PrivateGaze

Preserving User Privacy in Black-box Mobile Gaze Tracking Services

Du, Lingyu; Jia, Jinyuan; Zhang, Xucong; Lan, Guohao

DOI
10.1145/3678595

Publication date
2024

Document Version
Final published version

Published in
Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies

PrivateGaze: Preserving User Privacy in Black-box Mobile Gaze Tracking Services

LINGYU DU, Delft University of Technology, The Netherlands
JINYUAN JIA, The Pennsylvania State University, The United States
XUCONG ZHANG, Delft University of Technology, The Netherlands
GUOHAO LAN, Delft University of Technology, The Netherlands

Eye gaze contains rich information about human attention and cognitive processes. This capability makes the underlying technology, known as gaze tracking, a critical enabler for many ubiquitous applications and has triggered the development of easy-to-use gaze estimation services. Indeed, by utilizing the ubiquitous cameras on tablets and smartphones, users can readily access many gaze estimation services. In using these services, users must provide their full-face images to the gaze estimator, which is often a black box. This poses significant privacy threats to the users, especially when a malicious service provider gathers a large collection of face images to classify sensitive user attributes. In this work, we present PrivateGaze, the first approach that can effectively preserve users' privacy in black-box gaze tracking services without compromising gaze estimation performance. Specifically, we proposed a novel framework to train a privacy preserver that converts full-face images into obfuscated counterparts, which are effective for gaze estimation while containing no privacy information. Evaluation on four datasets shows that the obfuscated image can protect users' private information, such as identity and gender, against unauthorized attribute classification. Meanwhile, when used directly by the black-box gaze estimator as inputs, the obfuscated images lead to comparable tracking performance to the conventional, unprotected full-face images.

CCS Concepts: • **Security and privacy → Privacy protections**; • **Human-centered computing → Ubiquitous and mobile computing**.

Additional Key Words and Phrases: Mobile gaze estimation, black-box gaze tracking service, privacy preserving.

ACM Reference Format:
Lingyu Du, Jinyuan Jia, Xucong Zhang, and Guohao Lan. 2024. PrivateGaze: Preserving User Privacy in Black-box Mobile Gaze Tracking Services. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 8, 3, Article 99 (September 2024), 28 pages. https://doi.org/10.1145/3678595

1 Introduction

Human eye gaze information is pivotal in understanding human attention and the inner workings of the human brain. This unique capability makes the underlying technology, known as gaze tracking, a critical enabler for a wide range of ubiquitous and interaction applications [5, 42]. Examples include human activity recognition [7, 53, 54, 80], cognitive workload estimation [19, 28, 47], early detection of autism spectrum disorder [18, 25], and gaze-based human-computer interaction [23, 61, 65, 86], among many other applications [5, 41, 42].

Authors' Contact Information: Lingyu Du, Department of Software Technology, Delft University of Technology, Delft, The Netherlands, Lingyu.Du@tudelft.nl; Jinyuan Jia, College of Information Sciences and Technology, The Pennsylvania State University, State College, The United States, jinyuan@psu.edu; Xucong Zhang, Department of Intelligent Systems, Delft University of Technology, Delft, The Netherlands, xucong.zhang@tudelft.nl; Guohao Lan, Department of Software Technology, Delft University of Technology, Delft, The Netherlands, g.lan@tudelft.nl.

ACM 2474-9567/2024/9-ART99
https://doi.org/10.1145/3678595

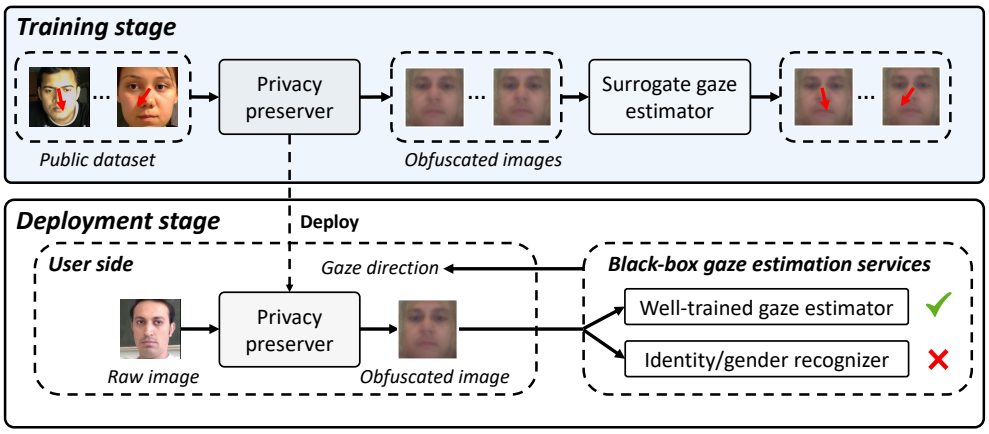


Fig. 1. An illustration of PrivateGaze, a framework to preserve users' privacy when they are using black-box gaze estimation services. The core of PrivateGaze is the privacy preserver, which transforms the original privacy-sensitive full-face image into an obfuscated version as input for the untrusted gaze estimation services. During the training stage, we train the privacy preserver with the assistance of a pre-trained surrogate gaze estimator. After training, the privacy preserve is deployed on the user's device to generate obfuscated images that can be used by the black-box gaze estimation services. This ensures accurate gaze estimation while preventing the user's private attributes, such as gender and identity, from being inferred by the service provider.

Thanks to the increasing demand for gaze-based applications, many vendors now provide affordable and easy-to-use gaze estimation services [12, 17, 24, 29, 73, 76, 85]. By utilizing general-purpose cameras, such as those embedded in tablets [35], mobile phones [49], and public displays [103], users can access gaze estimation service either through the cloud server, e.g., RealEye [73] and GazeRecorder [29], or by installing the software directly on their local devices, e.g., SeeSo [76], WeGaze [17], and EyeWare [24]. In both cases, users are required to share their full-face images with the service provider, who then takes the images as input to further locate eye regions [49, 59] or directly leverage the whole image [99, 100] for gaze estimation.

Typically, gaze estimation services are managed by commercial entities. This makes the end-to-end system, i.e., the image processing pipeline and the learning-based estimation model, an opaque black box to the user. When querying the gaze estimation services, users do not have any knowledge of how their face images are being processed, stored, or utilized. This problem becomes even more concerning given that facial images contain rich information about users' private attributes, such as identity and gender. Thus, when a malicious service provider has access to a large collection of unprotected face images, it can easily infer sensitive user information beyond the intended purpose, posing significant privacy threats to the users [50].

The privacy implications of gaze tracking have started to attract attention from the pervasive computing and eye-tracking communities [30, 57]. A large body of work [13, 22, 38, 39] focuses on the design of obfuscation techniques. These methods aim to eliminate sensitive user information from eye images to prevent iris-based user re-identification without compromising the utility of the estimated gaze data for downstream applications. Although these works have made significant progress in designing privacy-aware gaze-based applications, the question of how to preserve user's privacy in black-box gaze-tracking services remains an open challenge.

Besides the methods specifically designed for eye tracking, many privacy-preserving techniques have been developed for general-purpose image recognition and detection tasks. A popular solution is to train an encoder-based feature extractor via adversarial learning, so that it can capture task-related features from the original images while eliminating features related to the user's private attributes [58, 71, 93, 94]. However, these works

assume the privacy protector and the user have the full knowledge of the deep learning models utilized by the service provider, which is impractical for the real-world black-box gaze estimation services we considered. Another line of work adds small perturbations in the face images to obstruct unauthorized deep learning models from inferring private user attributes [69, 77]. However, the obfuscated images generated by these methods still contain a substantial amount of private user information that can be visually recognized by human eyes.

In this paper, we propose PrivateGaze, the first approach that can effectively preserve user's privacy in black-box gaze tracking services without compromising the estimation performance. An overview of PrivateGaze is shown in Figure 1. The core component is the proposed *privacy preserver*, which operates on the user's side to convert privacy-sensitive full-face images into privacy-enhanced obfuscated images that remain effective for gaze estimation. PrivateGaze comprises a suite of techniques we developed to tackle two major challenges.

First, the privacy preserver should eliminate features related to the user's private attributes, i.e., identity and gender, from the original full-face images. To resolve this challenge, we introduce a novel method that generates an average full-face image from a public dataset and leverages it as a template to transform images of different users, ensuring that the transformed versions exhibit a similar facial appearance akin to the average full-face image. Second, to maintain good gaze estimation performance, we need to ensure that the essential gaze-related information in the original images is preserved in the obfuscated images. Moreover, the obfuscated images should be readily compatible with the black-box gaze estimator, requiring no additional adaption from the service provider. To achieve this goal, we train a *surrogate gaze estimator* on a public dataset. As shown in Figure 1, we leverage the well-trained surrogate gaze estimator to encourage the privacy preserver to generate obfuscated images that contain features leading to the same gaze direction as the original images.

In summary, our major contributions are three-fold:

- We propose PrivateGaze, the first approach that can effectively preserve users' privacy when using black-box gaze tracking services without compromising the gaze estimation performance.
- We propose a novel framework to train a privacy preserver that can be deployed on the user's side to convert private-sensitive full-face images into privacy-enhanced obfuscated counterparts. The obfuscated images are effective for gaze estimation while containing no information about the user's private attributes.
- We conduct extensive experiments on four benchmark datasets to demonstrate the effectiveness of PrivateGaze. The results show that the obfuscated images produced by PrivateGaze can effectively protect users' private information, such as identity and gender, against unauthorized attribute inference. This protection remains robust even when the malicious attribute recognizer is trained on obfuscated images with correct attribute labels. Moreover, when used directly by the black-box gaze estimator as inputs, the obfuscated images achieve comparable tracking performance to the conventional, unprotected full-face images. Lastly, our system profiling shows that the proposed privacy preserver can be deployed on various computation platforms with low system costs in latency and memory usage.

Paper roadmap. The rest of the paper is organized as follows. In Section 2, we review related work and discuss the research gaps. We then present the detailed design of PrivateGaze in Section 3. Subsequently, we evaluate PrivateGaze in Section 4, followed by the conclusion in Section 5. The implementation of PrivateGaze is available at https://github.com/LingyuDu/PrivateGaze.

2 Related Work

2.1 Gaze Estimation

Gaze estimation methods can be generally categorized into model-based and appearance-based methods. Model-based methods [31, 32, 64, 107] infer gaze directions by constructing geometric models of the eyes from eye

images. By contrast, appearance-based methods [49, 83, 100, 101] directly estimate gaze directions from eye images or facial images captured by conventional cameras, such as those embedded in mobile devices [35, 49].

While earlier works in appearance-based methods [34, 100, 107] only took eye images as inputs, recent advancements [48, 49, 87, 99, 101] demonstrate that appearance-based methods can greatly benefit from information contained in facial regions and directly use full-face images for gaze estimation. More recently, the availability of large-scale datasets, such as GazeCapture [49] and ETHXGaze [99], combined with advances in deep-learning techniques, have significantly propelled appearance-based gaze estimation forward, enabling more accurate gaze prediction in complex environments with diverse backgrounds and lighting conditions.

The benefits of appearance-based gaze estimation are expected to encourage more users to utilize third-party appearance-based gaze estimation services for developing gaze-based applications, such as user attention estimation [1], cognitive context sensing [52, 89], and analyzing user interactions [8, 63]. However, these gaze estimation services often appear as a black box to users for commercial purposes, where users can only utilize the services without access to detailed information about how they operate. Such black-box gaze estimation services raise significant privacy concerns, as facial images contain rich private information about users. Our work is the first to preserve user privacy when utilizing black-box gaze estimation services.

2.2 Privacy-Preserving Methods in the Image Domain

There have been several approaches to preserving user privacy in images across various application scenarios. For example, Oh et al. [69] employ adversarial image perturbations to raw images to confuse deep learning-based classifiers from recognizing the user's identity. Shan et al. [77] propose adding imperceptible perturbations to images before their release, causing identity recognizers trained on these perturbed images to misidentify normal images. However, these perturbed images still retain private attributes and can be easily identified by recognizers trained on the perturbed dataset. In our work, we pursue a stringent privacy objective where attackers cannot infer private attributes from obfuscated images, even when deep learning classifiers have been trained on these obfuscated images with correct attribute labels.

Previous works [58, 71] have explored approaches where instead of adding perturbations to raw images, they focus on training an encoder to extract features from raw images that are useful for utility tasks while excluding information related to private attributes. For instance, Liu et al. [58] utilize adversarial learning to jointly train an encoder, a task-related model, and a private attribute recognizer. The task-related model and the private attribute recognizer are built upon the features extracted by the encoder from raw images. Wu et al. [93] propose DAPter, a method aimed at preserving user privacy during the utilization of task-related inference services on cloud platforms. However, their approach requires access to the task-related model and involves modifying its parameters for effective training, which is impractical for black-box models in real-world scenarios. Our work differs from these approaches by focusing on a more practical scenario where users have no knowledge of the deep learning model employed by the service provider for gaze estimation.

Face swapping [6, 11, 43, 66–68, 70, 91, 106] and face de-identification [37, 51, 62, 96] are potential methods to preserve user privacy by altering the identities of subjects in raw images. However, these techniques may retain other user attributes in the synthesized images [6, 51, 66, 91, 96], such as facial expressions and emotions, which are also privacy-sensitive information that users may wish to protect. By contrast, our method exclusively maintains gaze features in the synthesized images, offering robust privacy protection for users when utilizing gaze estimation services. Moreover, we observe that synthesized images generated by face swapping from target images of different subjects often exhibit distinct appearances [6, 43, 66], and similarly, de-identified images for different subjects may show visual differences [37]. We believe these inherent properties could be exploited by adversaries to classify users' identities if they can stealthily collect some synthesized images. In our approach,

synthesized images from different subjects maintain similar appearances, and our experiments show that the adversaries cannot correctly classify users' identities even with access to stealthily collected synthesized images.

2.3 Privacy-preserving Solutions for Eye-tracking Systems

Privacy-preserving eye tracking has emerged as a significant research topic in recent years due to growing privacy concerns associated with various stages of the eye-tracking pipeline. We categorize existing works [4, 14, 21, 22, 38, 39, 55, 81, 82] into three groups.

The first group focuses on the data collection stage, aiming to protect users' privacy-sensitive data by preventing its transmission to a central server. For example, Elfares et al. [21] utilize federated learning to train a deep learning-based gaze estimator. This approach retains raw data locally with users to preserve privacy and sends only minimal updates necessary for gaze estimation to the central server. Steil et al. [82] propose a method to protect the privacy of users and bystanders from scene images captured by a head-mounted eye tracker. Specifically, they develop a method to disable the scene camera upon detecting privacy-sensitive situations and automatically reactivate it when eye movement patterns change.

The second group of works focuses on preserving user privacy in the gaze estimation stage, aiming to mitigate the presence of private attributes in the images used by gaze estimators. For example, John et al. [39] introduce pixel-level noise to eye images captured by eye-tracking cameras, effectively thwarting iris authentication attacks. Eskildsen et al. [22] propose various methods to obfuscate eye images, including adding noise and applying non-linear low-pass filters, to prevent identification based on iris patterns. Additionally, John et al. [38] propose a hardware-based solution to remove bio-metric information from eye images by inducing optical defocus. This is achieved by increasing the distance between the eyes and the eye-tracking cameras, thereby intentionally blurring the iris region. Bozkir et al. [4] train a support vector regression model to estimate gaze direction from synthetic eye images, thereby preserving personal information for users.

The last group focuses on preserving the private attributes of users contained in the gaze data obtained by gaze estimators. David-John et al. [14] explore various privacy mechanisms such as adding Gaussian noise and temporal downsampling to reduce user identification accuracy based on gaze data features like fixations and saccades. Steil et al. [81] apply differential privacy by adding noise to features extracted from gaze data. They demonstrate that their approach prevents attackers from accurately identifying the user's identity and gender from gaze trajectories, while still maintaining good performance in gaze-based document type recognition tasks. Li et al. [55] propose a framework that directly applies differential privacy to raw gaze data. Their method can integrate with existing eye-tracking ecosystems and operate in real time, enhancing privacy protection during data processing stage.

In this paper, we focus on addressing privacy concerns during the gaze estimation stage. We propose a novel method to remove private attributes from full-face images while maintaining comparable performance in gaze estimation for black-box gaze estimation services.

3 Method

In this section, we introduce a novel method, PrivateGaze, that can convert a normal full-face image into a privacy-perceived full-face image. With the privacy information removed, the privacy-perceived full-face image still contains sufficient information for a black-box gaze estimation service to perform the gaze estimation task. In the following, we first define the threat model to formulate the problem and then detail the design of PrivateGaze.

3.1 Threat Model

Black-box gaze estimator. With recent developments in gaze estimation, it has become common to include the full-face image of the user as input to the methods [49, 101]. While existing works in privacy preservation [58,

71, 93] assume the details of the deep neural network used by the service provider are known, we consider a more practical case where the gaze estimator $\mathcal{G}_b(\cdot)$ is performed by a black-box, deep learning-based model. Specifically, the black-box gaze estimator $\mathcal{G}_b(\cdot)$ *is trained on an unknown dataset* $\mathcal{D}_b$ that contains raw full-face images and gaze annotations. Users can access gaze estimation services either through the cloud server or by installing the system directly on their local devices. In both cases, the user can only query and request $\mathcal{G}_b(\cdot)$ for service and has no knowledge about its implementation and training details.

Privacy concerns. The end-to-end gaze estimation system, including the processing pipeline and the deep learning-based gaze estimation model, makes the gaze estimation services untrustworthy. The full-face image of the user can be illegally used for purposes beyond gaze estimation, such as classifying the user's private attributes like identity and gender. This risk persists even if the gaze estimation system is running on local devices, as the gaze estimation services can stealthily share the data collected from the user with the malicious service provider when the local devices are connected to the internet. Therefore, the user would like to remove the private information contained in the full-face images before using them to call the gaze estimation services, without sacrificing gaze estimation performance.

Capabilities and goals of the malicious service provider. We assume the malicious service provider can stealthily collect a dataset $\mathcal{D}_p$, comprising images submitted by users for gaze estimation service, along with annotations of private user attributes such as identity and gender. Subsequently, $\mathcal{D}_p$ is used to train classifiers aimed at discerning users' private attributes from images that do not belong to $\mathcal{D}_p$.

Our goals. In this work, we envision a trustworthy party that provides a privacy preserver $\mathcal{P}(\cdot)$ to protect the user's privacy. As shown in Figure 1, during the deployment stage, $\mathcal{P}(\cdot)$ converts the user's original full-face images x into obfuscated images x' that do not contain information related to the user's attributes, such as identity and gender. The user then directly calls the black-box gaze estimator $\mathcal{G}_b(\cdot)$ with the obfuscated image x'. Formally, the obfuscated image x' must fulfill the objectives of preserving the user's privacy while ensuring good gaze estimation performance:

- **Privacy goal:** The obfuscated image x' cannot be used to correctly classify private attributes of the user, such as identity and gender, even if the malicious service provider trains deep learning-based classifiers on $\mathcal{D}_p$, i.e., a set of x' with accurate labels for these confidential user attributes.
- **Utility goal:** The obfuscated image x' can be directly used by $\mathcal{G}_b(\cdot)$ without any adaption needed from the service provider's side. The gaze estimation performance of $\mathcal{G}_b(\cdot)$ with x' should be similar to the original full-face images.

Assumption. We assume a public gaze estimation dataset $\mathcal{D}_w$ is available, which contains training examples (x_i, g_i), where x_i is the full-face image and g_i is the corresponding gaze annotation. The dataset $\mathcal{D}_w$ will be used to train $\mathcal{P}(\cdot)$. Note that $\mathcal{D}_w$ is different from $\mathcal{D}_b$, as we do not know which dataset has been used by the service provider to train the black-box gaze estimator $\mathcal{G}_b(\cdot)$.

3.2 PrivateGaze

To achieve the design goals, we propose a novel framework PrivateGaze consisting of a privacy preserver, an anchor image generation module, and the surrogate gaze estimator as shown in Figure 2. The privacy preserver $\mathcal{P}(\cdot)$ converts unprotected raw images x into obfuscated images x' to protect the private information of users, such as gender and identity contained in x. To achieve this goal, the privacy preserver ensures that x', converted from different x (images from different subjects), will exhibit similar facial appearances akin to a pre-generated average full-face image called the anchor image $\hat{x}$. We devise the anchor image generation module (in Section 3.2.1) to generate the $\hat{x}$ from the $\mathcal{D}_w$.

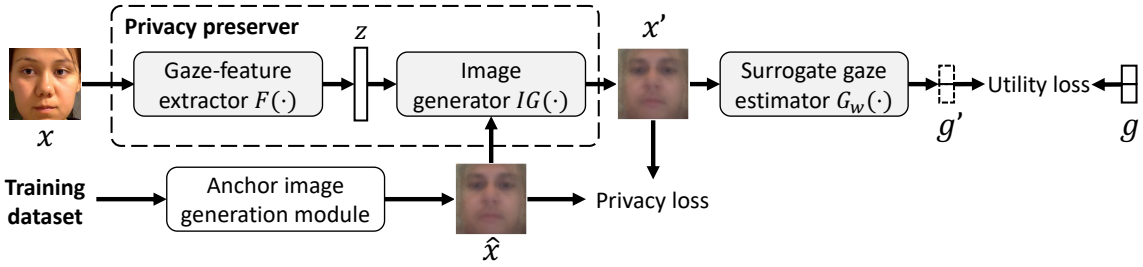


Fig. 2. An overview of PrivateGaze, which comprises the privacy preserver, the anchor image generation module, and the surrogate gaze estimator $\mathcal{G}_w(\cdot)$ trained on the training dataset $\mathcal{D}_w$. The privacy preserver includes the gaze-feature extractor $F(\cdot)$ and the image generator $IG(\cdot)$. $F(\cdot)$ extracts gaze features z from the raw images x in the training dataset. $IG(\cdot)$ takes z and a pre-generated image $\hat{x}$ as inputs to form the obfuscated images x'. $\hat{x}$ serves as the *anchor image* and is crafted from the training dataset using the proposed anchor image generation module. Subsequently, we compute the *privacy loss* based on $\hat{x}$ and x' to train $\mathcal{P}(\cdot)$ for the privacy objective. x' is then passed to $\mathcal{G}_w(\cdot)$ to obtain the estimated gaze direction g'. Finally, we calculate the *utility loss* based on the gaze annotations g and g' to train the privacy preserver for the utility objective.

To achieve the utility goal, the privacy preserver $\mathcal{P}(\cdot)$ is designed to extract gaze features z from x and generate x' that maintains these features for effective gaze estimation. Specifically, $\mathcal{P}(\cdot)$ consists of the gaze-feature extractor $F(\cdot)$ and the image generator $IG(\cdot)$. $F(\cdot)$ extracts gaze features z from the input x (detailed in Section 3.2.2). $IG(\cdot)$ takes z along with $\hat{x}$ as inputs to generate the privacy-preserved x' (detailed in Section 3.2.3). The generated x' has a similar appearance to $\hat{x}$ while preserving the gaze-related information z from x for accurate gaze estimation. To train $\mathcal{P}(\cdot)$, we construct a surrogate gaze estimator $\mathcal{G}_w(\cdot)$ trained on the $\mathcal{D}_w$, which performs the gaze estimation training with input x'. In this way, we are able to maximize the information in x' for the gaze estimation task.

3.2.1 Anchor image generation module. Below, we present a novel method for generating the anchor image $\hat{x}$ from a public dataset. The anchor image serves as a template for the obfuscated images x', ensuring they exhibit a facial appearance similar to $\hat{x}$. This allows us to manipulate the appearances of x' to preserve user's privacy while achieving the utility goal.

A major challenge in achieving this utility goal is training $\mathcal{P}(\cdot)$ with the surrogate gaze estimator $\mathcal{G}_w(\cdot)$, while aiming for good gaze estimation performance on the black-box gaze estimator $\mathcal{G}_b(\cdot)$. To address this challenge, we carefully generate the anchor image $\hat{x}$ to ensure that both $\mathcal{G}_w(\cdot)$ and $\mathcal{G}_b(\cdot)$ yield similar gaze estimation results on $\hat{x}$. This strategy enables $\mathcal{G}_w(\cdot)$ and $\mathcal{G}_b(\cdot)$ to achieve comparable gaze estimation performance on the obfuscated images x', as they share similar appearances with the anchor image.

In detail, $\hat{x}$ is an average full-face image created by blending facial images selected from the training dataset. The design ensures that x' synthesized from $\hat{x}$ does not closely resemble any individual subject. Then, we design the method to obtain $\hat{x}$ that can lead to similar gaze estimation results from $\mathcal{G}_w(\cdot)$ and $\mathcal{G}_b(\cdot)$. Specifically, we first randomly sample N raw images $\{x_i\}_{i=1}^N$ from the $\mathcal{D}_w$ and use them to query both $\mathcal{G}_w(\cdot)$ and $\mathcal{G}_b(\cdot)$, where the two gaze estimators return the corresponding gaze estimation results $\{\mathcal{G}_w(x_i)\}_{i=1}^N$ and $\{\mathcal{G}_b(x_i)\}_{i=1}^N$, respectively. We then calculate the $L1$ norm between the gaze directions estimated by $\mathcal{G}_w(\cdot)$ and $\mathcal{G}_b(\cdot)$ for each image x_i to obtain the list $\{D_g(x_i)\}_{i=1}^N$, where $D_g(x_i) = |\mathcal{G}_w(x_i) - \mathcal{G}_b(x_i)|_1$. After that, we sort the list $\{D_g(x_i)\}_{i=1}^N$ in the ascending order. We use $\{x_k\}_{k=1}^N$ to denote the set of raw images after sorting, which satisfies $D_g(x_k) \leq D_g(x_{k+1})$. We use $Ave(m)$ to denote the average full-face image calculated from the first m images in $\{x_k\}_{k=1}^N$ by $Ave(m) = \frac{1}{m}\sum_{k=1}^m x_k$. We generate M average full-face images by varying m from $K_1 + 1$ to $K_1 + M$, where $K_1 > 1$ and $K_1 + M \leq N$. Finally, we query $\mathcal{G}_w(\cdot)$ and $\mathcal{G}_b(\cdot)$ with the M average full-face images and select

Algorithm 1 Anchor image generation

Input: $\mathcal{D}_w$, $\mathcal{G}_b(\cdot)$, $\mathcal{G}_w(\cdot)$, $K_1 = 15$, $m = 1$, $M = 35$, $N = 500$, and a list of candidate anchor image $Anc = \{\}$.

Randomly sample N images from $\mathcal{D}_w$ to form $\{x_i\}_{i=1}^N$;
Query both $\mathcal{G}_b(\cdot)$ and $\mathcal{G}_w(\cdot)$ with $\{x_i\}_{i=1}^N$ to obtain the list $\{D_g(x_i)\}_{i=1}^N$, where $D_g(x_i) = |\mathcal{G}_w(x_i) - \mathcal{G}_b(x_i)|_1$;
Sort $\{D_g(x_i)\}_{i=1}^N$ in the ascending order;
Obtain the set of raw images after sorting $\{x_k\}_{k=1}^N$, where $D_g(x_k) \leq D_g(x_{k+1})$;
while $K_1 + m \leq M$ **do**
    Calculate the average full-face image $Ave(m) = \frac{1}{m}\sum_{k=1}^{m} x_k$;
    Anc.append($Ave(m)$);
    $m \leftarrow m + 1$;
end while
Query $\mathcal{G}_b(\cdot)$ and $\mathcal{G}_w(\cdot)$ with $Anc = \{Ave(m)\}_{m=1}^M$ to obtain the list $\{D_g(Ave(m))\}_{m=1}^M$;
$\hat{x} = Ave(\hat{m})$, where $\hat{m} = \arg\min_m \{D_g(Ave(m))\}_{m=1}^M$;

Output: The anchor image $\hat{x}$.

the one that leads to the minimum L_1 norm between the outputs of $\mathcal{G}_w(\cdot)$ and $\mathcal{G}_b(\cdot)$ as the anchor image $\hat{x}$. We set $K_1 = 15$, $M = 35$, and $N = 500$. The number of queries for $G_b(\cdot)$ is determined by N. The parameters K and M determine the minimum and maximum number of full-face images utilized in generating the anchor image, respectively. A higher value of K ensures the anchor image to be distinct from any single subject, whereas a larger M may include undesired full-face images, potentially resulting in significantly different gaze estimation results for $\mathcal{G}_w(\cdot)$ and $\mathcal{G}_b(\cdot)$ during the anchor image generation process. We show the $\hat{x}$ obtained from GazeCapture dataset [49] in Figure 3 and summarize the procedure of generating the anchor image in Algorithm 1.

3.2.2 Gaze-feature extractor. To ensure the obfuscated images x' are effective for gaze estimation, $\mathcal{P}(\cdot)$ extracts the gaze features z from x using the gaze-feature extractor $F(\cdot)$. As shown in the left part of Figure 3, $F(\cdot)$ comprises the gaze-aware encoder $E(\cdot)$ and the gaze projector $G(\cdot)$. Specifically, $E(\cdot)$ takes x as input and outputs a feature map z. To encourage $E(\cdot)$ in capturing the most essential and meaningful gaze-related features, z is fed into a nonlinear gaze projector $G(\cdot)$ to estimate gaze direction. We define the *gaze estimation loss* in training $E(\cdot)$ and $G(\cdot)$ as follow:

$$\mathcal{L}_g = \sum_{(x_i, g_i) \in \mathcal{D}_w} \ell(G(E(x_i)), g_i), \tag{1}$$

where $\ell(\cdot)$ is the L_1 loss function. The $G(\cdot)$ will be discarded in the deployment stage, and only the gaze features z are sent to the image generator $IG(\cdot)$.

3.2.3 Image generator. The image generator $IG(\cdot)$ is designed to synthesize the obfuscated images x' from gaze features z such that x' can be effectively used for gaze estimation by $\mathcal{G}_b(\cdot)$ while not containing private attributes from z. To achieve this goal, $IG(\cdot)$ generates x' with appearances similar to the anchor image $\hat{x}$ while preserving the gaze features extracted from x. Specifically, $IG(\cdot)$ takes both z and $\hat{x}$ as inputs to generate x'.

The structure of $IG(\cdot)$ is depicted in the right part of Figure 3, which adopts an encoder-decoder architecture. The encoder comprises several convolutional blocks (Conv Block) and takes $\hat{x}$ as input to produce a feature map $\hat{z}$ that has the same spatial dimension as the gaze features z extracted by the gaze-aware encoder $E(\cdot)$. $\hat{z}$ and z are concatenated and then fed into the decoder. Similar to the encoder, the decoder comprises several up-convolutional blocks (Up-conv Block). Each up-convolutional block involves upsampling the feature map followed by several convolutional layers. In our current design, each convolutional block consists of two convolutional layers, and each up-convolutional block has three convolutional layers. As shown in Figure 3, to ensure that x' can closely

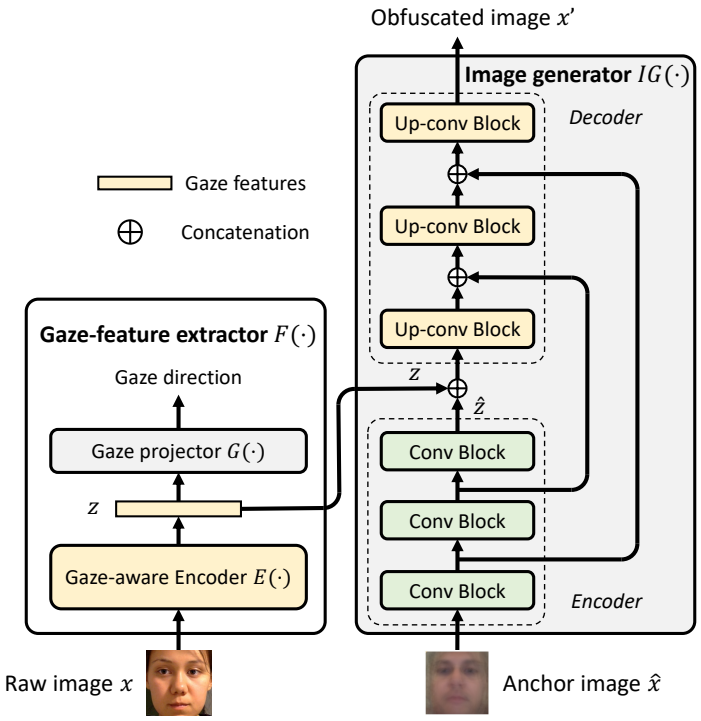


Fig. 3. The overall design of the privacy preserver $\mathcal{P}(\cdot)$, which consists of the gaze-feature extractor $F(\cdot)$ and the image generator $IG(\cdot)$. $F(\cdot)$ extracts gaze features z from the raw image x of the user. $IG(\cdot)$ takes the extracted gaze features z and the anchor image $\hat{x}$ as inputs to generate the privacy-preserved obfuscated image x'. x' has a similar appearance to $\hat{x}$ while retaining the gaze features extracted from x. Only the components with color-coded yellow will be deployed on the user's device after training for privacy preservation.

resemble $\hat{x}$, we concatenate the output of each up-convolutional block with the corresponding feature map from the encoder. The resulting combined feature map is then used as the input for the next up-convolutional block in the sequence.

We denote the process of generating x' from z and $\hat{x}$ as $x' = IG(\{z, \hat{x}\})$. Note that in Figure 3, we use three convolutional blocks and three up-convolutional blocks to illustrate the structure of $IG(\cdot)$. However, in practice, the number of these convolutional and up-convolutional blocks can vary based on design considerations and we are using four convolutional blocks and four up-convolutional blocks in our current design.

3.3 Training of the Privacy Preserver

First, to achieve the privacy objective, it is crucial that x', generated from different raw images, maintains a uniform appearance similar to $\hat{x}$. Therefore, we define the privacy loss as follows:

$$\mathcal{L}_{privacy} = \sum_{(x_i, g_i) \in \mathcal{D}_w} 1 - \text{MS-SSIM}\left(\hat{x}, IG\left(\{E(x_i), \hat{x}\}\right)\right), \tag{2}$$

where MS-SSIM$(\cdot, \cdot)$ is a function that calculates the multi-scale structural similarity [88], measuring the similarity between two images with values in the range of $[0, 1]$. A larger value of MS-SSIM indicates greater similarity between the two images.

Algorithm 2 Training algorithm of the privacy preserver

Input: $\mathcal{D}_w$, $\mathcal{G}_w(\cdot)$, and $\hat{x}$ generated by Algorithm 1.

1: Randomly initialize the parameters ψ of $\mathcal{P}(\cdot)$;
2: **for** each training step **do**
3: Sample the minibatch B of N_b images from $\mathcal{D}_w$;
4: Calculate the privacy loss over B by $\mathcal{L}_{privacy} = \frac{1}{N_b} \sum_{(x_i,g_i)\in B} 1 - \text{MS-SSIM}\left(\hat{x}, IG\left(\{E(x_i), \hat{x}\}\right)\right)$;
5: Calculate the utility loss over B by $\mathcal{L}_{utility} = \frac{1}{N_b} \sum_{(x_i,g_i)\in B} \ell\left(\mathcal{G}_w\left(IG\left(\{E(x_i), \hat{x}\}\right)\right), g_i\right) + \mathcal{L}_g$;
6: Update ψ by gradient descent: minimize $\mathcal{L}_{utility} + \lambda \mathcal{L}_{privacy}$;

Output: $\mathcal{P}(\cdot)$ with trained parameters ψ.

Second, to achieve the utility objective, we utilize x' as input to $\mathcal{G}_w(\cdot)$ to obtain the estimated gaze direction g' for x', and then train $\mathcal{P}(\cdot)$ to ensure g' closely approximates g. $\mathcal{G}_w(\cdot)$ is trained on raw images, and its parameters are frozen during the training of $\mathcal{P}(\cdot)$. Besides, the gaze estimation loss $\mathcal{L}_g$, defined in Equation 1, encourages the extraction of gaze features from raw images, thereby contributing to the utility objective. For training the privacy preserver to achieve the utility objective, we define the following utility loss:

$$\mathcal{L}_{utility} = \sum_{(x_i,g_i)\in\mathcal{D}_w} \ell\left(\mathcal{G}_w\left(IG\left(\{E(x_i), \hat{x}\}\right)\right), g_i\right) + \mathcal{L}_g, \tag{3}$$

where $\ell(\cdot)$ is the L_1 loss function. The first term of $\mathcal{L}_{utility}$ aims to minimize the L_1 norm between $\mathcal{G}_w(x')$ and g. The second term encourages the extraction of gaze features to generate x'.

Putting them all together, the final optimization objective for training $\mathcal{P}(\cdot)$ is the weighted sum of $\mathcal{L}_{utility}$ and $\mathcal{L}_{privacy}$:

$$\mathcal{L} = \mathcal{L}_{utility} + \lambda \mathcal{L}_{privacy}, \tag{4}$$

where λ is the weight that balances the trade-off between the utility and privacy objectives. PrivateGaze optimizes $\mathcal{P}(\cdot)$ by minimizing $\mathcal{L}$. During training, PrivateGaze samples a minibatch from $\mathcal{D}_w$ to calculate $\mathcal{L}$, and trains $\mathcal{P}(\cdot)$ by gradient descent. We summarize the training procedure for $\mathcal{P}(\cdot)$ in Algorithm 2.

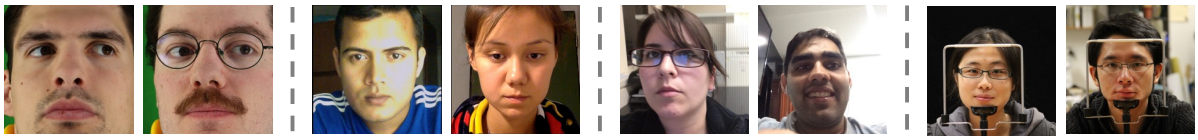
3.4 Deployment of the Privacy Preserver

In the deployment stage, the raw images collected from users are initially transformed into obfuscated images using the trained privacy preserver. Users then use these obfuscated images when calling the black-box gaze estimation services to protect their privacy. To reduce the computational cost, only specific components of the privacy preserver (highlighted in yellow in Figure 3) need to be deployed on the user's device. Specifically, for the gaze-feature extractor, only the gaze-aware encoder needs to be deployed, as the image generator exclusively utilizes gaze features from this encoder. For the image generator, because the anchor image remains consistent for different raw images, the encoder can be omitted. Instead, the feature maps generated by each convolutional block of the encoder are preserved as *appearance features*. Consequently, deployment requires only the decoder of the image generator and the appearance features to generate obfuscated images.

By deploying only these essential components, computational resources are optimized while achieving the system goal. The evaluation section includes a latency measurement of the privacy preserver, demonstrating its suitability for deployment across various hardware platforms without introducing much computational latency.

4 Evaluation

In this section, we conduct a comprehensive evaluation of PrivateGaze. We first introduce the datasets, followed by the methods for comparison and evaluation metrics. We then present the evaluation results on privacy and



(a) ETHXGaze (b) MPIIFaceGaze (c) GazeCapture (d) ColumbiaGaze

Fig. 4. Illustration of images sampled from the four gaze estimation datasets. Our selection of datasets covers a broad spectrum of mobile gaze tracking scenarios: from smartphone usage (GazeCapture) to laptop use cases (MPIIFaceGaze), and to ubiquitous web cameras (ETHXGaze and ColumbiaGazze) that widely appear in many daily devices.

utility objectives. Next, we conduct ablation studies to investigate the impact of different design choices on the performance of PrivateGaze. Finally, we evaluate system performance, measuring the processing time and memory usage when deploying the proposed privacy preserver on various hardware platforms.

4.1 Datasets

We consider the following four public gaze estimation datasets in our evaluation:

ETHXGaze [99] is a comprehensive dataset collected from 110 subjects in a laboratory environment, showcasing a wide range of head positions, lighting conditions, and individual appearances. It includes one training set and two testing sets. Our evaluation only utilizes the training set, which contains 80 subjects, as it is the only subset with gaze annotations. The images in this set have a resolution of 224×224.

GazeCapture [49] is a large-scale dataset collected from over 1,450 individuals in real-world environments. It includes nearly 2.5 million images taken with the front-facing cameras of smartphones, displaying a wide array of lighting conditions, head poses, user appearances, and backgrounds. In our preprocessing of this dataset, we adopted the method outlined in [105] to normalize the facial images, initially bringing them to a resolution of 128×128. Subsequently, we resized these images to a resolution of 224×224.

MPIIFaceGaze [101] is a dataset of full-face images from 15 subjects, including nine males and six females. The images were captured during the participants' routine laptop usage. It includes 3,000 images per subject, featuring a diverse range of head positions, lighting environments, and backgrounds. For our purposes, we utilize the normalized version of the dataset, as released by the authors. This normalized dataset maintains an image resolution of 224×224.

ColumbiaGaze [79] was gathered in a controlled laboratory setting, and features data from 56 subjects, including 32 males and 24 females. It is unique for its structured capture of five distinct head poses for each subject. The dataset comprises a total of 105 images per subject, representing combinations of three vertical and seven horizontal gaze directions. In our processing of this dataset, we extract the facial region from the original images and subsequently resize these cropped patches to a uniform resolution of 224×224.

Figure 4 showcases full-face images sampled from the four datasets under discussion. These images represent a wide range of scenarios in mobile gaze tracking, from the use of front-facing smartphone cameras (GazeCapture) to laptop-based interactions (MPIIFaceGaze). Moreover, the ETHXGaze and ColumbiaGaze datasets represent scenarios where web cameras, commonly found in a variety of ubiquitous devices, are used for gaze tracking.

4.2 Comparison Methods

We compare PrivateGaze with the following six methods:

Targeted Projected Gradient Descent Attack (TPGD): In this method, we first feed the original image x_i from the testing dataset to the surrogate gaze estimator $\mathcal{G}_w(\cdot)$ to output the targeted gaze direction $\overline{g}_i$. We then

implement the targeted projected gradient descent attack (TPGD) [60] to create perturbations $\Delta(x_i)$ for the anchor image $\hat{x}$ so that $\mathcal{G}_w(\cdot)$ outputs $\overline{g}_i$ for $\hat{x} + \Delta(x_i)$. This self-created baseline falls under the same category as the proposed PrivateGaze as the targeted attack [16, 27, 36], where the goal of PrivateGaze is to make the surrogate gaze estimator $\mathcal{G}_w(\cdot)$ output a targeted gaze annotation g_i. Finally, we take $\hat{x} + \Delta(x_i)$ as the obfuscated image and feed it to the black-box gaze estimator $\mathcal{G}_b(\cdot)$.

Gaussian Differential Privacy (GauDP): The local model of differential privacy [20] is a potential way to preserve the privacy of user when the central server is not trusted. In this baseline method, we introduce Gaussian noise to each color component $x[p, q, t]$ of every pixel in the raw image x. Specifically, we define the mechanism operating on $x[p, q, t]$ as $M(x[p, q, t]) = x[p, q, t] + \xi$, where $\xi \sim \mathcal{N}(0, Sen^2/\epsilon^2)$, with Sen representing the sensitivity of the color component's value, and ϵ denoting the privacy parameter. Since the value of each color component in the image ranges between 0 and 1, Sen is set to 1. As demonstrated in [15], M satisfies ϵ-Gaussian Differential Privacy. In our evaluation, we explore GauDP with ϵ ranging from 0.1 to 0.5.

Image Pixelization with Differential Privacy (IP-DP): IP-DP [26] is the state-of-the-art differential privacy method operating on images. Specifically, IP-DP first performs pixelization on the raw images, then adds noise sampled from a Laplace distribution to the pixelized images. The Laplace distribution has a mean of 0 and a scale of $\Delta P_b/\epsilon$, where ΔP_b is the global sensitivity of the pixelized images and ϵ is the privacy parameter. We evaluate IP-DP with different values of ϵ, including 0.3, 3.0, and 5.0.

Feature-space Differential Privacy (FS-DP): A variety of state-of-the-art differential privacy techniques [10, 56, 90, 95] add perturbations to features extracted by an encoder from raw images, rather than directly modifying the raw images themselves. We adapt these techniques for gaze estimation by training a variational auto-encoder (VAE) [46] with the surrogate gaze estimator $\mathcal{G}_w(\cdot)$. During training, the VAE takes raw images x as inputs and outputs reconstructed images $\hat{x}$, which are then processed by $\mathcal{G}_w(\cdot)$ to obtain gaze directions g'. In addition to the original VAE loss function, we introduce an extra loss term to optimize for utility, aiming to minimize the L_1 norm between g' and the gaze annotation g. In the deployment stage, the encoder of the VAE extracts d dimensional features f from x. We define the mechanism operating on f as $M(f) = f + \xi$, where ξ is sampled from a Laplace distribution $Lap(\Delta f/(d\epsilon))$. Here, Δf represents the sensitivity of the features and ϵ denotes the privacy parameter. The perturbated features are then fed into the decoder of the VAE to generate the obfuscated images. Following Xue et al. [95], we calculate the sensitivity as $\Delta f = \max_{f_i, f_j} \|f_i - f_j\|_1$, where f_i and f_j are features extracted from different raw images. We examine FS-DP with ϵ ranges from 1.0 to 3.0.

B-DAP: This method is adapted from DAPter [93], which aims to preserve user privacy in a white-box setting where the user possesses full knowledge of the deep learning model used by the service provider. The original DAPter employs a generative model-based image converter to generate obfuscated images. The image converter is trained to minimize an entropy reduction loss for privacy preservation and a task-related loss defined on the outputs of the target model for the utility objective. We adapt this method by using the surrogate gaze estimator $\mathcal{G}_w(\cdot)$ as the target model during the training stage. Specifically, the adapted method B-DAP trains the image converter by minimizing both the entropy reduction loss and the L_1 loss between g'_i and the annotation g_i. During testing, we evaluate the effectiveness of B-DAP using the black-box gaze estimator $\mathcal{G}_b(\cdot)$.

MaxP: A straightforward baseline is to minimize the similarity between the obfuscated and the raw image while ensuring the surrogate gaze estimator can still perform gaze estimation with the obfuscated image. Specifically, we train an auto-encoder that takes the raw image x_i as input and outputs the obfuscated image x'_i. During training, we feed x'_i to the surrogate gaze estimator to obtain the estimated gaze direction g'_i. The auto-encoder is trained to minimize the similarity between x_i and x'_i to achieve the privacy objective, and to minimize the L_1 loss between g'_i and the annotation g_i for the utility objective. In the testing stage, we feed the obfuscated image to the black-box gaze estimator $\mathcal{G}_b(\cdot)$.

4.3 Evaluation Setup and Metrics

In the evaluation, we designate ETHXGaze as the unknown dataset $\mathcal{D}_b$ to train the black-box gaze estimator $\mathcal{G}_b(\cdot)$, as it contains the most diverse head poses and gaze variations. Note that, $\mathcal{G}_b(\cdot)$ is trained on the raw images of $\mathcal{D}_b$. We use GazeCapture to train both the privacy preserver $\mathcal{P}(\cdot)$ and the surrogate gaze estimator $\mathcal{G}_w(\cdot)$, with 80% of the images used for training and 20% for validation. We select MPIIFaceGaze and ColumbiaGaze to evaluate the performance of $\mathcal{P}(\cdot)$. Our primary goal is to preserve the private attributes, i.e., identity and gender, of the individuals in these datasets while ensuring that gaze estimation performance remains robust when using the black-box gaze estimator.

For utility measurement, we measure the gaze estimation error, i.e., the average angular error, via the black-box gaze estimator $\mathcal{G}_b(\cdot)$. We take the obfuscated images generated by our PrivateGaze and other baselines as the inputs to assess the utility performance. We conduct this evaluation on both MPIIFaceGaze and ColumbiaGaze datasets. To evaluate the performance in privacy protection, we focus on preserving gender and identity as the two user attributes. For both the MPIIFaceGaze and ColumbiaGaze datasets, we start by randomly selecting 80% of the images that have accurate identity and gender labels. We then apply each of the seven methods under examination, i.e., PrivateGaze, GauDP, IP-DP, FS-DP, MaxP, B-DAP, and TPGD, to generate obfuscated images and form seven corresponding $\mathcal{D}_p$ sets. The $\mathcal{D}_p$ is used to train an identity recognizer and a gender recognizer. Lastly, we apply each of the seven methods on the remaining 20% of images to create corresponding testing set comprising obfuscated images. We report the recognition accuracy on this testing set for performance evaluation.

4.4 Implementation

To demonstrate the generalization of PrivateGaze, we employed various neural network architectures to construct the black-box gaze estimator $\mathcal{G}_b(\cdot)$ including ResNet18 [33], MobileNetV2 [74], ShuffleNet [102], VGG11 [78], and EfficientNet [84]. Since PrivateGaze and TPGD generate different obfuscated images for each black-box gaze estimator, the reported results are averaged across these five architectures. By contrast, the obfuscated images generated using GauDP, IP-DP, FS-DP, MaxP, and B-DAP are independent of the black-box gaze estimator used. The surrogate gaze estimator $\mathcal{G}_w(\cdot)$ is implemented using the ResNet18 architecture [33]. The classifiers used for identity and gender recognition are implemented using ResNet18. The encoder of the image generator $IG(\cdot)$ shares the same structure as $E(\cdot)$, ensuring that features extracted from the raw image x_i and the anchor image $\hat{x}$ have identical spatial dimensions. The decoder of the image generator consists of four up-convolutional blocks, each containing an upsampling of feature maps followed by three convolutional layers.

We develop PrivateGaze using the PyTorch framework and use the Adam optimizer [45]. The standard learning rate is set to 0.001, unless specified otherwise. The privacy preserver is trained over 12,000 steps with a mini-batch size of 25 for all evaluation scenarios. The classifiers for identity and gender recognition are trained for 20 and 5 epochs, respectively. The surrogate gaze estimators, trained on the GazeCapture dataset, undergo 5 epochs of training. The black-box gaze estimators, using different structures, are trained for 25 epochs. The learning rates for training both the surrogate and black-box gaze estimators are set to 0.0001. We fix the value of λ in Equation 4 at 75 and conduct ablation studies to assess the impact of λ on the performance of PrivateGaze.

4.5 Performance in the Privacy Goal

The evaluation results for the effectiveness of various privacy-preserving methods are summarized in Table 1. For PrivateGaze, it is important to note that different anchor images are generated when the black-box estimator follows different network architectures. This variation occurs because the image generation module queries $\mathcal{G}_b(\cdot)$ to generate the anchor image $\hat{x}$. We report the averaged identity recognition accuracy and gender recognition accuracy for PrivateGaze over these different structures of $\mathcal{G}_b(\cdot)$.

Table 1. Identity and gender recognition accuracies (%) on obfuscated images generated by PrivateGaze and other baseline methods evaluated on (a) MPIIFaceGaze and (b) ColumbiaGaze datasets. We report the results of GauDP, IP-DP, and FS-DP with different ϵ values. The proposed PrivateGaze can effectively preserve the private attributes against the attackers who train their classifiers on obfuscated images annotated with correct identity and gender labels. Lower recognition accuracy indicates better performance in privacy protection. "W/o Defense" denotes the scenario where unprotected original images are used for gaze estimation services, with attribute classifiers trained and tested on the original images.

Attributes	w/o Defense	GauDP			IP-DP			FS-DP			MaxP	BDAP	TPGD	PrivateGaze
		0.1	0.3	0.5	0.3	3.0	5.0	1.0	2.0	3.0				
Identity	99.7	38.2	84.2	93.1	17.9	94.3	96.2	29.6	75.6	90.3	99.3	98.9	6.50	6.31
Gender	99.8	73.2	87.0	95.7	65.9	96.2	98.6	63.0	78.4	88.5	99.6	96.3	62.0	62.0

(a) MPIIFaceGaze

Attributes	w/o Defense	GauDP			IP-DP			FS-DP			MaxP	BDAP	TPGD	PrivateGaze
		0.1	0.3	0.5	0.3	3.0	5.0	1.0	2.0	3.0				
Identity	99.9	4.85	95.1	99.6	6.20	86.4	97.4	5.86	38.9	73.8	99.7	91.1	2.82	1.13
Gender	99.9	57.5	87.2	96.9	57.0	96.9	99.2	57.6	67.7	82.2	99.5	95.6	56.9	56.9

(b) ColumbiaGaze

From Table 1, it is evident that the average identity and gender recognition accuracies on images obfuscated by PrivateGaze are notably low, with values of 6.3% and 1.1% respectively, and minimal standard deviations of 0.2 and 0.1 across different structures of $\mathcal{G}_b(\cdot)$. Additionally, the average gender recognition accuracies are 62.0% for the MPIIFaceGaze dataset and 53.9% for the ColumbiaGaze dataset, both with a standard deviation of 0.0 across different structures of $\mathcal{G}_b(\cdot)$. The results underscore the effectiveness of PrivateGaze in significantly reducing the recognizability of identity and gender attributes in the obfuscated images. The attacker cannot train an effective identity recognizer or gender recognizer on the obfuscated images generated by PrivateGaze, even when they have access to correct identity and gender labels of the obfuscated images. The minimal standard deviations indicate that PrivateGaze achieves consistent privacy performance across various architectures of $\mathcal{G}_b(\cdot)$, demonstrating its robustness and generalizability.

For GauDP, IP-DP, and FS-DP, we observe that the identity and gender recognition accuracies decrease as the values of ϵ drop, indicating that these methods can effectively preserve user privacy with smaller ϵ values. For example, the identity recognition accuracy for GauDP with $\epsilon = 0.1$ is 4.85% on the ColumbiaGaze dataset, while for FS-DP is 5.86% with $\epsilon = 1.0$. However, as we will demonstrate in Section 4.6, the utility performance of GauDP, IP-DP, and FS-DP with small ϵ values is significantly worse compared to PrivateGaze.

For obfuscated images obtained by MaxP and B-DAP, the identity recognition accuracy exceeds 90%, and the gender recognition accuracy is higher than 95% on both datasets. This outcome implies that the obfuscated images generated by MaxP and B-DAP still retain discernible features, which could potentially be exploited by attackers to infer the private attributes of the users. Lastly, TPGD achieves results similar to PrivateGaze on both testing datasets, as it is designed based on our framework.

4.5.1 In-depth analysis and discussion. As shown in Table 1, the gender recognizers for PrivateGaze and TPGD have the same recognition accuracy, i.e., 62.0% and 56.9% on MPIIFaceGaze and ColumbiaGaze datasets, respectively. We observe that when obfuscated images contain no discernible gender cues, the trained recognizers tend to output the same gender label for any inputs. Essentially, without gender-related information in the obfuscated images, the best knowledge the recognizers can obtain is an approximation of the gender distribution in the

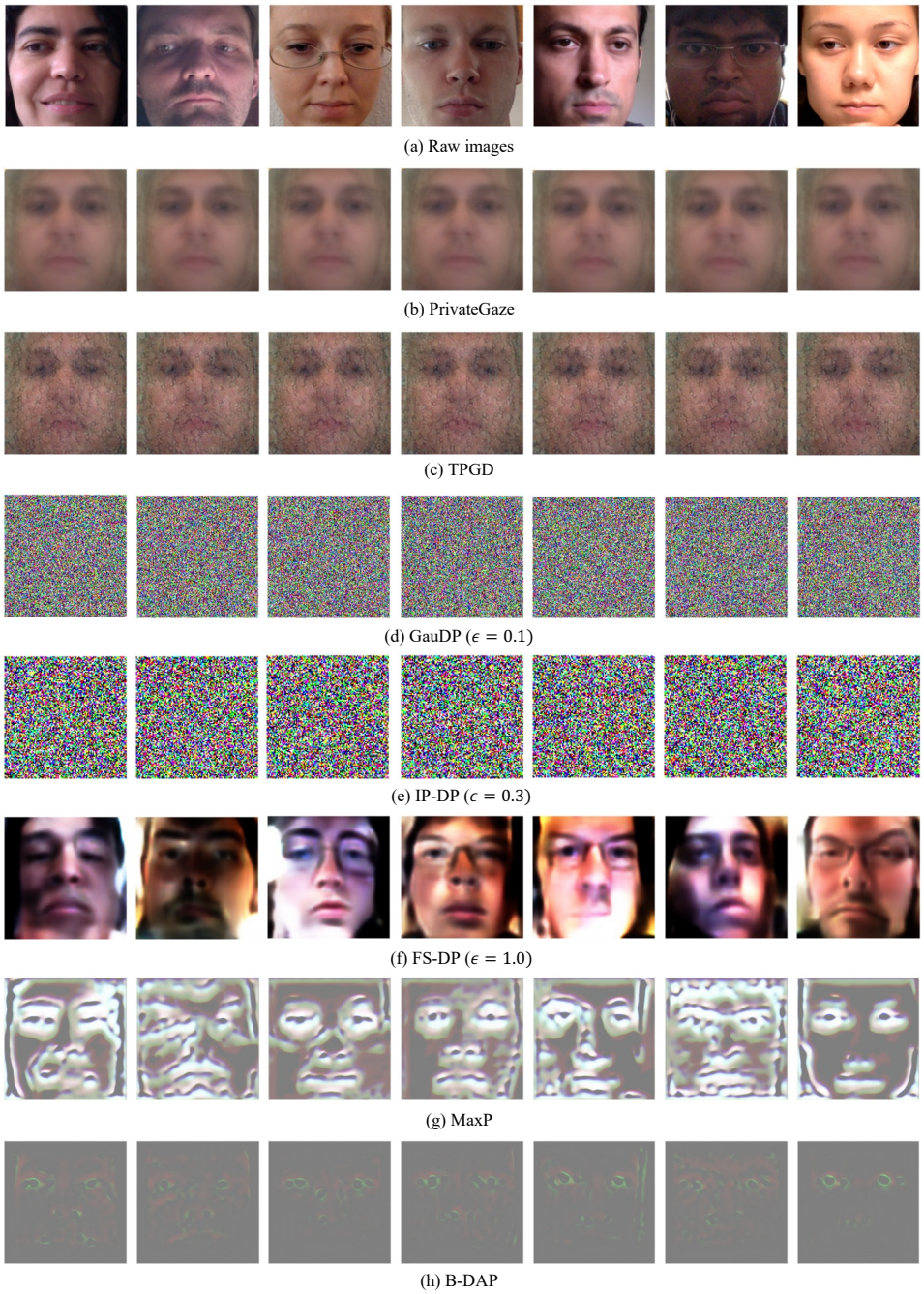


Fig. 5. Illustration of (a) raw images of different subjects and obfuscated images generated by (b) PrivateGaze, (c) TPGD, (d) GauDP ($\epsilon = 0.1$), (e) IP-DP ($\epsilon = 0.3$), (f) FS-DP ($\epsilon = 1.0$), (g) MaxP, and (h) B-DAP. The obfuscated images obtained by PrivateGaze and TPGD have similar appearances, making it challenging for attackers to infer user identity and gender from the obfuscated images.

training dataset $\mathcal{D}_p$. Consequently, during inference, leveraging prior probability distributions, the recognizers predict the gender of a given testing image to be the one with the highest prior probability, thereby achieving a better classification accuracy than random guessing.

For instance, in the MPIIFaceGaze dataset, the gender recognizers for both PrivateGaze and TPGD consistently classify subjects in all testing images as male. This outcome results from the imbalanced gender distribution within the training dataset $\mathcal{D}_p$, where 59.5% of images feature male subjects. Consequently, given that 62.0% of the testing images contain male subjects, this results in an equal gender recognition accuracy of 62.0%. Thus, these evaluation results demonstrate the effectiveness of PrivateGaze in preserving gender information.

In contrast to the imbalanced gender distribution, the identity distribution within the training set $\mathcal{D}_p$ for both MPIIFaceGaze and ColumbiaGaze datasets is well-balanced, with each subject having an equal number of training images. Thus, when the identity recognizers cannot learn any identity-related information from the obfuscated images and are faced with equal prior probabilities, the identity recognition accuracy is similar to that of random guess, i.e., 6.31% and 1.13% for MPIIFaceGaze and ColumbiaGaze datasets, respectively. This result further demonstrates the capability of PrivateGaze in preserving user information from potential malicious service providers.

4.5.2 Visualization of obfuscated images. We also perform a visual comparison between the raw full-face images and the obfuscated images generated by the proposed PrivateGaze and other baseline methods. As shown in Figures 5 (b) and (c), the obfuscated images generated by PrivateGaze and TPGD exhibit similar visual characteristics among themselves, making it extremely challenging for an attacker to infer the user private attributes, even when correctly labeled obfuscated images are used to train the deep learning-based classifiers (as showcased in Table 1). In contrast, as shown in Figure 5 (g), the user attributes are much more discernible in the obfuscated images produced by MaxP. Moreover, despite it might be difficult for human observers to identify the subjects in images obfuscated by B-DAP, the results shown in Table 1 indicate that an attacker can successfully train a classifier on these obfuscated images to accurately classify user private attributes.

As shown in Figures 5 (d) and (e), GauDP ($\epsilon = 0.1$) and IP-DP ($\epsilon = 0.3$) apply significant perturbations to the raw images, making it difficult for the malicious service provider to classify identities and genders from them. Furthermore, FS-DP ($\epsilon = 1.0$) generates obfuscated images that have significantly different appearances from raw images by perturbing features with strong random noises, effectively preserving user privacy.

4.5.3 Residual map. To better understand why the obfuscated images generated by PrivateGaze have similar appearances yet lead the black-box gaze estimators to output varied gaze directions, we examine the residual maps between the obfuscated image and the anchor image. As shown in Figure 6, the privacy preserver adds perturbations to the eye regions of the anchor image, which are the most critical parts of the facial image for gaze estimation [101]. These perturbations are learned by the privacy preserver and contain the gaze-related features of the raw images. In this way, the obfuscated images retain similar appearances while leading to different gaze directions when fed into the black-box gaze estimator.

4.6 Performance in the Utility Goal

Below, we evaluate the performance of different methods in achieving the utility goal, i.e., gaze estimation task. The results are reported in Table 2, which demonstrate that PrivateGaze consistently outperforms all compared methods and achieves the lowest average angular error across two datasets and with five different neural network architectures for the black-box gaze estimator. On average, PrivateGaze improves gaze estimation performance by 41.79%-78.47%, and 34.13%-64.68% on MPIIFaceGaze and ColumbiaGaze, respectively.

For differential privacy-based methods, i.e., GauDP, IP-DP, and FS-DP, while they achieve strong privacy performance by setting ϵ to small values, their utility performance notably lags behind PrivateGaze. For instance,

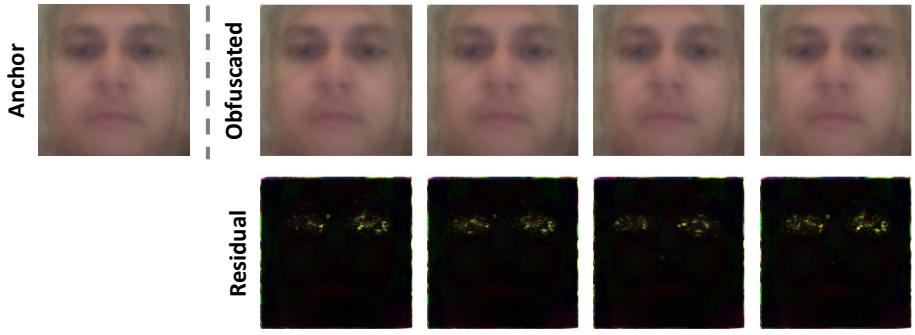


Fig. 6. Illustration of the residual maps (scaled by a factor of 30) generated from the obfuscated images (with different gaze directions) compared to the anchor image. The privacy preserver introduces perturbations into the eye regions of the anchor image when generating obfuscated images, resulting in different gaze directions.

Table 2. Performance in the utility goal is measured by the average angular error (in degree) of the black-box gaze estimator across five different structures, using obfuscated images generated by different methods as inputs for gaze estimation. We present results for GauDP, IP-DP, and FS-DP with different ϵ values, respectively. PrivateGaze consistently outperforms all compared methods in all examined scenarios. Overall, PrivateGaze achieves an average performance improvement in gaze estimation of 49.86%-77.10%, and 34.13%-60.67% on MPIIFaceGaze and ColumbiaGaze, respectively.

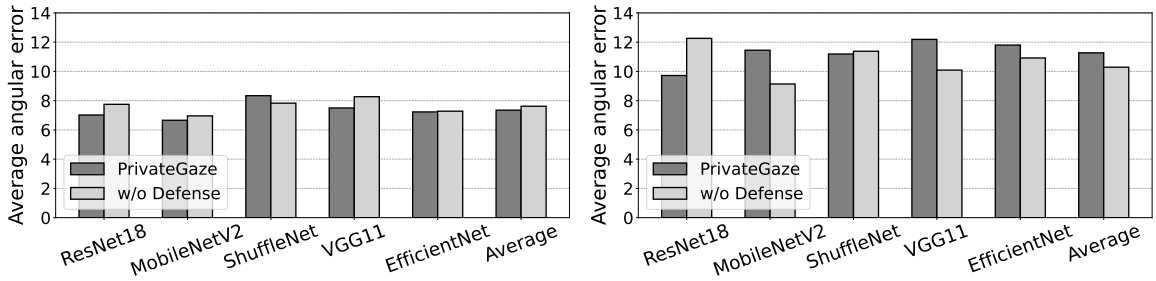
Sructures	GauDP			IP-DP			FS-DP			MaxP	B-DAP	TPGD	PrivateGaze
	0.1	0.3	0.5	0.3	3.0	5.0	1.0	2.0	3.0				
ResNet18	20.73	20.37	20.09	20.94	26.22	28.90	16.09	15.63	14.90	13.65	20.07	17.28	**7.02**
MobileNetV2	25.70	25.72	25.69	30.03	30.29	30.15	19.86	18.16	16.46	29.95	46.52	17.82	**6.66**
ShuffleNet	13.93	13.62	13.33	13.77	15.22	15.50	19.68	17.95	16.16	18.93	23.38	11.71	**8.34**
VGG11	14.06	14.82	15.77	14.32	15.76	16.64	20.99	19.63	18.23	17.72	13.43	15.89	**7.50**
EfficientNet	86.02	44.52	26.51	91.68	69.92	40.25	19.36	17.36	15.58	13.32	13.08	9.99	**7.23**
Average	32.09	23.81	20.28	34.14	31.48	26.28	19.19	17.74	16.26	18.71	23.29	14.66	**7.35**
Improvement	77.10%	69.13%	63.75%	78.47%	76.65%	72.03%	41.79%	58.56%	54.79%	60.72%	68.44%	49.86%	

(a) MPIIFaceGaze

Sructures	GauDP			IP-DP			FS-DP			MaxP	B-DAP	TPGD	PrivateGaze
	0.1	0.3	0.5	0.3	3.0	5.0	1.0	2.0	3.0				
ResNet18	14.46	14.89	15.65	19.42	26.75	30.80	19.35	18.90	18.80	15.42	15.22	21.45	**9.72**
MobileNetV2	17.14	17.27	17.33	20.55	21.02	21.03	19.30	18.96	18.71	22.69	41.32	23.17	**11.45**
ShuffleNet	13.58	13.48	13.31	12.86	12.55	12.36	19.94	20.44	20.26	20.96	24.87	19.35	**11.19**
VGG11	14.60	13.75	13.39	18.42	15.78	15.11	24.62	23.54	22.99	27.24	18.50	20.81	**12.19**
EfficientNet	83.52	44.61	25.86	86.91	63.39	36.51	21.95	21.39	20.68	15.72	12.64	13.67	**11.80**
Average	28.66	20.8	17.11	31.63	27.89	23.16	21.03	20.64	20.29	20.40	22.50	19.69	**11.17**
Improvement	60.67%	43.75%	34.13%	64.68%	59.94%	51.77%	46.88%	45.88%	44.94%	44.75%	49.91%	42.76%	

(b) ColumbiaGaze

GauDP with $\epsilon = 0.1$ results in an average angular error over various structures of $\mathcal{G}_b(\cdot)$ of 32.09° on MPIIFaceGaze, nearly five times higher than that of PrivateGaze. Increasing the values of ϵ improves the utility performance of differential privacy-based methods by reducing the intensity of the added perturbation for privacy protection. However, as shown in Table 1, this improvement comes at the cost of deteriorating privacy performance.



(a) MPIIFaceGaze (b) ColumbiaGaze

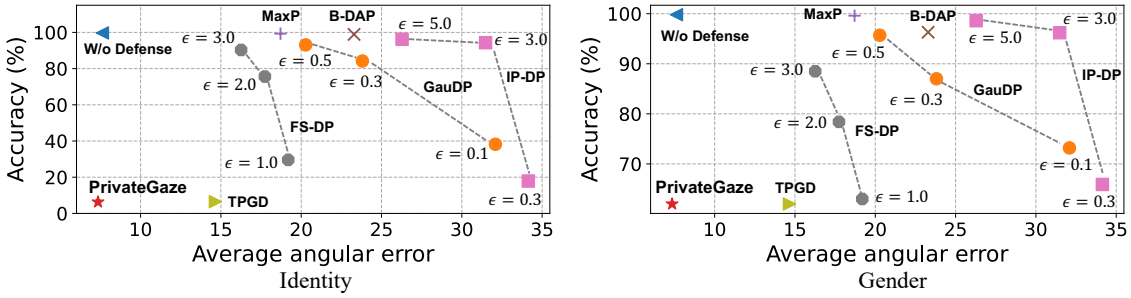
Fig. 7. Average angular error (in degree) of the black-box gaze estimator when using obfuscated images generated by PrivateGaze compared to raw images, i.e., w/o Defense, as inputs for gaze estimation. Our method demonstrates performance comparable to that of raw images across different structures and datasets.

On MPIIFaceGaze dataset, the average utility performance for MaxP and B-DAP is 18.71° and 23.29° respectively. These results indicate that the autoencoder and image converter trained by MaxP and B-DAP, respectively, with the surrogate gaze estimator cannot generalize well to black-box gaze estimators. Despite TPGD achieving similar performance in privacy protection as PrivateGaze (as shown previously in Table 1), its average utility performance over different black-box gaze estimators is 14.66°, which is significantly worse than that for PrivateGaze, at 7.35°. Similar observations hold for the evaluation results on the ColumbiaGaze dataset.

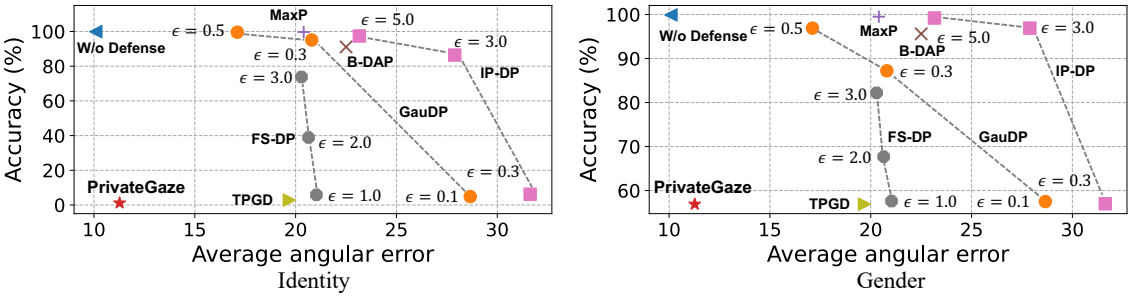
To further investigate the performance of PrivateGaze on the utility goal, we compare the average angular error of the black-box gaze estimator when using obfuscated images generated by PrivateGaze versus raw images (*w/o Defense*) as the inputs. The results are shown in Figure 7. On the MPIIFaceGaze dataset, the average performance of PrivateGaze among different structures is 7.35°, which is better than that of *w/o Defense* at 7.62°. On the ColumbiaGaze dataset, PrivateGaze achieves an average utility performance of 11.27° across different structures, which is slightly higher than *w/o Defense* at 10.29°. Overall, these results indicate that PrivateGaze maintains comparable utility performance for the black-box gaze estimator even when compared to a method that does not employ any privacy protection.

4.6.1 Discussion on utility performance. As shown in Figure 7, on the MPIIFaceGaze dataset, PrivateGaze achieves superior gaze estimation performance compared to *w/o Defense* on average. To explore the reason behind this improvement, we observed a reduction in the average angular error of $\mathcal{G}_w(\cdot)$ during training on obfuscated images generated by $\mathcal{P}(\cdot)$ from the validation set. We observe a decrease from 12° to 3.52°, which is lower than the error observed on raw images, at 5.31°. This suggests that the privacy preserver acts as an *image filter* that eliminates redundant features from raw images, thereby enhancing gaze estimation performance.

Moreover, we observe that the average gaze estimation performance of PrivateGaze is better than *w/o Defense* on MPIIFaceGaze, while it shows marginally inferior performance on ColumbiaGaze. This difference can be attributed to the training data used for the privacy preserver, which was trained on the GazeCapture dataset and evaluated on both MPIIFaceGaze and ColumbiaGaze. As shown in Figure 4, both GazeCapture and MPIIFaceGaze are acquired under real-world conditions employing front-facing cameras in mobile devices, whereas ColumbiaGaze is collected in a more controlled laboratory environment using web cameras. The images from GazeCapture are more similar to those from MPIIFaceGaze than to those from ColumbiaGaze. Consequently, the image filter, i.e., privacy preserver, operates more effectively on MPIIFaceGaze compared to ColumbiaGaze, and leads to better gaze estimation performance on MPIIFaceGaze.



(a) MPIIFaceGaze



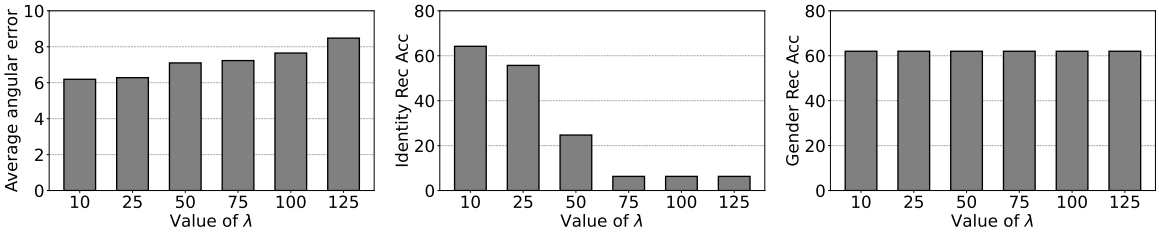
(b) ColumbiaGaze

Fig. 8. Overall performance comparison between PrivateGaze and the compared methods on (a) MPIIFaceGaze and (b) ColumbiaGaze datasets. The X-axis is the utility performance, i.e., the average angular error, while the Y-axis is the privacy performance, i.e., the identity recognition accuracy and gender recognition accuracy. The plots are the identity (gender) recognition accuracy and the average angular error for PrivateGaze and the compared methods. The overall performance of PrivateGaze lies on the lower left corner in all the evaluation scenarios, which indicates the superiority of PrivateGaze on the privacy-utility trade-off.

4.6.2 In-depth analysis on the generalization ability of PrivateGaze. PrivateGaze demonstrates good generalization ability, i.e., $\mathcal{P}(\cdot)$ is trained with $\mathcal{G}_w(\cdot)$ but yields good gaze estimation performance on $\mathcal{G}_b(\cdot)$. We attribute this capability primarily to the proposed anchor image generation module and privacy protection mechanism. Specifically, the privacy protection mechanism ensures that the obfuscated images x' closely resemble the anchor image $\hat{x}$, resulting in similar gaze estimation results for both $\mathcal{G}_w(\cdot)$ and $\mathcal{G}_b(\cdot)$. This alignment encourages the obfuscated images to produce similar outputs for both $\mathcal{G}_w(\cdot)$ and $\mathcal{G}_b(\cdot)$. Consequently, while PrivateGaze optimizes the x' to facilitate accurate gaze direction inference by $\mathcal{G}_w(\cdot)$, it also achieves good gaze estimation performance on $\mathcal{G}_b(\cdot)$. Our experiments validate this analysis. In contrast, TPGD utilizes the anchor image to generate obfuscated images without considering their similarity to the anchor image, leading to substantially inferior performance compared to PrivateGaze in terms of utility. Moreover, as detailed in Section 4.8.2, neglecting the outputs of $\mathcal{G}_w(\cdot)$ and $\mathcal{G}_b(\cdot)$ during the generation of $\hat{x}$ results in a significant performance decline for PrivateGaze in terms of utility.

4.7 Overall Performance Comparison

Below, we compare the overall performance, i.e., the privacy-utility trade-off, between PrivateGaze and the compared methods. The results are shown in Figure 8. Across all evaluation scenarios, PrivateGaze consistently



(a) Utility (b) Preserving users identities (c) Preserving users genders

Fig. 9. Impact of λ on the performance trade-off between utility and privacy. (a) utility, (b) preserving users' identities, and (c) preserving users' genders. The value of λ trades off the performance of PrivateGaze on utility against privacy. PrivateGaze sets $\lambda = 75$, which allows PrivateGaze to consistently preserve user privacy while maintaining comparable utility performance to raw images across different datasets.

Table 3. Impact of the proposed anchor image generation algorithm on the performance of PrivateGaze. Our method of generating anchor images is essential for maintaining the utility performance of PrivateGaze.

Task	PrivateGaze	RandomAnchor
Utility: Gaze estimation	7.23°	9.06°
Privacy: Identity recognition accuracy	6.31%	6.40%
Privacy: Gender recognition accuracy	62.0%	62.0%

occupies the lower left corner, indicating its effectiveness in preserving user privacy while maintaining good utility performance. Notably, TPGD, GauDP ($\epsilon = 0.1$), FS-DP ($\epsilon = 1.0$), and IP-DP ($\epsilon = 0.3$) demonstrate privacy-preserving performance comparable to PrivateGaze, yet their utility performances significantly lag behind. Moreover, among the differential privacy-based methods, i.e., GauDP, IP-DP, and FS-DP, FS-DP achieves the most favorable privacy-utility trade-off. This is because FS-DP adopts state-of-the-art DP techniques and involves the gaze estimator in the training stage. Lastly, MaxP and B-DAP exhibit overall performances within the upper middle region, indicating their inability to achieve both utility and privacy goals simultaneously.

4.8 Ablation Studies

We conduct ablation studies to investigate the impact of different design choices on the system performance. We use MPIIFaceGaze as the testing set and implement a black-box gaze estimator using EfficientNet.

4.8.1 Impact of λ on the performance trade-off between utility and privacy. For the optimization problem described in Equation 4, the parameter λ trades off the utility objective and the privacy-preserving objective. To explore how changes in λ affect the performance of PrivateGaze, we experiment with varying λ within the range from 10 to 125 and present results in Figure 9.

Overall, reducing the value of λ improves the utility performance of PrivateGaze yet degrades its privacy performance when λ is set below 75. Particularly, there is a decrease in the average angular error from 8.48° to 6.10° when λ is reduced from 125 to 10. This indicates an increased weighting of the utility objective in the optimization problem. For privacy objective, the identity recognition accuracy is higher than 20% when the value of λ is smaller than 75, while the gender recognition accuracy is stable for all the examined values of λ. In conclusion, we have chosen to set λ to 75 in our implementation. This setting effectively balances the need to preserve user privacy in the obfuscated images while maintaining comparable utility performance to raw images across various datasets.

(a) Female (b) Male

Fig. 10. Two anchor images that are generated using images containing (a) female-only subjects and (b) male-only subjects, respectively.

4.8.2 Impact of anchor image generation. We investigate the impact of the proposed anchor image generation module (described in Section 3.2.1 and Algorithm 1) on the performance of PrivateGaze. Specifically, instead of querying the black-box gaze estimator and the surrogate gaze estimator to find a suitable set of images, we randomly sample 50 images from the training set to form an average facial image as the anchor image. We use the term *RandomAnchor* to denote the method of generating anchor images through randomly sampled images. The results are shown in Table 3.

First, the anchor image does not affect the performance of PrivateGaze on the privacy objective, as the identity recognition accuracy and the gender recognition accuracy of RandomAnchor and PrivateGaze are similar. Second, the average angular error of RandomAnchor is 20% higher than that of PrivateGaze. This indicates that the proposed method for generating the anchor image can improve the generalizability of the privacy preserver. In other words, when trained with the surrogate gaze estimator, the privacy preserver can still achieve good gaze estimation performance when the obfuscated images are used by the black-box gaze estimator. Moreover, we observe that the utility performance of RandomAnchor is superior to that of the other baseline methods shown in Table 2. This highlights the effectiveness of our method in achieving the utility objective.

4.8.3 Impact of gender of the anchor image on preserving gender information. In general, the anchor image is constructed by averaging full-face images from both male and female subjects, ensuring it does not portray a specific gender. When using the anchor image as the base in generating the obfuscated images, as illustrated in Figure 6, the resulting obfuscated images will resemble the anchor image and also avoid portraying a specific gender. Nevertheless, there is a possibility that the images used to generate the anchor image predominantly belong to subjects of a specific gender, especially in cases where the gender distribution within the training dataset is imbalanced. In such instances, the anchor image may inadvertently exhibit characteristics of that specific gender.

To study how the gender of the anchor image affects the performance of PrivateGaze in preserving gender information, we generate an anchor image $\hat{x}_f$ using only images containing female subjects. The resulting anchor image is shown in Figure 10 (a). In this case, we consider the genders of $\hat{x}_f$ and the corresponding obfuscated images as "female". We find that the gender classification accuracy on the obfuscated images remains at 62.0%. As discussed in Section 4.5.1, this indicates the effectiveness of PrivateGaze in preserving user's gender information. Similarly, we conduct experiments using only male images to generate the anchor image (illustrated in Figure 10 (b)), and we obtain the same result. These results demonstrate that the efficacy of PrivateGaze in preserving users' gender information is not compromised even when the anchor image exhibits a specific gender.

4.9 System Performance on Different Computation Platforms

We measure the processing time and memory usage of the privacy preserver when implemented with different architectures and deployed on different computation platforms. In addition to the default structure, which consists of four convolutional and up-convolutional blocks, we also evaluate a variant with such blocks. We assess the

Table 4. The processing time (in ms) on different hardware platforms. PrivateGaze does not introduce too much processing latency.

Platforms	3 Blocks	4 Blocks
Desktop (RTX 3080Ti)	2.7	3.8
Laptop (RTX 3060)	9.6	10.9
Laptop (RTX 1050Ti)	46.7	53.5

Table 5. The memory usage (in MB) on different hardware platforms. PrivateGaze consumes similar memories on different hardware platforms.

Platforms	3 Blocks	4 Blocks
Desktop (RTX 3080Ti)	2193	2267
Laptop (RTX 3060)	1964	2043
Laptop (RTX 1050Ti)	1902	2003

Table 6. The utility and privacy performance of PrivateGaze in different structures.

Task	3 Blocks	4 Blocks	w/o Defense
Utility: Gaze estimation	8.12°	7.23°	7.28°
Privacy: Identity recognition accuracy	6.36%	6.31%	99.8%
Privacy: Gender recognition accuracy	62.0%	62.0%	99.4 %

performance of PrivateGaze on three hardware platforms, including a desktop equipped with an NVIDIA GeForce RTX 3080Ti GPU, a laptop featuring an NVIDIA GeForce RTX 3060 GPU, and a laptop equipped with an NVIDIA GeForce RTX 1050Ti GPU. These platforms are chosen to represent a wide range of common computational devices used in daily scenarios.

Processing time. We measure the latency introduced by the privacy preserver in generating obfuscated images. We randomly sample one image from the MPIIFaceGaze dataset and feed it into the privacy preserver. We repeat the experiment 1000 times, and report the average processing time on different hardware platforms in Table 4.

Specifically, for the privacy preserver consisting of four convolutional and up-convolutional blocks, the average processing time on the desktop with an NVIDIA GeForce RTX 3080Ti GPU is less than 4 ms. The processing time on the laptops with an NVIDIA GeForce RTX 3060 GPU and with an NVIDIA GeForce RTX 1050Ti GPU is 10.9 ms and 53.5 ms, respectively. When reducing the number of the convolutional and up-convolutional blocks to three, the average processing time decreases. On the desktop, it is reduced to 2.7 ms, on the laptop with an NVIDIA GeForce RTX 3060 GPU it is 9.6 ms, and the laptop with an NVIDIA GeForce RTX 1050Ti GPU it is 46.7 ms. These results indicate that the deployed privacy preserver introduces minimal processing latency.

Memory usage. To measure memory usage, we follow the method described in [97] by reporting the memory allocated specifically to the privacy preserver. This is determined by subtracting the memory usage before loading the privacy preserver from the run-time memory usage. The results are shown in Table 5, indicating similar memory usage across different scenarios, approximately 2,000 MB for the privacy preserver.

Utility and privacy performance. We report the utility and privacy performance of the privacy preserver with different structures in Table 6. When reducing the number of convolutional and up-convolutional blocks, the utility performance of PrivateGaze is decreased to 8.12°, which is approximately 1° higher than w/o defense, while the privacy performance maintains stable. Therefore, although using three convolutional and up-convolutional blocks can slightly reduce the processing time, we opt to design the privacy preserver with four convolutional and up-convolutional blocks to ensure comparable utility performance to w/o defense.

4.10 Discussions

Below, we discuss the key findings of this paper and the impacts of PrivateGaze. We also discuss its limitations and propose future research directions for enhancing user privacy in black-box mobile services.

Key findings. This work presents three major findings. First, we demonstrate effective user privacy preservation by transforming different raw images into obfuscated images that have similar appearances with a pre-generated

anchor image. Second, leveraging the anchor image allows us to control the appearance of obfuscated images, thereby achieving our utility goal. Specifically, since the anchor image, i.e., the average full-face image, produces consistent outputs for both $\mathcal{G}_w(\cdot)$ and $\mathcal{G}_b(\cdot)$, obfuscated images that closely resemble the anchor image also yield consistent results for gaze estimators $\mathcal{G}_w(\cdot)$ and $\mathcal{G}_b(\cdot)$. This alignment enables $\mathcal{P}(\cdot)$ trained with $\mathcal{G}_w(\cdot)$ to perform accurate gaze estimation on $\mathcal{G}_b(\cdot)$. Lastly, our well-designed $\mathcal{P}(\cdot)$ structure and training objective allow us to manipulate the behaviours of gaze estimators through imperceptible modification applied to the anchor image. This finding underscores vulnerabilities in deep learning-based gaze estimation systems.

Impacts of PrivateGaze. Compared to existing works [58, 71, 93], PrivateGaze addresses a more practical scenario where the deep learning-based model used by the service provider remains a black box to users. PrivateGaze introduces a novel framework designed to preserve user privacy while maintaining good gaze estimation performance on such black-box models. While our current evaluation focuses on preserving identity and gender as private attributes, the framework's flexibility allows for the preservation of other private attributes, such as ages, emotions, and details of the user's surroundings. Moreover, PrivateGaze can be extended to preserve user privacy in various applications, including head pose estimation [75] and emotion recognition [92], by adapting the utility goals accordingly.

Limitations. Our experiments have demonstrated that PrivateGaze outperforms DP-based methods in achieving both privacy and utility goals. However, it is important to note that unlike DP-based methods, the current design of PrivateGaze does not provide a theoretical privacy guarantee.

Future research directions. A promising avenue for future research involves extending PrivateGaze to other applications, such as hand pose estimation [75] Another intriguing direction is to develop privacy-preserving solutions tailored for wearable-based gaze estimation systems that either utilize near-eye pupil images[40, 44] or event streams [2, 3, 98, 104] as tracking inputs. These systems pose unique challenges due to the sensitivity of the data captured and the wide adoption of eye tracking in head-mounted platforms such as augmented/virtual reality devices [9, 72]. Designing effective privacy-preserving solutions for such systems could significantly enhance user trust and adoption in these technologies.

5 Conclusion

In this work, we present PrivateGaze, the first approach that can effectively preserve users' privacy information when calling black-box gaze tracking services without compromising the estimation performance. PrivateGaze trains a user-side privacy preserver to convert privacy-sensitive full-face images into privacy-enhanced obfuscated versions. The obfuscated images do not contain any information about the users' private attributes yet can be directly used by the black-box gaze estimator to obtain accurate gaze directions. Our comprehensive experiments on four benchmark datasets show that PrivateGaze can effectively protect users' private attributes, e.g., identity and gender, even when the attribute recognizers are trained on obfuscated images with accurate attribute labels. Meanwhile, the obfuscated images generated by PrivateGaze can achieve comparable gaze tracking performance to conventional, unprotected full-face images.

Acknowledgments

We would like to express our gratitude to the anonymous reviewers and the associate editors for their insightful comments and guidance. We also appreciate Koen G. Langendoen for his valuable comments and suggestions during our discussions. This work was supported in part by the Meta Research Award and by SURF Research Cloud grants EINF-2391, EINF-8964, and EINF-9272. The contents of this paper do not necessarily reflect the positions or policies of the funding agencies.

References

[1] Yomna Abdelrahman, Anam Ahmad Khan, Joshua Newn, Eduardo Velloso, Sherine Ashraf Safwat, James Bailey, Andreas Bulling, Frank Vetere, and Albrecht Schmidt. 2019. Classifying attention types with thermal imaging and eye tracking. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 3 (2019), 1–27.

[2] Anastasios N Angelopoulos, Julien NP Martel, Amit P Kohli, Jörg Conradt, and Gordon Wetzstein. 2021. Event-Based Near-Eye Gaze Tracking Beyond 10,000 Hz. *IEEE Transactions on Visualization and Computer Graphics (TVCG)* 27, 5 (2021), 2577–2586.

[3] Pietro Bonazzi, Sizhen Bian, Giovanni Lippolis, Yawei Li, Sadique Sheik, and Michele Magno. 2024. Retina: Low-Power Eye Tracking with Event Camera and Spiking Hardware. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 5684–5692.

[4] Efe Bozkir, Ali Burak Ünal, Mete Akgün, Enkelejda Kasneci, and Nico Pfeifer. 2020. Privacy Preserving Gaze Estimation Using Synthetic Images via a Randomized Encoding Based Framework. In *Proceedings of ACM Symposium on Eye Tracking Research and Applications (ETRA)*. 1–5.

[5] Andreas Bulling and Hans Gellersen. 2010. Toward mobile eye-based human-computer interaction. *IEEE Pervasive Computing* 9, 4 (2010), 8–12.

[6] Renwang Chen, Xuanhong Chen, Bingbing Ni, and Yanhao Ge. 2020. SimSwap: An Efficient Framework For High Fidelity Face Swapping. In *Proceedings of the 28th ACM International Conference on Multimedia (MM)*. 2003–2011.

[7] Xiuge Chen, Namrata Srivastava, Rajiv Jain, Jennifer Healey, and Tilman Dingler. 2023. Characteristics of Deep and Skim Reading on Smartphones vs. Desktop: A Comparative Study. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems (CHI)*. 1–14.

[8] Eunji Chong, Katha Chanda, Zhefan Ye, Audrey Southerland, Nataniel Ruiz, Rebecca M Jones, Agata Rozga, and James M Rehg. 2017. Detecting gaze towards eyes in natural social interactions and its use in child assessment. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 1–20.

[9] Viviane Clay, Peter König, and Sabine Koenig. 2019. Eye tracking in virtual reality. *Journal of Eye Movement Research* 12, 1 (2019).

[10] William L Croft, Jörg-Rüdiger Sack, and Wei Shi. 2022. Differentially private facial obfuscation via generative adversarial networks. *Future Generation Computer Systems* 129 (2022), 358–379.

[11] Kaiwen Cui, Rongliang Wu, Fangneng Zhan, and Shijian Lu. 2023. Face Transformer: Towards High Fidelity and Accurate Face Swapping. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. 668–677.

[12] Edwin S Dalmaijer, Sebastiaan Mathôt, and Stefan Van der Stigchel. 2014. PyGaze: An Open-source, Cross-platform Toolbox for Minimal-effort Programming of Eyetracking Experiments. *Behavior Research Methods* 46 (2014), 913–921.

[13] Brendan David-John, Kevin Butler, and Eakta Jain. 2022. For your eyes only: Privacy-preserving eye-tracking datasets. In *Proceedings of ACM Symposium on Eye Tracking Research and Applications (ETRA)*. 1–6.

[14] Brendan David-John, Diane Hosfelt, Kevin Butler, and Eakta Jain. 2021. A privacy-preserving approach to streaming eye-tracking data. *IEEE Transactions on Visualization and Computer Graphics (TVCG)* 27, 5 (2021), 2555–2565.

[15] Jinshuo Dong, Aaron Roth, and Weijie J Su. 2022. Gaussian differential privacy. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 84, 1 (2022), 3–37.

[16] Yinpeng Dong, Fangzhou Liao, Tianyu Pang, Hang Su, Jun Zhu, Xiaolin Hu, and Jianguo Li. 2018. Boosting Adversarial Attacks With Momentum. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 9185–9193.

[17] Laura Van Doore. 2020. WeGaze. https://lauravandoore.com/portfolio-item/wegaze/, accessed: 2023-11-09.

[18] Huiyu Duan, Guangtao Zhai, Xiongkuo Min, Zhaohui Che, Yi Fang, Xiaokang Yang, Jesús Gutiérrez, and Patrick Le Callet. 2019. A Dataset of Eye Movements for the Children with Autism Spectrum Disorder. In *Proceedings of the 10th ACM Multimedia Systems Conference (MMSys)*. 255–260.

[19] Andrew T Duchowski, Krzysztof Krejtz, Nina A Gehrer, Tanya Bafna, and Per Bækgaard. 2020. The low/high index of pupillary activity. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems (CHI)*. 1–12.

[20] Cynthia Dwork, Aaron Roth, et al. 2014. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science* 9, 3–4 (2014), 211–407.

[21] Mayar Elfares, Zhiming Hu, Pascal Reisert, Andreas Bulling, and Ralf Küsters. 2023. Federated Learning for Appearance-based Gaze Estimation in the Wild. In *Proceedings of The 1st Gaze Meets ML workshop*. PMLR, 20–36.

[22] Anton Mølbjerg Eskildsen and Dan Witzner Hansen. 2021. Analysis of iris obfuscation: Generalising eye information processes for privacy studies in eye tracking. In *Proceedings of ACM Symposium on Eye Tracking Research and Applications (ETRA)*. 1–10.

[23] Augusto Esteves, Eduardo Velloso, Andreas Bulling, and Hans Gellersen. 2015. Orbits: Gaze interaction for smart watches using smooth pursuit eye movements. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology (UIST)*. 457–466.

[24] EyeWare. 2016. EyeWare. https://eyeware.tech/, accessed: 2023-11-09.

[25] Terje Falck-Ytter, Sven Bölte, and Gustaf Gredebäck. 2013. Eye tracking in early autism research. *Journal of Neurodevelopmental Disorders* 5, 1 (2013), 1–13.

[26] Liyue Fan. 2018. Image Pixelization with Differential Privacy. In *32th IFIP Annual Conference on Data and Applications Security and Privacy (DBSec)*. Springer International Publishing, 148–162.
[27] Weiwei Feng, Nanqing Xu, Tianzhu Zhang, and Yongdong Zhang. 2023. Dynamic Generative Targeted Attacks With Pattern Injection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 16404–16414.
[28] Lex Fridman, Bryan Reimer, Bruce Mehler, and William T Freeman. 2018. Cognitive load estimation in the wild. In *Proceedings of the 2018 ACM CHI Conference on Human Factors in Computing Systems (CHI)*. 1–9.
[29] GazeRecorder. 2009. GazeRecorder. https://gazerecorder.com/, accessed: 2023-11-09.
[30] Céline Gressel, Rebekah Overdorf, Inken Hagenstedt, Murat Karaboga, Helmut Lurtz, Michael Raschke, and Andreas Bulling. 2023. Privacy-Aware Eye Tracking: Challenges and Future Directions. *IEEE Pervasive Computing* 22, 1 (2023), 95–102.
[31] Elias Daniel Guestrin and Moshe Eizenman. 2006. General theory of remote gaze estimation using the pupil center and corneal reflections. *IEEE Transactions on Biomedical Engineering* 53, 6 (2006), 1124–1133.
[32] Dan Witzner Hansen and Qiang Ji. 2009. In the eye of the beholder: A survey of models for eyes and gaze. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)* 32, 3 (2009), 478–500.
[33] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep Residual Learning for Image Recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 770–778.
[34] Qiuhai He, Xiaopeng Hong, Xiujuan Chai, Jukka Holappa, Guoying Zhao, Xilin Chen, and Matti Pietikäinen. 2015. OMEG: Oulu multi-pose eye gaze dataset. In *Proceedings of Scandinavian Conference on Image Analysis*. Springer, 418–427.
[35] Qiong Huang, Ashok Veeraraghavan, and Ashutosh Sabharwal. 2017. TabletGaze: dataset and analysis for unconstrained appearance-based gaze estimation in mobile tablets. *Machine Vision and Applications* 28, 5 (2017), 445–461.
[36] Andrew Ilyas, Logan Engstrom, Anish Athalye, and Jessy Lin. 2018. Black-box Adversarial Attacks with Limited Queries and Information. In *Proceedings of the 35th International Conference on Machine Learning (ICML)*. 2137–2146.
[37] Yonghyun Jeong, Jooyoung Choi, Sungwon Kim, Youngmin Ro, Tae-Hyun Oh, Doyeon Kim, Heonseok Ha, and Sungroh Yoon. 2021. FICGAN: facial identity controllable GAN for de-identification. *arXiv preprint arXiv:2110.00740* (2021).
[38] Brendan John, Sophie Jörg, Sanjeev Koppal, and Eakta Jain. 2020. The Security-Utility Trade-off for Iris Authentication and Eye Animation for Social Virtual Avatars. *IEEE Transactions on Visualization and Computer Graphics (TVCG)* 26, 5 (2020), 1880–1890.
[39] Brendan John, Ao Liu, Lirong Xia, Sanjeev Koppal, and Eakta Jain. 2020. Let It Snow: Adding Pixel Noise to Protect the User's Identity. In *Proceedings of ACM Symposium on Eye Tracking Research and Applications (ETRA)*. 1–3.
[40] Moritz Kassner, William Patera, and Andreas Bulling. 2014. Pupil: An Open Source Platform for Pervasive Eye Tracking and Mobile Gaze-Based Interaction. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*. 1151–1160.
[41] Christina Katsini, Yasmeen Abdrabou, George E Raptis, Mohamed Khamis, and Florian Alt. 2020. The role of eye gaze in security and privacy applications: Survey and future HCI research directions. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems (CHI)*. 1–21.
[42] Mohamed Khamis, Florian Alt, and Andreas Bulling. 2018. The past, present, and future of gaze-enabled handheld mobile devices: Survey and lessons learned. In *Proceedings of the ACM International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI)*. 1–17.
[43] Jiseob Kim, Jihoon Lee, and Byoung-Tak Zhang. 2022. Smooth-Swap: A Simple Enhancement for Face-Swapping With Smoothness. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 10779–10788.
[44] Joohwan Kim, Michael Stengel, Alexander Majercik, Shalini De Mello, David Dunn, Samuli Laine, Morgan McGuire, and David Luebke. 2019. NVGaze: An anatomically-informed dataset for low-latency, near-eye gaze estimation. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI)*. 1–12.
[45] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).
[46] Diederik P Kingma and Max Welling. 2013. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114* (2013).
[47] Thomas Kosch, Mariam Hassib, Paweł W Woźniak, Daniel Buschek, and Florian Alt. 2018. Your eyes tell: Leveraging smooth pursuit for assessing cognitive workload. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems (CHI)*. 1–13.
[48] Rakshit Kothari, Shalini De Mello, Umar Iqbal, Wonmin Byeon, Seonwook Park, and Jan Kautz. 2021. Weakly-Supervised Physically Unconstrained Gaze Estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 9980–9989.
[49] Kyle Krafka, Aditya Khosla, Petr Kellnhofer, Harini Kannan, Suchendra Bhandarkar, Wojciech Matusik, and Antonio Torralba. 2016. Eye Tracking for Everyone. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2176–2184.
[50] Jacob Leon Kröger, Otto Hans-Martin Lutz, and Florian Müller. 2020. What does your gaze reveal about you? On the privacy implications of eye tracking. *Privacy and Identity Management. Data for Better Living: AI and Privacy: 14th IFIP WG 9.2, 9.6/11.7, 11.6/SIG 9.2. 2 International Summer School, Windisch, Switzerland, August 19–23, 2019, Revised Selected Papers 14* (2020), 226–241.
[51] Zhenzhong Kuang, Huigui Liu, Jun Yu, Aikui Tian, Lei Wang, Jianping Fan, and Noboru Babaguchi. 2021. Effective De-identification Generative Adversarial Network for Face Anonymization. In *Proceedings of the 29th ACM International Conference on Multimedia (MM)*.

3182–3191.
[52] Hsin-Yu Lai, Charles G Sodini, Vivienne Sze, and Thomas Heldt. 2023. Individualized Tracking of Neurocognitive-State-Dependent Eye-Movement Features Using Mobile Devices. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 7, 1 (2023), 1–23.
[53] Guohao Lan, Bailey Heit, Tim Scargill, and Maria Gorlatova. 2020. GazeGraph: Graph-based few-shot cognitive context sensing from human visual behavior. In *Proceedings of the 18th Conference on Embedded Networked Sensor Systems (SenSys)*. 422–435.
[54] Guohao Lan, Tim Scargill, and Maria Gorlatova. 2022. EyeSyn: Psychology-inspired Eye Movement Synthesis for Gaze-based Activity Recognition. In *Proceedings of the 21st ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*. IEEE, 233–246.
[55] Jingjie Li, Amrita Roy Chowdhury, Kassem Fawaz, and Younghyun Kim. 2021. Kaleido: Real-Time Privacy Control for Eye-Tracking Systems. In *Proceedings of 30th USENIX Security Symposium (USENIX Security)*. 1793–1810.
[56] Tao Li and Chris Clifton. 2021. Differentially private imaging via latent space manipulation. *arXiv preprint arXiv:2103.05472* (2021).
[57] Daniel J Liebling and Sören Preibusch. 2014. Privacy considerations for a pervasive eye tracking world. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication (UbiComp)*. 1169–1177.
[58] Sicong Liu, Junzhao Du, Anshumali Shrivastava, and Lin Zhong. 2019. Privacy Adversarial Network: Representation Learning for Mobile Data Privacy. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 4 (2019), 1–18.
[59] Feng Lu, Yusuke Sugano, Takahiro Okabe, and Yoichi Sato. 2014. Adaptive linear regression for appearance-based gaze estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 36, 10 (2014), 2033–2046.
[60] Aleksander Madry, Aleksandar Makelov, Ludwig Schmidt, Dimitris Tsipras, and Adrian Vladu. 2017. Towards deep learning models resistant to adversarial attacks. *arXiv preprint arXiv:1706.06083* (2017).
[61] Alexander Mariakakis, Mayank Goel, Md Tanvir Islam Aumi, Shwetak N Patel, and Jacob O Wobbrock. 2015. SwitchBack: Using focus and saccade tracking to guide users' attention for mobile task resumption. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI)*. 2953–2962.
[62] Blaz Meden, Ziga Emersic, Vitomir Struc, and Peter Peer. 2017. k-Same-Net: Neural Network Based Face Deidentification. In *Proceedings of the 2017 International Conference and Workshop on Bioinspired Intelligence (IWOBI)*. 1–7.
[63] Nandini Modi and Jaiteg Singh. 2023. Understanding online consumer behavior at e-commerce portals using eye-gaze tracking. *International Journal of Human–Computer Interaction* 39, 4 (2023), 721–742.
[64] Atsushi Nakazawa and Christian Nitschke. 2012. Point of gaze estimation through corneal surface reflection in an active illumination environment. In *Proceedings of the European Conference on Computer Vision (ECCV)*. 159–172.
[65] Omar Namnakani, Yasmeen Abdrabou, Jonathan Grizou, Augusto Esteves, and Mohamed Khamis. 2023. Comparing Dwell time, Pursuits and Gaze Gestures for Gaze Interaction on Handheld Mobile Devices. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems (CHI)*. 1–17.
[66] Jacek Naruniec, Leonhard Helminger, Christopher Schroers, and Romann M Weber. 2020. High-resolution neural face swapping for visual effects. In *Computer Graphics Forum*, Vol. 39. Wiley Online Library, 173–184.
[67] Yuval Nirkin, Yosi Keller, and Tal Hassner. 2019. FSGAN: Subject Agnostic Face Swapping and Reenactment. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. 7184–7193.
[68] Yuval Nirkin, Iacopo Masi, Anh Tran Tuan, Tal Hassner, and Gerard Medioni. 2018. On Face Segmentation, Face Swapping, and Face Perception. In *Proceedings of the 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG)*. 98–105.
[69] Seong Joon Oh, Mario Fritz, and Bernt Schiele. 2017. Adversarial image perturbation for privacy protection a game theory perspective. In *Proceedings of IEEE International Conference on Computer Vision (ICCV)*. IEEE, 1491–1500.
[70] Ivan Perov, Daiheng Gao, Nikolay Chervoniy, Kunlin Liu, Sugasa Marangonda, Chris Umé, Mr Dpfks, Carl Shift Facenheim, Luis RP, Jian Jiang, et al. 2020. DeepFaceLab: Integrated, flexible and extensible face-swapping framework. *arXiv preprint arXiv:2005.05535* (2020).
[71] Francesco Pittaluga, Sanjeev Koppal, and Ayan Chakrabarti. 2019. Learning Privacy Preserving Encodings Through Adversarial Training. In *Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV)*. 791–799.
[72] Alexander Plopski, Teresa Hirzle, Nahal Norouzi, Long Qian, Gerd Bruder, and Tobias Langlotz. 2022. The eye in extended reality: A survey on gaze interaction and eye tracking in head-worn extended reality. *ACM Computing Surveys (CSUR)* 55, 3 (2022), 1–39.
[73] RealEye. 2017. RealEye. https://www.realeye.io/, accessed: 2023-11-09.
[74] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. 2018. MobileNetV2: Inverted Residuals and Linear Bottlenecks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 4510–4520.
[75] Andreas Schulz and Rainer Stiefelhagen. 2012. Video-based pedestrian head pose estimation for risk assessment. In *Proceedings of the 15th International IEEE Conference on Intelligent Transportation Systems*. 1771–1776.
[76] SeeSo. 2021. SeeSo. https://seeso.io/, accessed: 2023-11-09.
[77] Shawn Shan, Emily Wenger, Jiayun Zhang, Huiying Li, Haitao Zheng, and Ben Y. Zhao. 2020. Fawkes: Protecting Privacy against Unauthorized Deep Learning Models. In *Proceedings of the 29th USENIX Security Symposium (USENIX Security)*. 1589–1604.

[78] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).

[79] B.A. Smith, Q. Yin, S.K. Feiner, and S.K. Nayar. 2013. Gaze Locking: Passive Eye Contact Detection for Human?Object Interaction. In *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST)*. 271–280.

[80] Namrata Srivastava, Joshua Newn, and Eduardo Velloso. 2018. Combining low and mid-level gaze features for desktop activity recognition. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 4 (2018), 1–27.

[81] Julian Steil, Inken Hagestedt, Michael Xuelin Huang, and Andreas Bulling. 2019. Privacy-aware eye tracking using differential privacy. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications (ETRA)*. 1–9.

[82] Julian Steil, Marion Koelle, Wilko Heuten, Susanne Boll, and Andreas Bulling. 2019. Privaceye: privacy-preserving head-mounted eye tracking using egocentric scene image and eye movement features. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications (ETRA)*. 1–10.

[83] Kar-Han Tan, David J Kriegman, and Narendra Ahuja. 2002. Appearance-based eye gaze estimation. In *Sixth IEEE Workshop on Applications of Computer Vision, 2002.(WACV 2002). Proceedings.* IEEE, 191–195.

[84] Mingxing Tan and Quoc Le. 2019. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In *Proceedings of the 36th International Conference on Machine Learning (ICML)*. PMLR, 6105–6114.

[85] VicarVison. 2007. VicarVison. https://vicarvision.nl/blog/eyereader-webcam-based-eye-tracking-technology/, accessed: 2023-11-09.

[86] Bryan Wang and Tovi Grossman. 2020. BlyncSync: Enabling multimodal smartwatch gestures with synchronous touch and blink. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems (CHI)*. 1–14.

[87] Yaoming Wang, Yangzhou Jiang, Jin Li, Bingbing Ni, Wenrui Dai, Chenglin Li, Hongkai Xiong, and Teng Li. 2022. Contrastive Regression for Domain Adaptation on Gaze Estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 19376–19385.

[88] Z. Wang, E.P. Simoncelli, and A.C. Bovik. 2003. Multiscale structural similarity for image quality assessment. In *Proceedings of the Thrity-Seventh Asilomar Conference on Signals, Systems & Computers*, Vol. 2. 1398–1402.

[89] Chatchai Wangwiwattana, Xinyi Ding, and Eric C Larson. 2018. Pupilnet, measuring task evoked pupillary response using commodity rgb tablet cameras: Comparison to mobile, infrared gaze trackers for inferring cognitive load. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 4 (2018), 1–26.

[90] Yunqian Wen, Bo Liu, Ming Ding, Rong Xie, and Li Song. 2022. Identitydp: Differential private identification protection for face images. *Neurocomputing* 501 (2022), 197–211.

[91] Ethan Wilson, Frederick Shic, and Eakta Jain. 2023. Introducing Explicit Gaze Constraints to Face Swapping. In *Proceedings of the 2023 Symposium on Eye Tracking Research and Applications (ETRA)*. 1–7.

[92] Hao Wu, Jinghao Feng, Xuejin Tian, Edward Sun, Yunxin Liu, Bo Dong, Fengyuan Xu, and Sheng Zhong. 2020. EMO: real-time emotion recognition from single-eye images for resource-constrained eyewear devices. In *Proceedings of the 18th International Conference on Mobile Systems, Applications, and Services (MobiSys)*. 448–461.

[93] Hao Wu, Xuejin Tian, Yuhang Gong, Xing Su, Minghao Li, and Fengyuan Xu. 2021. DAPter: Preventing user data abuse in deep learning inference services. In *Proceedings of the Web Conference*. 1017–1028.

[94] Taihong Xiao, Yi-Hsuan Tsai, Kihyuk Sohn, Manmohan Chandraker, and Ming-Hsuan Yang. 2020. Adversarial learning of privacy-preserving and task-oriented representations. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, Vol. 34. 12434–12441.

[95] Hanyu Xue, Bo Liu, Ming Ding, Tianqing Zhu, Dayong Ye, Li Song, and Wanlei Zhou. 2021. Dp-image: Differential privacy for image data in feature space. *arXiv preprint arXiv:2103.07073* (2021).

[96] Hanyu Xue, Bo Liu, Xin Yuan, Ming Ding, and Tianqing Zhu. 2023. Face image de-identification by feature space adversarial perturbation. *Concurrency and Computation: Practice and Experience* 35, 5 (2023), e7554.

[97] Xiao Zeng, Kai Cao, and Mi Zhang. 2017. MobileDeepPill: A Small-Footprint Mobile Deep Learning System for Recognizing Unconstrained Pill Images. In *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys)*. 56–67.

[98] Tongyu Zhang, Yiran Shen, Guangrong Zhao, Lin Wang, Xiaoming Chen, Lu Bai, and Yuanfeng Zhou. 2024. Swift-Eye: Towards Anti-blink Pupil Tracking for Precise and Robust High-Frequency Near-Eye Movement Analysis with Event Cameras. *IEEE Transactions on Visualization and Computer Graphics (TVCG)* (2024).

[99] Xucong Zhang, Seonwook Park, Thabo Beeler, Derek Bradley, Siyu Tang, and Otmar Hilliges. 2020. ETH-XGaze: A Large Scale Dataset for Gaze Estimation under Extreme Head Pose and Gaze Variation. In *Proceedings of the European Conference on Computer Vision (ECCV)*. 365–381.

[100] Xucong Zhang, Yusuke Sugano, Mario Fritz, and Andreas Bulling. 2015. Appearance-based gaze estimation in the wild. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 4511–4520.

[101] Xucong Zhang, Yusuke Sugano, Mario Fritz, and Andreas Bulling. 2017. It's Written All Over Your Face: Full-Face Appearance-Based Gaze Estimation. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 51–60.

[102] Xiangyu Zhang, Xinyu Zhou, Mengxiao Lin, and Jian Sun. 2018. ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 6848–6856.

[103] Yanxia Zhang, Ming Ki Chong, Jörg Müller, Andreas Bulling, and Hans Gellersen. 2015. Eye tracking for public displays in the wild. *Personal and Ubiquitous Computing* 19 (2015), 967–981.

[104] Guangrong Zhao, Yurun Yang, Jingwei Liu, Ning Chen, Yiran Shen, Hongkai Wen, and Guohao Lan. 2024. EV-Eye: Rethinking high-frequency eye tracking through the lenses of event cameras. *Advances in Neural Information Processing Systems (NeurIPS)* 36 (2024), 62169–62182.

[105] Yufeng Zheng, Seonwook Park, Xucong Zhang, Shalini De Mello, and Otmar Hilliges. 2020. Self-Learning Transformations for Improving Gaze and Head Redirection. In *Proceedings of the Neural Information Processing Systems (NeurIPS)*, Vol. 33. 13127–13138.

[106] Yuhao Zhu, Qi Li, Jian Wang, Cheng-Zhong Xu, and Zhenan Sun. 2021. One Shot Face Swapping on Megapixels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 4834–4844.

[107] Zhiwei Zhu and Qiang Ji. 2007. Novel eye gaze tracking techniques under natural head movement. *IEEE Transactions on Biomedical Engineering* 54, 12 (2007), 2246–2260.