymized faces to avoid L2CS bias. As our dataset is annotated, we have filtered the results considering some image attributes. Finally, we compare the results qualitatively on each variation, considering arbitrary inputs first and after, compiling the worst result on each technique.

4.5.1 Metrics

To evaluate the gaze preservation of the selected anonymization techniques, we used Mean Absolute Error (MAE) to calculate the difference between the original face and the anonymized face gaze estimation in degrees. Similarly to the Disguise technique (CAI et al., 2024), we opt to perform the gaze estimation on L2CS-Net (ABDELRAHMAN et al., 2023) and the overall head orientation on Retinaface (DENG et al., 2020). The L2CS technique receives an image with faces (in all of our scenarios, only one face per image) and outputs its gaze estimation direction by returning a value in degrees for PITCH and YAW. The Retinaface finds 5 3D key points in the face, one for each eye, one for the nose, and two for the mouth.

We first compare the L2CS estimation and Retinaface on our dataset, MetaGaze, to validate its performance on synthetic faces. We have compared the L2CS gaze estimation with the ground truth from MetaGaze, which we calculated using the Unreal engine parameters. We do not calculate a ground truth. Instead, we used the face's key points positions before anonymization for comparison. From these results, we use it as a baseline to compare the subsequent experiments: anonymize our test set with each of the seven techniques used in this work and then estimate the gaze direction MAE by calculating its difference from the ground truth.

Figure 15 – Sample images from the DMD dataset. In this work, we selected only frontal camera images (first row) to estimate the gaze direction from the driver better.

4.5.2 Datasets

The datasets used in our experiments for gaze preservation on anonymized faces were: Our proposed dataset, MetaGaze, and the Driving Monitoring Dataset (DMD) (ORTEGA et al., 2020). The criteria we employed were to get a nonsynthetic vehicular-focused dataset. DMD has frontal cameras near the car panel to capture the driver from a frontal angle (see Figure 15). We extract 10 frames per second from 15 different person videos of 3 minutes each. The total number of images collected was almost 30,000.

As intended by our proposed dataset, Figure 16 shows that the MetaGaze could (both in the ground truth and L2CS gaze estimation) cover a more evenly represented gaze direction than the CelebA dataset. Thus, we do not use the CelebA dataset in our evaluation, even though it was used in the Disguise (CAI et al., 2024) experiments. The main reason is that the CelebA was not developed for this intent of gaze estimation and does not cover a reasonable amount of gaze pitch and yaw angle. We also show that even though the CelebA (LIU et al., 2015) dataset is used to train several GAN-related methods for anonymization; it lacks representation on some gaze directions classes. Besides, adding a variety of FOVs and Eye openness is beneficial as both a benchmark and a training dataset for face synthesis, making the model more robust for these conditions.

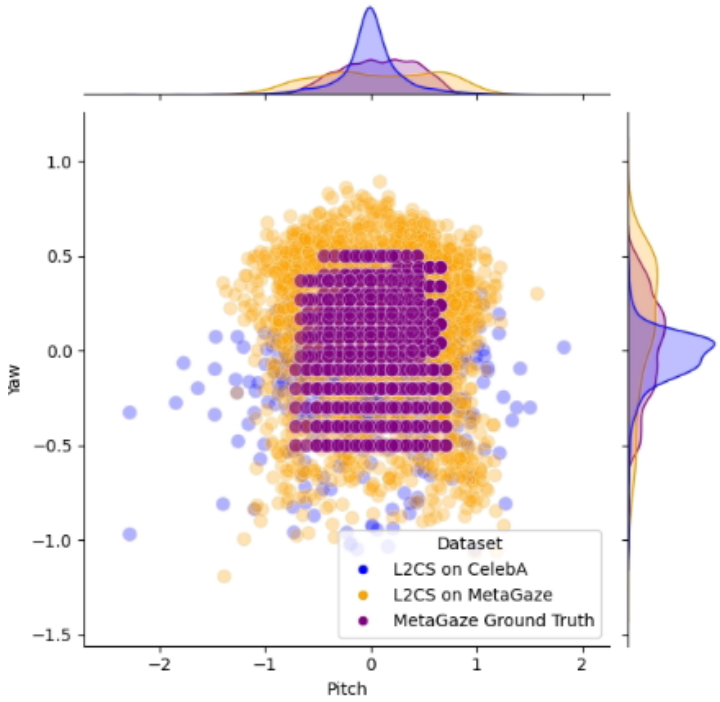

Figure 16 – Distribution of gaze estimation in degrees of two face datasets. In the blue is the CelebA dataset, gaze estimation using L2CS; in orange is our proposed dataset, MetaGaze, gaze estimation using L2CS; in purple MetaGaze ground truth.

4.5.3 Selected Techniques.

We compare a total of 7 models. The first two were used as baselines: the DeepPrivacy2 pretrained model, given by the authors, and the GANonymization model, trained by us, but with the same parameters and seed (to reproduce the random number generator) as the one used by the authors. The other 5 techniques (MetaGaze_model, CelebA+MetaGaze, MetaGaze+CelebA, Iris, Iris+Tesselation) were variations of the GANonymization we trained. We trained all models for 25 epochs.

Figure 17 synthesizes our variations used in the experiments. We have used the models trained by the authors of DeepPrivacy2 and GANonymization as a baseline. From GANonymization, we have derived our mesh variations: Iris and Iris+Tesselation. From the MetaGaze dataset, we trained a model, MetaGaze_model, along with two fine-tunings: CelebA+MetaGaze and MetaGaze+CelebA.

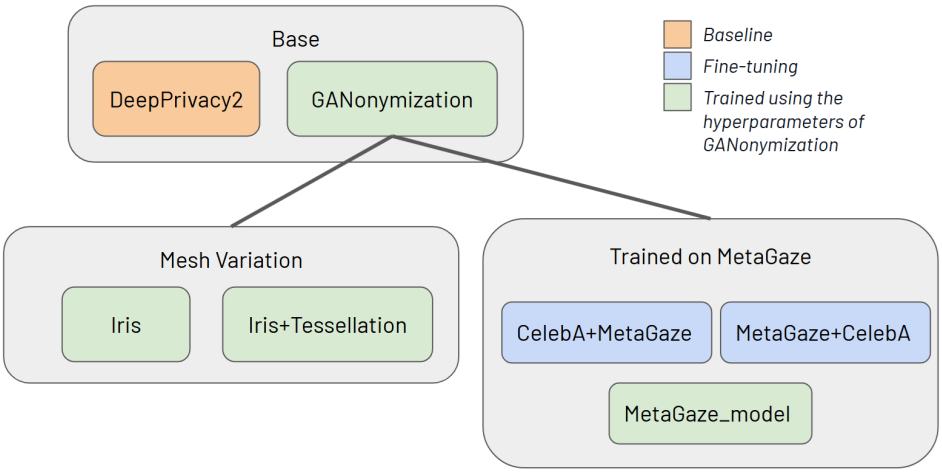

Figure 17 – Fluxogram of the 7 models used in experiments. We divided them into 3 groups: Baseline for the original techniques used as a comparison. We created mesh variations by modifying the input mesh image and training it on MetaGaze, either by fine-tuning or directly training.

4.5.4 Implementation Details.

We conducted our experiments on a desktop computer running Ubuntu 22.04 LTS, which has a 13th-generation Intel Core i9-13900 with 32 GB of RAM and a GeForce RTX 3090 with 24GB of VRAM. The models have taken approximately 48 hours to train for 25 epochs. The base models used during the experiments were from the following repositories: DeepPrivacy2[1], L2CS-Net[2], and GANonymization[3].

[1] https://github.com/hukkelas/deep_privacy2
[2] https://github.com/Ahmednull/L2CS-Net
[3] https://github.com/hcmlab/GANonymization

5 RESULTS AND DISCUSSION

In this Chapter, we have done a Quantitative Analysis (Section 5.1) of our proposed dataset, MetaGaze, and DMD dataset, comparing 7 different methods regarding gaze preservation using L2CS-NET as a gaze estimator. Table 6 displays the results relative to the annotated ground truth annotated by us, while Table 7 does the preservation estimation by subtracting the gaze direction after anonymization from their original value (pre-anonymization). In addition, along with Further Quantitative Analysis 5.2 to dig deeper into the results by comparing the variations by each dataset attribute. Finally, we perform a Qualitative Analysis (Section 5.3) to compare the faces generated visually. Finally, the main points are summarized in Section 5.4.

5.1 QUANTITATIVE ANALYSIS

Methods	**Gaze (MAE on L2CS°↓)**			
	Absolute		**Relative**	
	Pitch	**Yaw**	**Pitch**	**Yaw**
Non-anonymized	26.5	12.8	0.0	0.0
DeepPrivacy2	41.7	27.5	15.2	14.7
GANonymization	37.3	19.2	10.8	6.4
MetaGaze_model	40.6	20.3	14.1	7.5
CelebA+MetaGaze	**34.4**	**18.7**	**7.9**	**5.9**
MetaGaze+CelebA	42.3	19.9	15.8	7.1
Iris	36.0	21.1	9.5	8.3
Iris+Tesselation	40.2	20.3	13.7	7.5

Table 6 – Results of methods on MetaGaze considering the ground truth. The first two columns show the absolute value and the last two are the relative values (the method minus the non-anonymized value) regarding ground truth.

In Table 6, we compare the methods on MetaGaze by using the ground truth of the developed dataset. We added a non-anonymized line to evaluate the L2CS estimator on the original images and avoid the noise caused by the estimator's error in our evaluation. The absolute error is the direct comparison with the ground truth, and the last two are the relative subtraction of the non-anonymized value in the effort to annulate L2CS error.

With these relative values, the fine-tuning CelebA+MetaGaze improved the error from 10.8 to 7.9 in pitch and 6.4 to 5.9 in yaw compared to the GANonymization relative results. Not only that, but it was the only method to improve GANonymization yaw values. In contrast,

the Deepprivacy2 yaw's relative result of 14.7 % is far above the other ones; this may be due to its difficulty in estimating the gaze yaw direction by only seeing the head pose, as discussed earlier.

	Methods	Gaze (MAE on L2CS°↓)		Retinaface (L2 px distance↓)			
		Pitch	Yaw	All	Eyes	Nose	Mouth
MetaGaze	DeepPrivacy2	47.94	30.63	7.34	4.37	4.87	2.93
	GANonymization	35.25	18.96	6.94	3.93	4.46	3.21
	MetaGaze_model	32.76	17.31	9.65	5.47	5.97	4.82
	CelebA+MetaGaze	34.44	19.64	**6.01**	**3.43**	**3.87**	**2.67**
	MetaGaze+CelebA	**31.08**	**15.76**	9.94	5.51	6.25	5.00
	Iris	32.05	19.62	7.02	3.93	4.50	3.32
	Iris+Tesselation	35.46	18.61	7.26	4.09	4.61	3.45
DMD	DeepPrivacy2	19.24	15.93	17.65	10.35	12.41	5.66
	GANonymization	11.68	13.21	16.25	7.10	11.85	8.43
	MetaGaze_model	13.88	13.80	16.64	7.21	11.94	8.92
	CelebA+MetaGaze	12.33	13.16	16.29	7.02	11.90	8.51
	MetaGaze+CelebA	15.37	13.54	16.89	7.27	12.11	9.10
	Iris	**11.20**	**12.63**	**16.14**	**6.99**	**11.75**	**8.46**
	Iris+Tesselation	11.58	13.26	16.36	7.10	11.88	8.60

Table 7 – Mean absolute error (MAE) in degrees values for gaze preservation between MetaGaze and DMD datasets and their anonymized versions. The "Retinaface" technique locates 5 key points on the face; the column displays four categories: All, Eyes, Nose, and Mouth.

MetaGaze. The first comparison results evaluate the face orientation in two aspects: how the gaze direction estimation is preserved (metrified by L2CS) and how the overall face orientation is maintained (calculated by RetinaFace). We anonymized the test set on each method. We calculate the MAE of the difference in L2CS estimation of the non-anonymized face minus the face anonymized by each method (in degrees). We also calculate the Retinaface keypoints difference in pixels and displayed in Table 7.

Considering the MetaGaze experiment (first vertical half of table), the best result configuration on gaze estimation is MetaGaze+CelebA, with an MAE of 31.08 for pitch and 15.76 for yaw. Curiously, the best configuration regarding head pose (Retinaface) is CelebA+MetaGaze with a pixel error of 6.01, 3.43, 3.87, and 2.67 in all eyes, nose, and mouth, respectively, improving the baseline, GANonymization, at 13.4%. CelebA+MetaGaze was the same as in Table 6 as well, which shows that the eye position and gaze direction are directly related. Both best methods resulted from fine-tuning, demonstrating that this strategy is efficient and that the order of training (which dataset to use as base training and which to use in fine-tuning is

relevant). In Section 5.3, we demonstrate that the dataset used before fine-tuning is visually dominant (for example, the MetaGaze+CelebA faces are more similar to the MetaGaze dataset than the CelebA dataset); this is due to our strategy of fine-tuning learning rate parameter to be a quarter of the initial training. Following this hypothesis, the MetaGaze+CelebA exceeds in gaze estimation by using the base of MetaGaze, which covers more gaze directions, along with some adjustments from the CelebA fine-tuning. While CelebA+MetaGaze exceeds head pose preservation (Retinaface), this is observed in the second and third-best head pose estimations: Iris and Iris+Tesselations trained only on the CelebA dataset.

Following the best result in gaze estimation, we have the Iris method, with 32.05 pitch and 19.62 yaw, and MetaGaze_model, with 32.76 pitch and 17.31 yaw. The results suggest that both directions we aborded (fine-tuning and mesh variations) are valid. Our MetaGaze+CelebA variation, compared to the GANonymization baseline, improved the MAE in degrees in the pitch from 35.25 to 31.08 and the yaw from 18.96 to 15.76. DeepPrivacy2 got the worst result: 47.94 pitch 30.63 yaw. The DeepPrivacy2 performance is probably caused by its strategy of filling the face gap on the image; even though it is more robust, as it can be used even when no key points are found, the face generation lacks orientation information. All 5 of our variations reduce the MAE on either pitch or yaw. The pitch error is almost always near double the yaw error compared to GANonymization. The pitch amplitude is also the double of yaw (see Table 4), so its MAE is naturally greater.

Regarding Retinaface, after the best case, CelebA+MetaGaze, we have GANonymization and a slightly worse result in Iris and Iris+Tesselation. The results show that the Iris variations need refinement to improve the Retinaface evaluation. DeepPrivacy2, different from the gaze estimation, maintained a relatively low pixel error of 7.34, 4.37, 4.87, and 2.93 in all eyes, nose, and mouth, respectively. The worst results are from the two variations trained on MetaGaze: MetaGaze_model and MetaGaze+CelebA, with an error above 9.5 pixels. One hypothesis is that the nature of the pix2pix architecture (used as the base for GANonymization and all variations but DeepPrivacy2) of overfitting some head poses could overwrite some head pose variations. The overfitting makes the error bigger as it could been as some poses are 'deleted' during the training to improve the accuracy.

DMD. To complement our analysis, we have also displayed the non-synthetic and vehicular gaze DMD dataset results in Table 7. In this scenario, the best configuration is Iris, with 11.20 and 12.63 in pitch and yaw. Followed by Iris+Tesselation and GAnonymization, the 3 methods we trained exclusively on the CelebA dataset. Unlike in the MetaGaze dataset experiment, the

gaze direction values have been kept up with the head pose estimation. The overall errors on DMD were nearly three times smaller than in MetaGaze dataset. As in MetaGaze experiments, the worst case was also DeepPrivacy2, 19.24 pitch, and 15.93 yaw. But, unlike in MetaGaze experiment, in this case, only the mesh Iris variation surpasses the GANonymization baseline, which was by a slight amount. DMD dataset has videos instead of images (we extracted 10 frames per second) and covers the car's interior and the drivers instead of the close-on face, like on MetaGaze dataset. The DMD images have the faces not so close to the cameras (see Figure 15 for some samples), and, as we got 10 frames per second, we had a lot of similar images. Those factors could explain the difference in error between DMD and MetaGaze experiments.

Considering Retinaface, we behaved similarly to that in gaze evaluation. Iris was the only one to improve the results of GANonymization by a slight margin of 16.25 to 16.14. Additional data on DMD or even more real driver datasets would be beneficial for further analysis of this behavior.

5.2 FURTHER QUANTITATIVE ANALYSIS

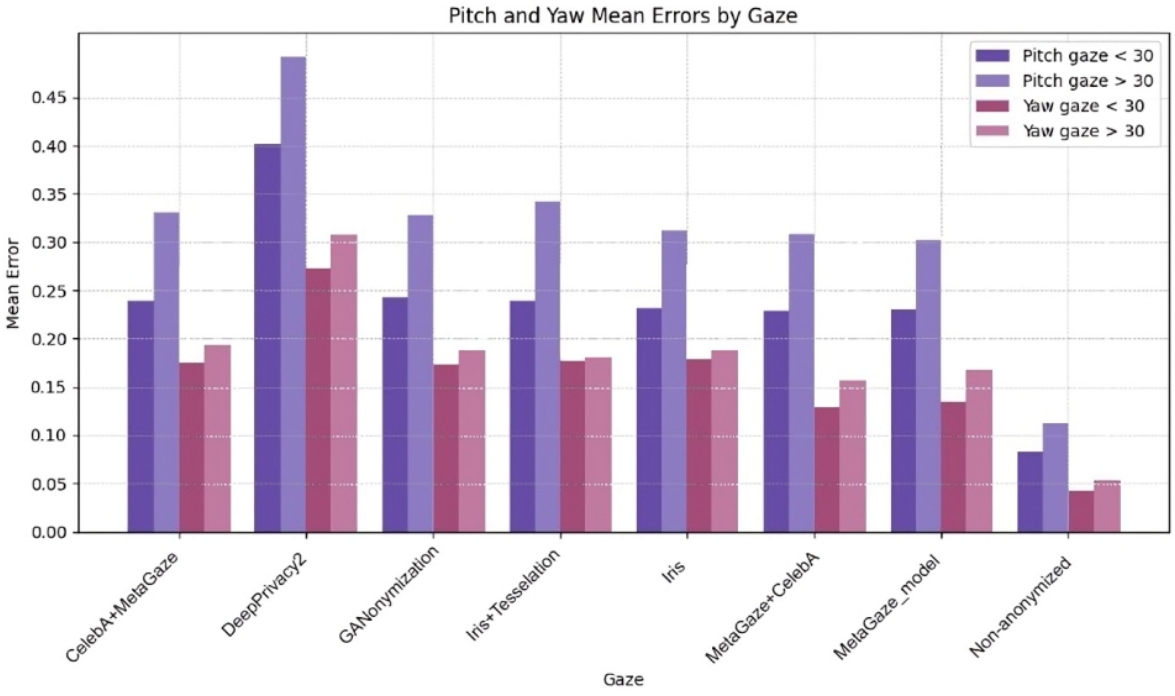

Figure 18 – Comparison of MAE in degrees between the ground truth gaze direction and the evaluation on L2CS gaze estimator for faces generated by each model on MetaGaze test set input. In purple, pitch errors; in pink, yaw errors. The lighter the color, the bigger the gaze angle.

Gaze. In Figure 18, we display the MAE between the prediction and the ground truth of

each pitch method (purple) and yaw (pink). The lighter color represents gaze angles above the absolute value of 30° degrees in either pitch or yaw (see Figure 11 for samples in gaze variation). A pattern occurs in all the cases: more extreme angles lead to higher errors. The difference in yaw is less discerning, as the range is smaller, but in pitch, this contrast goes from less than 10% in Non-anonymized to more than 40% in Iris+Tesselation. Regarding the methods, apart from the baseline Non-anonymized and the DeepPrivacy2 variation, the methods got similar results. DeepPrivacy2 got the worst results considering either of the categories from the gaze angles analyzed.

Two hypotheses to explain this expected behavior considering the two angle groups are: (1) more extreme gaze directions are less represented in training sets, and thus, the models have some difficulty generating them properly. (2) There is an inherent difficulty in generating and even recognizing faces that are not frontal for both models and people.

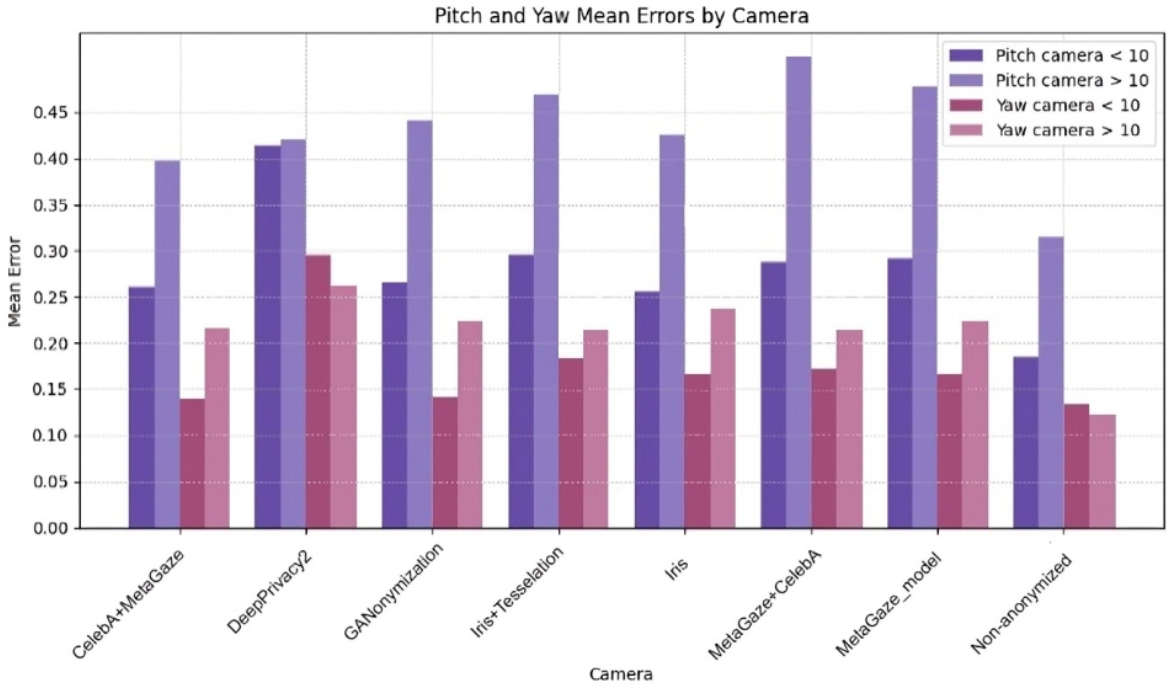

Figure 19 – Comparison of MAE in degrees between the ground truth gaze direction and the evaluation on L2CS gaze estimator for faces generated by each model on MetaGaze test set input. In purple, pitch errors; in pink, yaw errors. The lighter the color, the bigger the camera angle.

Camera. Figure 19 represents the MAE errors separated by camera angles (see Figure 11 for samples in camera angle variations). If any angle in the camera is above 10° degrees, it is in the lighter colors. Like Figure 18, we have noticed a similar pattern: more extreme angles lead to higher errors. In some methods, the angle camera causes the MAE almost to double; for example, in GANonymization, we see that small camera angles have an MAE of 25, while above

10° degree, the error goes to 44. DeepPrivacy2 and the baseline non-anonymized do not follow this pattern in yaw; they have better yaw results on higher camera angles. In DeepPrivacy2, the errors in smaller camera angles are higher than in any other method, so much so that there is no significant difference between small and high camera angles in DeepPrivacy2 anonymized results.

As in gaze, the hypothesis from this behavior is that higher camera angles are underrepresented in the training set, or even those faces on higher angles are less recognizable and thus more challenging to generate.

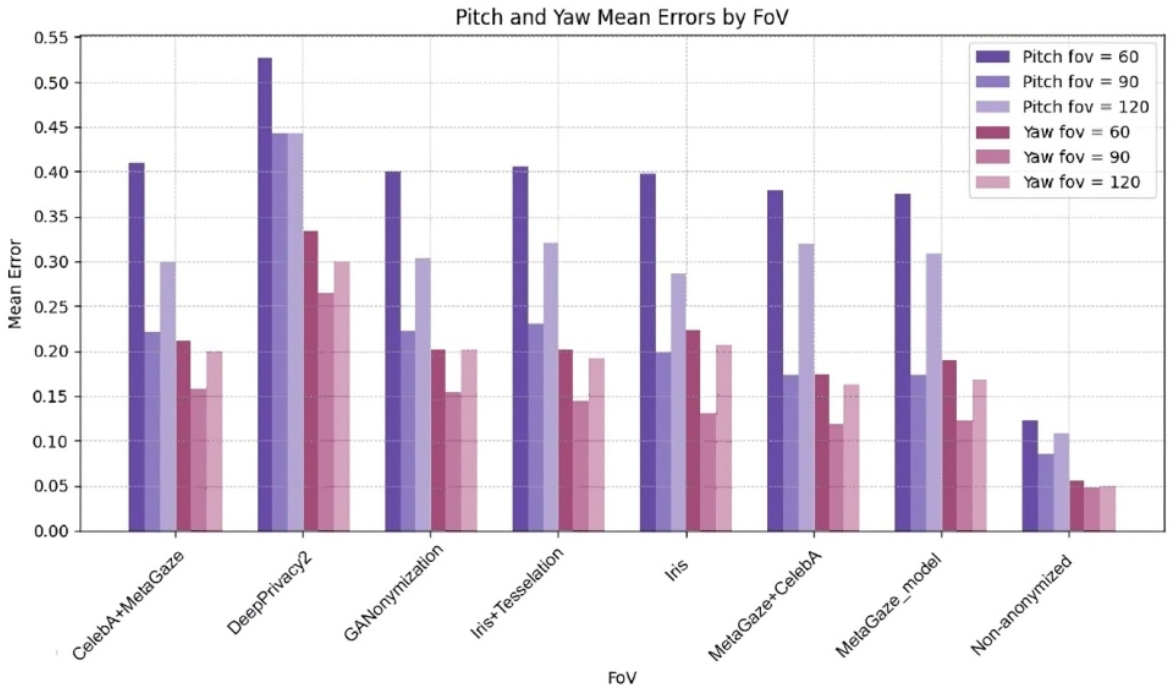

Figure 20 – Comparison of MAE in degrees between the ground truth gaze direction and the evaluation on L2CS gaze estimator for faces generated by each model on MetaGaze test set input. In purple, pitch errors; in pink, yaw errors. The lighter the color, the bigger the FOV angle.

FOV. Figure 20 shows the method's error by camera FOV angle (see Figure 11 for samples in FOV variations). The higher the angle, the lighter the color. Considering the baseline case, Non-anonymized, there is an increase in error for angles 60° and 120° of FOV, even though the distance in yaw for 90° and 120° is less expressive, the pattern persists. We point out that this pattern remains in all the other cases. All methods had more difficulty preserving the gaze direction when the camera FOVs 60° and 120° degrees. However, deepPrivacy2 does not differ in performance regarding 90° and 120° for pitch. Interestingly, the error is more significant in 60° degrees., even though the FOV of 120° degrees is more unusual for us than 60° degrees.

The possible explanation is that the camera FOV 90° is more usual and does not distort the face as much as the other two FOVs. This distortion explains geometrically why the meshes

seem weirder and less natural. In some cases, the distortion even harms the media pipe face mesh and the face generated upon that mesh (see Figure 11 in FOV label, cases of 60°, first row, and 120°, second row).

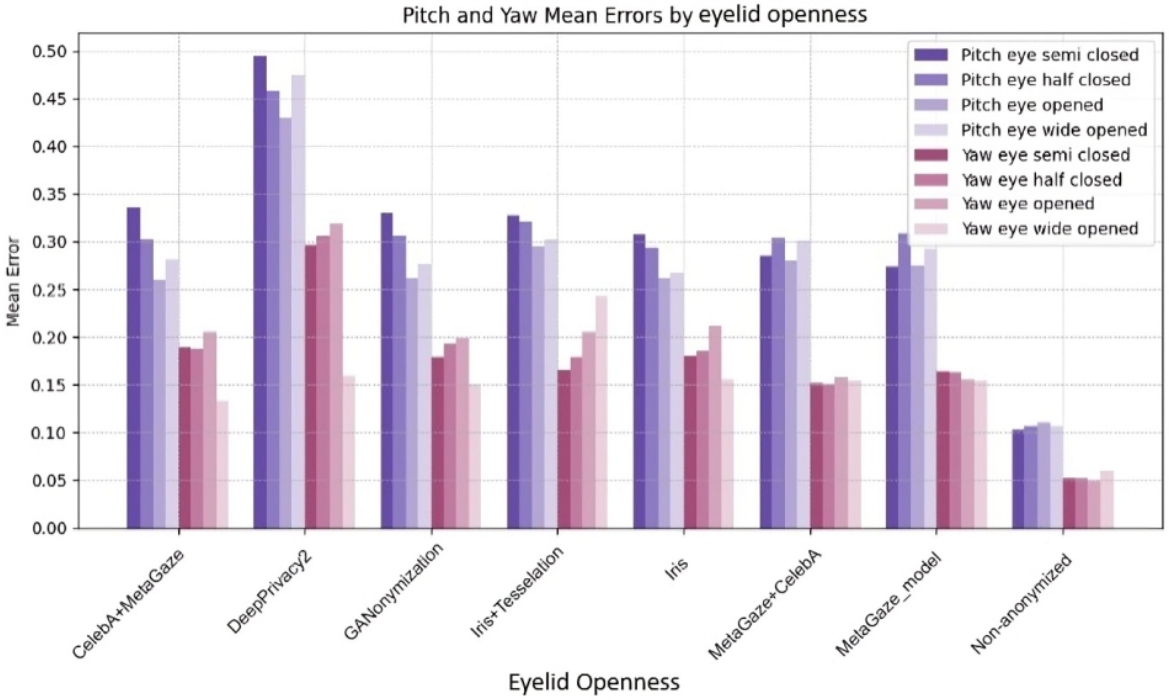

Figure 21 – Comparison of MAE in degrees between the ground truth gaze direction and the evaluation on L2CS gaze estimator for faces generated by each model on MetaGaze test set input. In purple, pitch errors; in pink, yaw errors. The lighter the color, the wider the eyelid openness.

Eyelid Openness. Figure 21 displays four variations of eyelid openness: semi-closed, half-closed, opened, and wide-opened (see Figure 11 for samples in eyelid openness variations). The pattern here is not as straightforward as in Figures 18, 19 and 20. Nonetheless, we notice that, in pitch, the order from greater error to smaller goes as follows: semi-opened, half-closed, wide-open, and open. This behavior is not true in the last three cases, MetaGaze+Celeba, MetaGaze_model, and Non-anonymized. Regarding yaw, we have the opposite direction. The error seems to increase from wide-opened to semi-closed, to half-closed, and to open. The DeepPrivacy2 was no exception in either pitch or yaw in this scenario.

Some hypotheses are: in our data, semi-closed and half-closed were not combined with all the gaze and camera angles, as some occluded the iris by the eyelid, and the estimation would not be viable; this is also why closed eyes are not evaluated in this graphic. Removing the worst-case scenarios may explain why the most natural behavior, higher errors on smaller eyelid openness, was unclear in the graphics. Another interesting note is that wide-opening usually performed worse than opening-in-pitch (maybe because of its naturality) but performed better

in yaw, as the eyelid occlusions were minor. Also, the methods that trained the majority on MetaGaze (MetaGaze+Celeba and MetaGaze_model) reduced the error in semi-closed eyes compared to the other methods, probably because of the presence of a more significant amount of images with semi-closed eyes on training.

5.3 QUALITATIVE ANALYSIS

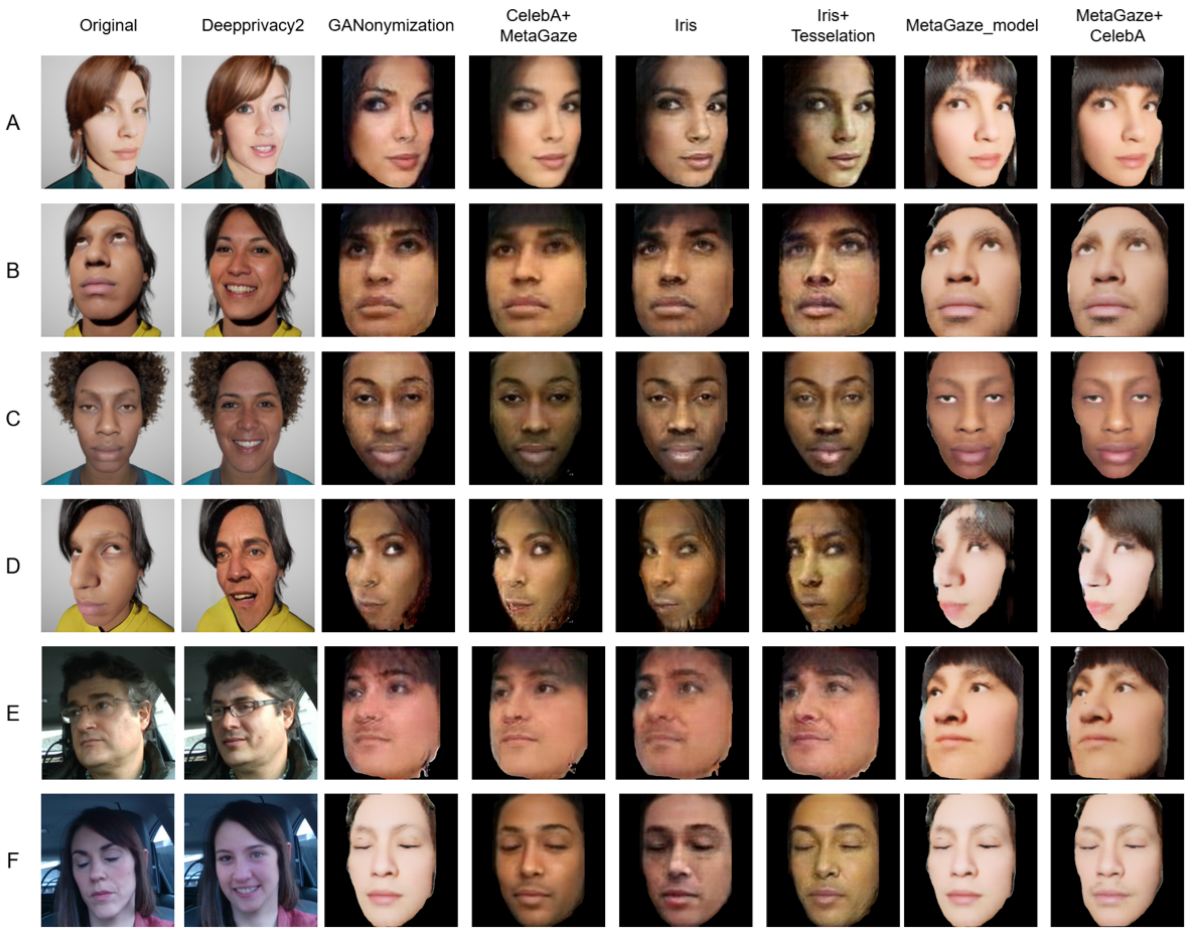

Figure 22 – A matrix containing the input faces (on the left) from MetaGaze (A-D) and DMD (E-F), the output from the two baselines (DeepPrivacy2 and GANonymization) evaluated on experiments, and the output from each of the five GANonymization variations we trained.

Considering Figure 22, we display faces generated from MetaGaze inputs (rows A to D) and DMD (rows E and F) from each of the seven configs evaluated in this dissertation. Considering the MetaGaze inputs, visually, the DeepPrivacy2 outputs were the most unique once all the others shared the same architecture model from GANonymization. Notably, the resolution in DeepPrivacy2 is higher, but the face seems biased to be frontal and smiling despite the input not sharing these features (all but E case). DeepPrivacy2 tackles the face generation problem

as an inpainting problem, so they do not need a face mesh to generate the output. Instead, the original face is removed, and the network should be able to fill this gap with a new face. However, some potential traits (gaze or facial expression) from the input face could be lost in the process.

The following four configs (GANonymization, CelebA+MetaGaze, Iris, and Iris+Tesselation), all but the fine-tuning (CelebA+MetaGaze), were trained exclusively on CelebA. The differences between the four are slight, and the most notable feature is a discrete color tone variation. The last two configurations (MetaGaze_model and MetaGaze+CelebA) were primarily trained on MetaGaze dataset and and output similar faces. However, the hair in the A and D rows had a hair "failure" on MetaGaze_model, which was fixed on MetaGaze+CelebA output. One relevant aspect of training in MetaGaze dataset is that the outputs look like the identity of one of the 24 synthetic human models used in training and, thus, are seen as a little more artificial and repetitive.

Now, considering rows E and F from DMD, we see that DeepPrivacy2 could preserve the man's glasses but not the woman's closed eyes. All GANonymization did not preserve the glasses but preserved the eyelid closed and the neutral mouth expression.

Besides gathering arbitrary images for the qualitative analysis, we displayed the highest gaze error for each method in Figure 23. The matrix contains the input faces (on the left) and the method's label from MetaGaze's worst scenarios. The main diagonal represents that class's worst case (highlighted) and the results from the other techniques for comparison reasons.

Corroborating with our Further Quantitative Analysis in the Section 5.2, the worst cases are usually generated by an input face with extreme camera angles or FOVs. The first row, DeepPrivacy2 (FOV of 120° degrees and extreme gaze direction), did not generate a preserved gaze; the eye is semi-closed and pointing down instead of up; even though the image resolution is high, it does not fit the head pose properly. Considering the other methods from the DeepPrivacy2 input, the gaze direction correctly points right and up. However, the methods not primary training on MetaGaze dataset, which are GANonymization, CelebA+MetaGaze, Iris, and Iris+Tesselation, have generated faces with much noise, while MetaGaze_model and MetaGaze+CelebA are more clean.

The second line has an input with extreme gaze direction, camera angle, and FOV of 60° degrees; this combination occluded one of the eyes and almost occluded the iris. DeepPrivacy2 generated a frontal and smiling face that did not fit the head pose, while the other methods, including GANonymization, were not able to generate a reasonable definition. All the observa-

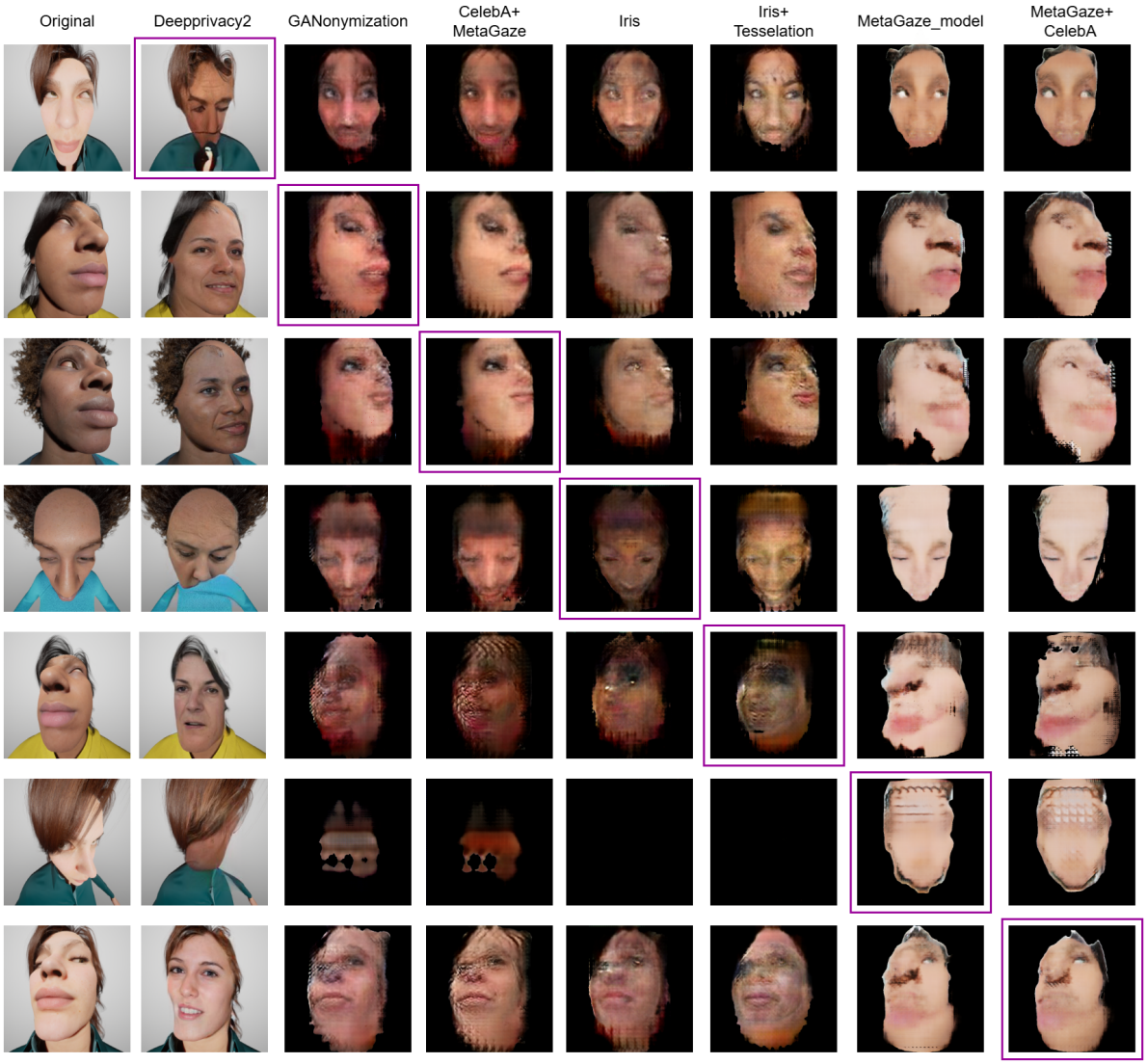

Figure 23 – A matrix containing the input faces (on the left) from MetaGaze worst scenarios. Each line represents the input that causes the highest MAE gaze error for highlighted technique (first line, Deepprivacy2, second line GANonymization and so on), along with the results from the other methods for comparison.

tions of GANonymization can be done to the third row, Celeba+MetaGaze, with the difference that DeepPrivacy2 fits the head pose accordingly, but not the gaze direction.

The most challenging case for the Iris, the Iris+Tesselation, and the MetaGaze_model inputs have similar characteristics: the camera angle is extreme, and the FOV is 120° degrees. This case is challenging as the FOV and camera angle distort the face, and the eyelid, the nose, or the hair occludes the iris. However, all methods generated the eyes semi-closed correctly in the Iris input. On Iris+Tesselation, all GANonymization methods generated very incomplete faces.

On MetaGaze_model input, the faces generated are even more primitive, with no traits of face aspects; in Iris and the Iris+Tesselation methods, the outputs were blank. Lastly,

MetaGaze+CelebA input had a FOV of 90° degrees, eyes half closed, and extreme camera and gaze angles. DeepPrivacy2 outputs a frontal smiling face with a frontal gaze, while the other methods generated faces with much noise, even considering the MetaGaze training methods.

5.4 KEY FINDINGS

In sum, considering the quantitative and qualitative analysis of the results from this work, we conclude that most methods have more difficulty preserving gaze direction in the following scenarios:

- Camera angles above 10° degrees;
- Gaze directions above 30° degrees;
- FOV of 60°, followed by FOV of 120°;
- Eyes semi-closed, followed by half-closed and wide-open;

To conclude, we summarize the pitfalls of gaze preservation in anonymized faces for vehicular scenarios. We validate that fine-tuning and refining the mesh input to highlight the iris region are both valid for reducing the gaze direction preservation error in the anonymized generated faces. We achieve the best result with the fine-tuning of CelebA+MetaGaze for the MetaGaze dataset and the Iris modification for the DMD dataset.

6 CONCLUSION

While anonymizing users in images, gaze preservation is a key capability for several application fields, such as the automotive sector. Our proposed annotated synthetic gaze dataset, MetaGaze, provides a benchmark for gaze preservation among face anonymization techniques. In addition to existing competitor datasets, MetaGaze covers various camera conditions and gaze angles, FOV, and eyelid openness variations in a well-balanced distribution, configuring a platform for investigating and potentially fine-tuning face generator models. In our experiments, we uncovered the scenarios where existing state-of-the-art methods show more difficulty in preserving gaze after anonymization, being: camera angles above 10° degrees, gaze directions above 30° degrees, FOV of 60° degrees, followed by 120° degrees, and eyes semi-closed, half-closed or even wide-opened. We also conducted additional experiments using a baseline anonymization method. We achieved improvements in reducing the error from 10.8° to 7.9° degrees in pitch and 6.4° to 5.9° degrees in yaw compared to the relative results of GANonymization regarding the preservation of the mean absolute error of gaze estimation, demonstrating the potential of MetaGaze in improving the performance of face generator models.

Limitations. The constraints of the GANonymization improvement suggestions still lie in their dependency on face mesh detection. The detection accuracy drops considerably as the pith and yaw get higher than 45 degrees. Although we trained the model using synthetic data from MetaGaze, results suggest it could surpass real training sets. However, we also include the DMD dataset (footage of drivers in real conditions) to evaluate beyond our benchmark. Regarding fairness, during the development of the MetaGaze, we selected 30 synthetic human models. The main criteria were to represent gender, ethnicity, and age group diversity. Further studies should be done to reduce the impact of lack of representation.

Future Work. Future work will involve expanding the benchmark validation by including a broader range of face anonymization methods. Additionally, there are plans to expand the number of attributes explored by MetaHumans (for example, light conditions or face expressions) and evaluate their impact on each face anonymization method. Another key area of exploration will be modifying the GANonymization loss function to consider the face geometry and make the GAN model training converge earlier and with more precision results, making the training more accessible for other researchers and developers.

REFERENCES

ABDELRAHMAN, A. A.; HEMPEL, T.; KHALIFA, A.; AL-HAMADI, A.; DINGES, L. L2cs-net: Fine-grained gaze estimation in unconstrained environments. In: IEEE. *2023 8th International Conference on Frontiers of Signal Processing (ICFSP)*. [S.l.], 2023. p. 98–102.

ADEBOYE, O.; DARGAHI, T.; BABAIE, M.; SARAEE, M.; YU, C.-M. Deepclean: a robust deep learning technique for autonomous vehicle camera data privacy. *IEEE Access*, IEEE, v. 10, p. 124534–124544, 2022.

BAEK, M.; MUN, J.; KIM, W.; CHOI, D.; YIM, J.; LEE, S. Driving environment perception based on the fusion of vehicular wireless communications and automotive remote sensors. *Sensors*, MDPI, v. 21, n. 5, p. 1860, 2021.

BAI, T.; FU, S.; YANG, Q. Privacy-preserving object detection with secure convolutional neural networks for vehicular edge computing. *Future Internet*, MDPI, v. 14, n. 11, p. 316, 2022.

BAI, T.; SHAO, D.; HE, Y.; FU, S.; YANG, Q. P^3: A privacy-preserving perception framework for building vehicle-edge perception networks protecting data privacy. In: *2023 32nd International Conference on Computer Communications and Networks (ICCCN)*. [S.l.: s.n.], 2023. p. 1–10.

BERA, S.; KHANDEPARKAR, K. Ai based real-time privacy-aware camera data processing in autonomous vehicles. In: *2023 IEEE International Symposium on Ethics in Engineering, Science, and Technology (ETHICS)*. [S.l.: s.n.], 2023. p. 1–5.

BODDETI, V. N.; SREEKUMAR, G.; ROSS, A. On the biometric capacity of generative face models. *arXiv preprint arXiv:2308.02065*, 2023.

BRKIć, K.; HRKAć, T.; KALAFATIć, Z. Protecting the privacy of humans in video sequences using a computer vision-based de-identification pipeline. *Expert Systems with Applications*, v. 87, p. 41–55, 2017. ISSN 0957-4174. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0957417417303986>.

CAI, Z.; GAO, Z.; PLANCHE, B.; ZHENG, M.; CHEN, T.; ASIF, M. S.; WU, Z. Disguise without disruption: Utility-preserving face de-identification. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. [S.l.: s.n.], 2024. v. 38, n. 2, p. 918–926.

CAI, Z.; XIONG, Z.; XU, H.; WANG, P.; LI, W.; PAN, Y. Generative adversarial networks: A survey toward private and secure applications. *ACM Computing Surveys (CSUR)*, ACM New York, NY, USA, v. 54, n. 6, p. 1–38, 2021.

CAO, J.; HU, Y.; YU, B.; HE, R.; SUN, Z. 3d aided duet gans for multi-view face image synthesis. *IEEE Transactions on Information Forensics and Security*, IEEE, v. 14, n. 8, p. 2028–2042, 2019.

CHEN, C.; LI, X.; YANG, L.; LIN, X.; ZHANG, L.; WONG, K.-Y. K. Progressive semantic-aware style transformation for blind face restoration. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2021. p. 11896–11905.

CHEN, J.; KONRAD, J.; ISHWAR, P. Vgan-based image representation learning for privacy-preserving facial expression recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. [S.l.: s.n.], 2018. p. 1570–1579.

Compliance Week. *Volkswagen Fined $11M Under GDPR for Unauthorized Data Collection*. 2023. <https://www.complianceweek.com/regulatory-enforcement/volkswagen-fined-11m-under-gdpr-for-unauthorized-data-collection/31903.article>. Accessed: 2024-08-15.

CONINCK, S. D.; WANG, W.-C.; LEROUX, S.; SIMOENS, P. Privacy-preserving visual analysis: training video obfuscation models without sensitive labels. *Applied Intelligence*, Springer, p. 1–12, 2024.

DAO, T. T.; VU, D. H.; PHAM, C.; TRAN, A. Efhq: Multi-purpose extremepose-face-hq dataset. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2024. p. 22605–22615.

DEMIR, U.; UNAL, G. Patch-based image inpainting with generative adversarial networks. *arXiv preprint arXiv:1803.07422*, 2018.

DENG, J.; GUO, J.; VERVERAS, E.; KOTSIA, I.; ZAFEIRIOU, S. Retinaface: Single-shot multi-level face localisation in the wild. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. [S.l.: s.n.], 2020. p. 5203–5212.

DUFRESNE-CAMARO, C.-O.; CHEVALIER, F.; AHMED, S. I. Computer vision applications and their ethical risks in the global south. In: *Graphics Interface 2020*. [S.l.: s.n.], 2020.

DWORK, C. Differential privacy. In: BUGLIESI, M.; PRENEEL, B.; SASSONE, V.; WEGENER, I. (Ed.). *Automata, Languages and Programming*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006. p. 1–12. ISBN 978-3-540-35908-1.

Epic Games. *MetaHuman Creator*. 2023. <https://www.unrealengine.com/en-US/metahuman>. Accessed: 2024-08-15.

Epic Games. *Unreal Engine*. 2023. <https://www.unrealengine.com/>. Accessed: 2024-08-15.

FAHMY, H.; PASTORE, F.; BAGHERZADEH, M.; BRIAND, L. Supporting deep neural network safety analysis and retraining through heatmap-based unsupervised learning. *IEEE Transactions on Reliability*, IEEE, v. 70, n. 4, p. 1641–1657, 2021.

FARD, A. P.; MAHOOR, M. H.; LAMER, S. A.; SWEENY, T. Ganalyzer: Analysis and manipulation of gans latent space for controllable face synthesis. *arXiv preprint arXiv:2302.00908*, 2023.

FU, B.; CHEN, C.; HENNIGER, O.; DAMER, N. A deep insight into measuring face image utility with general and face-specific image quality metrics. In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. [S.l.: s.n.], 2022. p. 905–914.

GEYER, J.; KASSAHUN, Y.; MAHMUDI, M.; RICOU, X.; DURGESH, R.; CHUNG, A. S.; HAUSWALD, L.; PHAM, V. H.; MÜHLEGG, M.; DORN, S. et al. A2d2: Audi autonomous driving dataset. *arXiv preprint arXiv:2004.06320*, 2020.

GHOSH, S.; DHALL, A.; SHARMA, G.; GUPTA, S.; SEBE, N. Speak2label: Using domain knowledge for creating a large scale driver gaze zone estimation dataset. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. [S.l.: s.n.], 2021. p. 2896–2905.

GOODFELLOW, I.; POUGET-ABADIE, J.; MIRZA, M.; XU, B.; WARDE-FARLEY, D.; OZAIR, S.; COURVILLE, A.; BENGIO, Y. Generative adversarial nets. In: GHAHRAMANI, Z.; WELLING, M.; CORTES, C.; LAWRENCE, N.; WEINBERGER, K. (Ed.). *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2014. v. 27. Disponível em: <https://proceedings.neurips.cc/paper_files/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf>.

GOODFELLOW, I.; POUGET-ABADIE, J.; MIRZA, M.; XU, B.; WARDE-FARLEY, D.; OZAIR, S.; COURVILLE, A.; BENGIO, Y. Generative adversarial nets. *Advances in neural information processing systems*, v. 27, 2014.

HE, X.; ZHU, M.; CHEN, D.; WANG, N.; GAO, X. Diff-privacy: Diffusion-based face privacy protection. *IEEE Transactions on Circuits and Systems for Video Technology*, IEEE, 2024.

HELLMANN, F.; MERTES, S.; BENOUIS, M.; HUSTINX, A.; HSIEH, T.-C.; CONATI, C.; KRAWITZ, P.; ANDRÉ, E. Ganonymization: A gan-based face anonymization framework for preserving emotional expressions. *ACM Transactions on Multimedia Computing, Communications and Applications*, ACM New York, NY, 2024.

HERASHCHENKO, D.; FARKAŠ, I. Appearance-based gaze estimation enhanced with synthetic images using deep neural networks. In: IEEE. *2023 IEEE Symposium Series on Computational Intelligence (SSCI)*. [S.l.], 2023. p. 129–134.

HO, J.; JAIN, A.; ABBEEL, P. *Denoising Diffusion Probabilistic Models*. 2020. Disponível em: <https://arxiv.org/abs/2006.11239>.

HONG, S.; RYU, J. Unsupervised face domain transfer for low-resolution face recognition. *IEEE Signal Processing Letters*, v. 27, p. 156–160, 2020.

HUANG, G. B.; MATTAR, M.; BERG, T.; LEARNED-MILLER, E. Labeled faces in the wild: A database forstudying face recognition in unconstrained environments. In: *Workshop on faces in'Real-Life'Images: detection, alignment, and recognition*. [S.l.: s.n.], 2008.

HUKKELÅS, H.; LINDSETH, F. Deepprivacy2: Towards realistic full-body anonymization. In: *Proceedings of the IEEE/CVF winter conference on applications of computer vision*. [S.l.: s.n.], 2023. p. 1329–1338.

HUKKELÅS, H.; LINDSETH, F. Does image anonymization impact computer vision training? In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2023. p. 140–150.

HUKKELÅS, H.; MESTER, R.; LINDSETH, F. Deepprivacy: A generative adversarial network for face anonymization. In: SPRINGER. *International symposium on visual computing*. [S.l.], 2019. p. 565–578.

HUYNH-THU, Q.; GHANBARI, M. Scope of validity of psnr in image/video quality assessment. *Electronics letters*, IET, v. 44, n. 13, p. 800–801, 2008.

ISOLA, P.; ZHU, J.-Y.; ZHOU, T.; EFROS, A. A. Image-to-image translation with conditional adversarial networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2017. p. 1125–1134.

JAIN, V.; LEARNED-MILLER, E. *FDDB: A Benchmark for Face Detection in Unconstrained Settings*. [S.l.], 2010.

JIANG, D.; SONG, D.; TONG, R.; TANG, M. Styleipsb: Identity-preserving semantic basis of stylegan for high fidelity face swapping. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2023. p. 352–361.

JO, B.; CHO, D.; PARK, I. K.; HONG, S. Ifqa: Interpretable face quality assessment. In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. [S.l.: s.n.], 2023. p. 3444–3453.

JOSHI, I.; GRIMMER, M.; RATHGEB, C.; BUSCH, C.; BREMOND, F.; DANTCHEVA, A. Synthetic data in human analysis: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, IEEE, 2024.

KARRAS, T.; AILA, T.; LAINE, S.; LEHTINEN, J. Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196*, 2017.

KARRAS, T.; AITTALA, M.; HELLSTEN, J.; LAINE, S.; LEHTINEN, J.; AILA, T. *Training Generative Adversarial Networks with Limited Data*. 2020.

KARRAS, T.; AITTALA, M.; LAINE, S.; HÄRKÖNEN, E.; HELLSTEN, J.; LEHTINEN, J.; AILA, T. Alias-free generative adversarial networks. *Advances in Neural Information Processing Systems*, v. 34, p. 852–863, 2021.

KARRAS, T.; LAINE, S.; AILA, T. *A Style-Based Generator Architecture for Generative Adversarial Networks*. 2019.

KARRAS, T.; LAINE, S.; AITTALA, M.; HELLSTEN, J.; LEHTINEN, J.; AILA, T. *Analyzing and Improving the Image Quality of StyleGAN*. 2020.

KARTYNNIK, Y.; ABLAVATSKI, A.; GRISHCHENKO, I.; GRUNDMANN, M. Real-time facial surface geometry from monocular video on mobile gpus. *arXiv preprint arXiv:1907.06724*, 2019.

KELLNHOFER, P.; RECASENS, A.; STENT, S.; MATUSIK, W.; TORRALBA, A. Gaze360: Physically unconstrained gaze estimation in the wild. In: *Proceedings of the IEEE/CVF international conference on computer vision*. [S.l.: s.n.], 2019. p. 6912–6921.

KLEMP, M.; RÖSCH, K.; WAGNER, R.; QUEHL, J.; LAUER, M. Ldfa: Latent diffusion face anonymization for self-driving applications. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2023. p. 3199–3205.

KLEMP, M.; RÖSCH, K.; WAGNER, R.; QUEHL, J.; LAUER, M. Ldfa: Latent diffusion face anonymization for self-driving applications. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2023. p. 3199–3205.

KOLODNY, L. *Tesla Workers Shared Sensitive Images Recorded by Customer Cars*. 2023. <https://www.reuters.com/technology/tesla-workers-shared-sensitive-images-recorded-by-customer-cars-2023-04-06/>. Accessed: 2024-08-15.

KUMAR, S. A.; YAGHOUBI, E.; DAS, A.; HARISH, B.; PROENÇA, H. The p-destre: A fully annotated dataset for pedestrian detection, tracking, and short/long-term re-identification from aerial devices. *IEEE Transactions on Information Forensics and Security*, IEEE, v. 16, p. 1696–1708, 2020.

KUNCHALA, A.; BOUROCHE, M.; SCHOEN-PHELAN, B. Towards a framework for privacy-preserving pedestrian analysis. In: *Proceedings of the IEEE/CVF winter conference on applications of computer vision*. [S.l.: s.n.], 2023. p. 4370–4380.

LI, A.; LI, G.; SUN, L.; WANG, X. *FaceFormer: Scale-aware Blind Face Restoration with Transformers*. 2022.

LI, H.; CHAUDHARI, P.; YANG, H.; LAM, M.; RAVICHANDRAN, A.; BHOTIKA, R.; SOATTO, S. *Rethinking the Hyperparameters for Fine-tuning*. 2020. Disponível em: <https://arxiv.org/abs/2002.11770>.

LI, T.; LIN, L. Anonymousnet: Natural face de-identification with measurable privacy. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. [S.l.: s.n.], 2019.

LI, Z.; YU, R.; DAS, A.; ZHANG, S.; GU, H.; WANG, X.; ZHOU, F.; SABIR, A.; AHMED, D.; ZAFAR, A. Inspire: Instance-level privacy-pre serving transformation for vehicular camera videos. In: *2023 32nd International Conference on Computer Communications and Networks (ICCCN)*. [S.l.: s.n.], 2023. p. 1–10.

LIU, C.; LIN, Q.; ZENG, Z.; PAN, Y. Emoface: Audio-driven emotional 3d face animation. In: IEEE. *2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*. [S.l.], 2024. p. 387–397.

LIU, X.; WEIJER, J. van de; BAGDANOV, A. D. Rankiqa: Learning from rankings for no-reference image quality assessment. In: *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. [S.l.: s.n.], 2017.

LIU, Z.; LUO, P.; WANG, X.; TANG, X. Deep learning face attributes in the wild. In: *Proceedings of International Conference on Computer Vision (ICCV)*. [S.l.: s.n.], 2015.

LYNSKEY, O. *The foundations of EU data protection law*. [S.l.]: Oxford University Press, 2015.

MAXIMOV, M.; ELEZI, I.; LEAL-TAIXÉ, L. Ciagan: Conditional identity anonymization generative adversarial networks. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. [S.l.: s.n.], 2020. p. 5447–5456.

MCMURROUGH, C. D.; METSIS, V.; RICH, J.; MAKEDON, F. An eye tracking dataset for point of gaze detection. In: *Proceedings of the Symposium on Eye Tracking Research and Applications*. [S.l.: s.n.], 2012. p. 305–308.

MEDEN, B.; ROT, P.; TERHöRST, P.; DAMER, N.; KUIJPER, A.; SCHEIRER, W. J.; ROSS, A.; PEER, P.; ŠTRUC, V. Privacy–enhancing face biometrics: A comprehensive survey. *IEEE Transactions on Information Forensics and Security*, v. 16, p. 4147–4183, 2021.

MIRJALILI, V.; RASCHKA, S.; NAMBOODIRI, A.; ROSS, A. Semi-adversarial networks: Convolutional autoencoders for imparting privacy to face images. In: IEEE. *2018 International Conference on Biometrics (ICB)*. [S.l.], 2018. p. 82–89.

MIRJALILI, V.; RASCHKA, S.; ROSS, A. Privacynet: Semi-adversarial networks for multi-attribute face privacy. *IEEE Transactions on Image Processing*, IEEE, v. 29, p. 9400–9412, 2020.

MIRZA, M.; OSINDERO, S. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014.

MORA, K. A. F.; MONAY, F.; ODOBEZ, J.-M. Eyediap: A database for the development and evaluation of gaze estimation algorithms from rgb and rgb-d cameras. In: *Proceedings of the symposium on eye tracking research and applications*. [S.l.: s.n.], 2014. p. 255–258.

MORVAY, B. T.; BéRES, B.; TORMA, S.; SZEGLETES, L. Diffusion probabilistic model based face anonymization in embedded environments. In: *2023 14th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*. [S.l.: s.n.], 2023. p. 000135–000140.

Mozilla Foundation. *It's Official: Cars Are the Worst Product Category We Have Ever Reviewed for Privacy*. 2023. <https://foundation.mozilla.org/en/privacynotincluded/articles/its-official-cars-are-the-worst-product-category-we-have-ever-reviewed-for-privacy/>. Accessed: 2024-08-15.

MY, K.; BAGDANOV, A.; BERTINI, M.; BIMBO, A. Domain adaptation for privacy-preserving pedestrian detection in thermal imagery. In: _____. [S.l.: s.n.], 2019. p. 203–213. ISBN 978-3-030-30644-1.

NEFF, C.; MENDIETA, M.; MOHAN, S.; BAHARANI, M.; ROGERS, S.; TABKHI, H. Revamp2t: Real-time edge video analytics for multicamera privacy-aware pedestrian tracking. *IEEE Internet of Things Journal*, v. 7, n. 4, p. 2591–2602, 2020.

OHNO, M.; UKYO, R.; AMANO, T.; RIZK, H.; YAMAGUCHI, H. Privacy-preserving pedestrian tracking with path image inpainting and 3d point cloud features. *Pervasive and Mobile Computing*, v. 100, p. 101914, 2024. ISSN 1574-1192. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1574119224000403>.

ORTEGA, J. D.; KOSE, N.; CAñAS, P.; CHAO, M.-A.; UNNERVIK, A.; NIETO, M.; OTAEGUI, O.; SALGADO, L. Dmd: A large-scale multi-modal driver monitoring dataset for attention and alertness analysis. In: BARTOLI, A.; FUSIELLO, A. (Ed.). *Computer Vision – ECCV 2020 Workshops*. [S.l.]: Springer International Publishing, 2020. p. 387–405. ISBN 978-3-030-66823-5.

PAN, X.; ZHAN, X.; DAI, B.; LIN, D.; LOY, C. C.; LUO, P. *Exploiting Deep Generative Prior for Versatile Image Restoration and Manipulation*. 2020.

POIRIER-GINTER, Y.; LALONDE, J.-F. Robust unsupervised stylegan image restoration. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2023. p. 22292–22301.

PURTOVA, N. The law of everything. broad concept of personal data and future of eu data protection law. *Law, Innovation and Technology*, Taylor & Francis, v. 10, n. 1, p. 40–81, 2018.

QIU, H.; JIANG, Y.; ZHOU, H.; WU, W.; LIU, Z. Stylefacev: Face video generation via decomposing and recomposing pretrained stylegan3. *arXiv preprint arXiv:2208.07862*, 2022.

REN, Z.; LEE, Y. J.; RYOO, M. S. Learning to anonymize faces for privacy preserving action detection. In: *Proceedings of the European Conference on Computer Vision (ECCV)*. [S.l.: s.n.], 2018.

RONNEBERGER, O.; FISCHER, P.; BROX, T. U-net: Convolutional networks for biomedical image segmentation. In: SPRINGER. *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*. [S.l.], 2015. p. 234–241.

SCHöNFELD, E.; SCHIELE, B.; KHOREVA, A. A u-net based discriminator for generative adversarial networks. In: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2020. p. 8204–8213.

SEITZER, M. *pytorch-fid: FID Score for PyTorch*. 2020. <https://github.com/mseitzer/pytorch-fid>. Version 0.3.0.

SHAMSHAD, F.; NASEER, M.; NANDAKUMAR, K. Clip2protect: Protecting facial privacy using text-guided makeup via adversarial latent search. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2023. p. 20595–20605.

SHAO, F.; LI, K.; LIN, W.; JIANG, G.; YU, M.; DAI, Q. Full-reference quality assessment of stereoscopic images by learning binocular receptive field properties. *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, v. 24, 05 2015.

SHEIKH, H.; BOVIK, A.; VECIANA, G. de. An information fidelity criterion for image quality assessment using natural scene statistics. *IEEE Transactions on Image Processing*, v. 14, n. 12, p. 2117–2128, 2005.

SIDDIQUI, J. R. Fexgan-meta: Facial expression generation with meta humans. *arXiv preprint arXiv:2203.05975*, 2022.

SILVA, T. S. A short introduction to generative adversarial networks. *https://sthalles.github.io*, 2017. Disponível em: <https://sthalles.github.io/intro-to-gans/>.

SKOROKHODOV, I.; TULYAKOV, S.; ELHOSEINY, M. Stylegan-v: A continuous video generator with the price, image quality and perks of stylegan2. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2022. p. 3626–3636.

SMITH, B. A.; YIN, Q.; FEINER, S. K.; NAYAR, S. K. Gaze locking: passive eye contact detection for human-object interaction. In: *Proceedings of the 26th annual ACM symposium on User interface software and technology*. [S.l.: s.n.], 2013. p. 271–280.

SUGANO, Y.; MATSUSHITA, Y.; SATO, Y. Learning-by-synthesis for appearance-based 3d gaze estimation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2014. p. 1821–1828.

SUN, Q.; MA, L.; OH, S. J.; GOOL, L. V.; SCHIELE, B.; FRITZ, M. *Natural and Effective Obfuscation by Head Inpainting*. 2018.

TREMBLAY, J.; PRAKASH, A.; ACUNA, D.; BROPHY, M.; JAMPANI, V.; ANIL, C.; TO, T.; CAMERACCI, E.; BOOCHOON, S.; BIRCHFIELD, S. Training deep networks with synthetic data: Bridging the reality gap by domain randomization. In: *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. [S.l.: s.n.], 2018. p. 969–977.

VILLANUEVA, A.; PONZ, V.; SESMA-SANCHEZ, L.; ARIZ, M.; PORTA, S.; CABEZA, R. Hybrid method based on topography for robust detection of iris center and eye corners. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, ACM New York, NY, USA, v. 9, n. 4, p. 1–20, 2013.

VORA, S.; RANGESH, A.; TRIVEDI, M. M. Driver gaze zone estimation using convolutional neural networks: A general framework and ablative analysis. *IEEE Transactions on Intelligent Vehicles*, IEEE, v. 3, n. 3, p. 254–265, 2018.

WANG, W.; NIU, L.; ZHANG, J.; YANG, X.; ZHANG, L. Dual-path image inpainting with auxiliary gan inversion. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2022. p. 11421–11430.

WANG, Z.; BOVIK, A. C.; SHEIKH, H. R.; SIMONCELLI, E. P. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, IEEE, v. 13, n. 4, p. 600–612, 2004.

WANG, Z.; ZHANG, J.; CHEN, R.; WANG, W.; LUO, P. *RestoreFormer: High-Quality Blind Face Restoration from Undegraded Key-Value Pairs*. 2022.

WEIDENBACHER, U.; LAYHER, G.; STRAUSS, P.-M.; NEUMANN, H. A comprehensive head pose and gaze database. In: IET. *2007 3rd IET International Conference on Intelligent Environments*. [S.l.], 2007. p. 455–458.

WEN, Y.; LIU, B.; DING, M.; XIE, R.; SONG, L. Identitydp: Differential private identification protection for face images. *Neurocomputing*, v. 501, p. 197–211, 2022. ISSN 0925-2312. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0925231222007597>.

WENGER, E.; SHAN, S.; ZHENG, H.; ZHAO, B. Y. Sok: Anti-facial recognition technology. In: IEEE. *2023 IEEE Symposium on Security and Privacy (SP)*. [S.l.], 2023. p. 864–881.

WU, Y.; YANG, F.; XU, Y.; LING, H. Privacy-protective-gan for privacy preserving face de-identification. *Journal of Computer Science and Technology*, Springer, v. 34, p. 47–60, 2019.

XIANG, A. Being" seen" versus" mis-seen": Tensions between privacy and fairness in computer vision. *Harv. JL & Tech.*, HeinOnline, v. 36, p. 1, 2022.

XIANG, A. Being'seen'vs.'mis-seen': Tensions between privacy and fairness in computer vision. *Harvard Journal of Law & Technology*, v. 36, n. 1, 2022.

XIONG, Z.; LI, W.; HAN, Q.; CAI, Z. Privacy-preserving auto-driving: A gan-based approach to protect vehicular camera data. In: *2019 IEEE International Conference on Data Mining (ICDM)*. [S.l.: s.n.], 2019. p. 668–677.

YANG, B.; YAN, J.; LEI, Z.; LI, S. Z. Fine-grained evaluation on face detection in the wild. In: IEEE. *Automatic Face and Gesture Recognition (FG), 11th IEEE International Conference on*. [S.l.], 2015.

YANG, J.; QUAN, X.; ZHANG, H. Novel gan inversion model with latent space constraints for face reconstruction. In: MANTORO, T.; LEE, M.; AYU, M. A.; WONG, K. W.; HIDAYANTO, A. N. (Ed.). *Neural Information Processing*. Cham: Springer International Publishing, 2021. p. 620–631.

YANG, S.; LUO, P.; LOY, C.-C.; TANG, X. Wider face: A face detection benchmark. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2016. p. 5525–5533.

YANG, T.; REN, P.; XIE, X.; ZHANG, L. Gan prior embedded network for blind face restoration in the wild. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2021. p. 672–681.

YANG, X.; DONG, Y.; PANG, T.; SU, H.; ZHU, J.; CHEN, Y.; XUE, H. Towards face encryption by generating adversarial identity masks. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. [S.l.: s.n.], 2021. p. 3897–3907.

YANG, Y.; GUPTA, A.; FENG, J.; SINGHAL, P.; YADAV, V.; WU, Y.; NATARAJAN, P.; HEDAU, V.; JOO, J. Enhancing fairness in face detection in computer vision systems by demographic bias mitigation. In: *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*. [S.l.: s.n.], 2022. p. 813–822.

YANG, Z.; CAI, Z.; MEI, H.; LIU, S.; CHEN, Z.; XIAO, W.; WEI, Y.; QING, Z.; WEI, C.; DAI, B.; WU, W.; QIAN, C.; LIN, D.; LIU, Z.; YANG, L. Synbody: Synthetic dataset with layered human models for 3d human perception and modeling. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. [S.l.: s.n.], 2023. p. 20282–20292.

YU, X.; PORIKLI, F. Ultra-resolving face images by discriminative generative networks. In: LEIBE, B.; MATAS, J.; SEBE, N.; WELLING, M. (Ed.). *Computer Vision - ECCV 2016 - 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part V*. Springer, 2016. (Lecture Notes in Computer Science, v. 9909), p. 318–333. Disponível em: <https://doi.org/10.1007/978-3-319-46454-1_20>.

ZHANG, R.; ISOLA, P.; EFROS, A. A.; SHECHTMAN, E.; WANG, O. The unreasonable effectiveness of deep features as a perceptual metric. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2018. p. 586–595.

ZHANG, W.; ZHAI, G.; WEI, Y.; YANG, X.; MA, K. Blind image quality assessment via vision-language correspondence: A multitask learning perspective. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2023. p. 14071–14081.

ZHANG, X.; SUGANO, Y.; FRITZ, M.; BULLING, A. Mpiigaze: Real-world dataset and deep appearance-based gaze estimation. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 41, n. 1, p. 162–175, 2017.

ZHANG, Y.; TSANG, I.; LUO, Y.; CHANGHUI, H.; LU, X.; YU, X. Recursive copy and paste gan: Face hallucination from shaded thumbnails. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PP, p. 1–1, 02 2021.

ZHAO, H.; GOU, Y.; LI, B.; PENG, D.; LV, J.; PENG, X. Comprehensive and delicate: An efficient transformer for image restoration. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2023. p. 14122–14132.

ZHAO, S.; LIU, Z.; LIN, J.; ZHU, J.-Y.; HAN, S. *Differentiable Augmentation for Data-Efficient GAN Training*. 2020.

ZHOU, S.; CHAN, K.; LI, C.; LOY, C. C. Towards robust blind face restoration with codebook lookup transformer. *Advances in Neural Information Processing Systems*, v. 35, p. 30599–30611, 2022.

ZHOU, Y.; XIONG, J.; BI, R.; TIAN, Y. Secure yolov3-spp: Edge-cooperative privacy-preserving object detection for connected autonomous vehicles. In: *2022 International Conference on Networking and Network Applications (NaNA)*. [S.l.: s.n.], 2022. p. 82–89.

ZHU, H.; WU, W.; ZHU, W.; JIANG, L.; TANG, S.; ZHANG, L.; LIU, Z.; LOY, C. C. CelebV-HQ: A large-scale video facial attributes dataset. In: *ECCV*. [S.l.: s.n.], 2022.

ZHU, S.; LIU, S.; LOY, C. C.; TANG, X. Deep cascaded bi-network for face hallucination. In: LEIBE, B.; MATAS, J.; SEBE, N.; WELLING, M. (Ed.). *Computer Vision - ECCV 2016 - 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part V*. Springer, 2016. (Lecture Notes in Computer Science, v. 9909), p. 614–630. Disponível em: <https://doi.org/10.1007/978-3-319-46454-1_37>.