This CVPR Workshop paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

Outsmarting Biometric Imposters: Enhancing Iris-Recognition System Security through Physical Adversarial Example Generation and PAD Fine-Tuning

Yuka Ogino, Kazuya Kakizaki, Takahiro Toizumi, Atsushi Ito
NEC Corporation
yogino@nec.com, kazuya1210@nec.com, t-toizumi_ct@nec.com, ito-atsushi@nec.com

Abstract

In this paper, we address the vulnerabilities of iris recognition systems to both image-based impersonation attacks and Presentation Attacks (PAs) in physical environments. While existing Presentation Attack Detection (PAD) methods have been effective against PAs, they remain susceptible to adversarial examples. We propose a combination of physical adversarial attacks tailored to iris recognition and PAD, and also propose a defense method against them. Our attack methods involve a physical impersonation attack using adversarial perturbation on the iris region and a physical PAD evading attack using an adversarial patch on the pupil region. We demonstrate the high transferability and effectiveness of our attacks on multiple PA instruments in digital and distinct physical environments using multiple recognition engines. To counteract these attacks, we develop a defense method for PAD involving adversarial fine-tuning against both the physical attacks. This defense method successfully reduces the PAD evasion attack success rate from 71.5% to 21.0% in physical environments and ultimately lowers the overall physical impersonation success rate from 58.0% to 19.5%. Our proposed method lays the groundwork for developing more robust and secure iris recognition systems with increased protection against sophisticated PAs.

1. Introduction

Iris recognition is one of the most reliable biometric recognition methods, utilizing the unique texture patterns of the iris region to verify an individual [18]. However, iris recognition systems face two main types of potential threats: image-based impersonation attacks [57, 58] that target digital images and subsequent Presentation Attacks (PAs) [12, 13] executed on physically captured images.

Image-based impersonation attacks manipulate the query iris image using target iris image (or feature), leading the system to incorrect verification. Various techniques

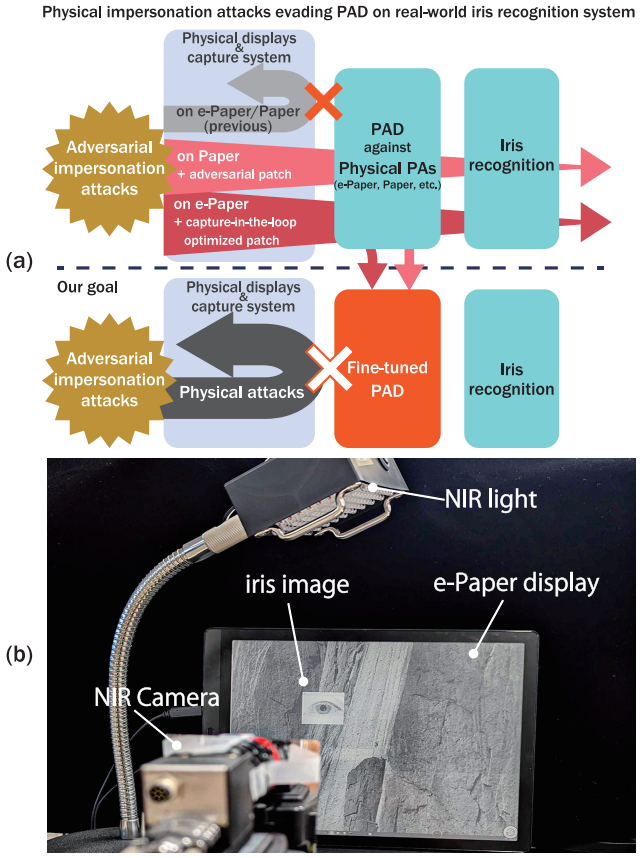


Figure 1. (a) Our real-world attack scenario and goal. (b) An example of our scenario. The target iris image is displayed on the e-Paper illuminated using the near-infrared (NIR) light. NIR camera for iris recognition captures displayed iris images.

are employed in these attacks which include the utilization of adversarial perturbations [58] and image morphing [57]. However, these methods have only been evaluated on digital images and their effectiveness in physical environments, such as on printed paper or displayed screens, remains unassessed. Therefore, confirming the effectiveness of these attacks in real-world physical environments

becomes crucial.

Presentation Attacks (PAs) [12, 13] typically make use of presentation attack instruments (PAIs) such as prosthetic eyes, textured contact lenses, printed iris images, or iris images displayed on electronic paper (e-Paper) screens to deceive iris recognition systems. These PAIs are captured by the system, misleading the verification process. PAs can co-occur with image-based impersonation attacks, for instance, by displaying a generated image on e-Paper screens or printed papers. Thus, Presentation Attack Detection (PAD) plays a vital role in iris recognition systems to prevent both types of attacks on physically-captured images.

PAD methods detects PAs by deriving PA scores from cropped iris images.PAD methods [1, 3, 7, 8, 12, 14, 15, 17, 23, 26, 27, 44, 47, 56, 59, 60], have been developed and researched to preemptively identify such attacks. Conventional studies [1, 7, 8, 15, 26, 27, 44, 56, 60] have demonstrated that using Convolutional Neural Networks (CNNs) can significantly enhance the performance of PAD. Especially, D-NetPAD [56] consists of simple DenseNet121 [29] structure achieved most accurate detection in LivDet-Iris2020 Challenge [13].

However, CNN-based PAD methods remain susceptible to adversarial examples[6, 22, 38]. Despite this, exploration into the vulnerabilities within the PAD models themselves and the development of defensive strategies against these are still limited. It is necessary to consider the vulnerabilities of PAD in terms of operational environments in real-world iris recognition systems and to develop corresponding defences. Generally, fine-tuning with adversarial examples is effective [22, 38]. We believe that it is paramount for this fine-tuning process to utilize attacks that hold valid and effective implications in the real world.

Recent studies [2, 4, 21, 28, 36, 54, 64] have proposed physical adversarial attack scenarios on real-world sensing systems. Digitally generated adversarial attacks tend to be ineffective against physically sensed adversarial instances due to distributional shifts due to image degradation from the sensing process. Solutions to these domain shifts are sought through digital optimization techniques in the conduct of physical adversarial attacks. However, these methods have yet to take into account to the sensing system itself. Therefore, it is crucial to examine the efficacy of adversarial attacks and their defenses in the context of iris recognition systems, which operate under near-infrared (NIR) lighting and necessitate high-resolution capturing.

In this paper, we demonstrate physical adversarial examples effective on both iris recognition and PAD. We then apply adversarial training [38] in PAD models to protect from these attacks as illustrated in Figure 1 (a). We propose novel physical adversarial attacks and demonstrate their effectiveness on both iris recognition and PAD.

Our contributions are as follows:

- Physical Impersonation Attack: We propose an attack using adversarial perturbation on the iris region against iris recognition systems. We demonstrate its high transferability and effectiveness in digital and two physical environments with two PAIs.
- Physical PAD Evading Attack: We propose an attack employing an adversarial patch on the pupil region without preventing impersonation attacks. We demonstrate the effectiveness of our capturing-in-the-loop (CIL) optimization method for lower-dpi PAIs such as e-Paper displays. We apply our optimization on one capturing environment and demonstrate that optimized patches are effective on another capturing environment.
- Defense Method: We develop an effective defense method for PAD involving adversarial fine-tuning [22, 38], which protects against both the physical impersonation and PAD evading attacks, as shown through experiments.

We use three iris recognition engines [10, 43, 49] and a publicly available PAD method, D-NetPAD [56] for evaluation of physical attack, and fine-tune D-NetPAD. Our experimental results show the existence of adversarial examples for both iris recognition and PAD in real-world scenarios of iris recognition systems shown in Figure 1 (b), as well as the significant robustness improvement achievable by fine-tuning D-NetPAD. Our proposed method paves the way for developing more robust and secure iris recognition systems with enhanced protection against sophisticated PAs.

2. Related Work

2.1. Iris Recognition

Iris recognition involves four main stages: segmentation, mask creation, normalization, and feature extraction. The process starts by isolating the iris region from the eye image and generating a binary mask to filter out noisy pixels. The annular-shaped iris and its mask are then normalized using the rubber sheet model [18] into a rectangular shape. The normalized images are fed into the feature extractor. Many studies [18, 43, 49] have used hand-crafted features such as two-dimensional (2D) Gabor or wavelet filters and have used the Hamming distance for matching. Other studies [10, 19, 25, 41, 42, 50, 63, 69] have used CNNs as feature extractors, producing 1D feature vectors and leveraging cosine similarity for matching. Iris recognition systems use NIR cameras and require high-resolution images, e.g., 10 pixels per millimeter as per ISO/IEC 39794-6 [30].

2.2. Impersonation Attacks on Iris Images

Two primary types of digital impersonation attacks exist in iris recognition: morphing-based techniques [57] and adversarial example-based methods [58]. Sharma et al. [57] proposed a morphing technique that aligns the iris regions

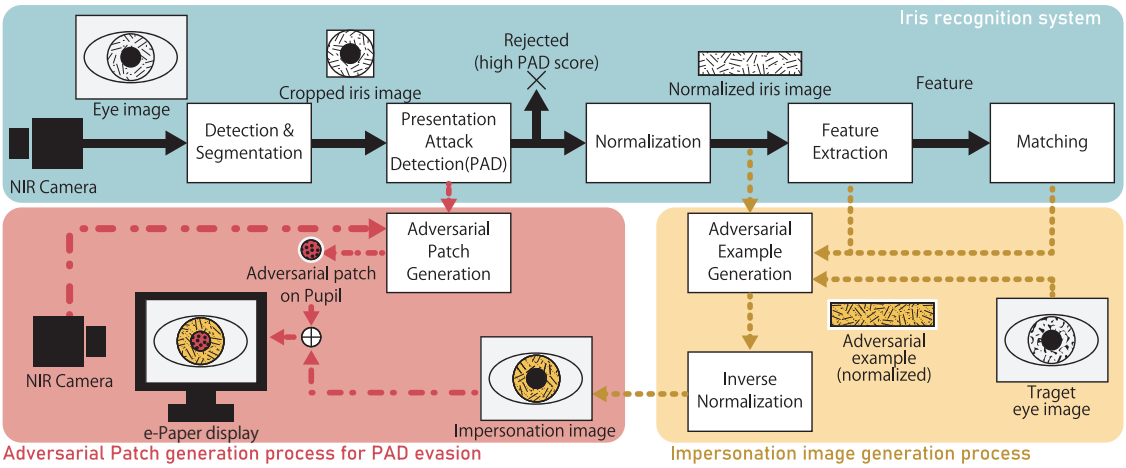


Figure 2. Overview of iris-recognition process (blue region) and adversarial-attacks generation processes. Our method first applies perturbations to iris region for impersonation (yellow region) and optimizes an adversarial patch on pupil region to evade PAD (red region).

of target and query images based on the pupil and iris position information, demonstrating a high recognition success rate for both target and query images in a digital environment. Soleymani et al. [58] targeted the hand-crafted iris recognition method OSIRIS, trained a U-Net [51] based surrogate model on normalized images, and conducted adversarial attacks on the surrogate model. However, it remains unclear whether these attacks remain effective when applied to original annular-shaped iris images or images captured in real-world sensing environments.

2.3. Iris Presentation Attack Detection

PAD protects iris recognition against impersonating specific identities or other PAs [12]. As shown in the upper part of Figure 2, PAD is generally applied on cropped iris images after the segmentation process within the iris recognition pipeline. To enhance the performance of PAD, the LivDet-Iris Challenge [13, 66–68] is held periodically and facilitates the development of PAD methods by focusing on detecting various PAIs such as artificial eyes, GAN-generated images [33], printed paper, cosmetic contact lenses and e-Paper displays in captured images. Similar to iris recognition, PAD methods fall into two categories: hand-crafted and CNN-based methods. Hand-crafted methods use features such as frequency [14], binarized statistical image features (BSIF) [17, 47], and local descriptors [23, 59]. CNN-based methods [1, 7, 8, 15, 26, 27, 44, 56, 60] demonstrated high accuracy in previous LivDet Iris competitions. For example, Sharma et al. [56] proposed D-NetPAD based on DenseNet121 [29]. D-NetPAD achieved the most accurate detection in LivDet-Iris2020. While iris recognition framework solely extracts the iris region from images, PAD framework inputs cropped regions, including the pupil. We believe that this difference could make PAD systems more vulnerable.

2.4. Adversarial Attacks and Defense

CNN models have been found to be vulnerable to adversarial examples. Prior research has investigated generating adversarial examples by either adding small, human-imperceptible perturbations to images [22, 38] or introducing adversarial patches that modify specific areas in images without being confined to imperceptible perturbations [4, 16, 35, 55, 61, 65]. Moreover, several studies [2, 4, 21, 28, 36, 54, 64] have proposed physical adversarial attacks on real-world sensing systems. Generated adversarial examples in digital environments do not consistently fool in physical environments due to factors such as viewpoint changes and camera noise. As a technique for adapting to physical environments, Expectation over Transformation (EOT) [2] introduces image transformations and noise while the attack generation process enhances the robustness of adversarial examples. Physical adversarial examples have been used for attack evaluations by printing them on paper [2, 4, 36, 64] or projecting them using projectors [21, 28, 54]. In the context of iris PAD, there is a commonality in printing on paper, and e-Paper displays can serve as substitutes for projectors.

Various defense methods have been proposed against adversarial attacks, such as detection methods [5, 24], saliency-map based [11, 40], or enhancing model by training and fine-tuning with adversarial examples [9, 20, 22, 34, 37, 38, 48]. Consequently, it is possible to improve defense capabilities by generating image-based weaknesses in PAD, fine-tuning the models, and reinforcing their resilience and detection performance.

3. Proposed Method

We propose two physical adversarial techniques targeting iris recognition and PAD. In the first phase, we develop a physical impersonation attack that leverages internal representations of a normalized iris image to generate adversarial examples. This approach makes the perturbations robust against image deformations and degradations in physical environments. In the second phase, we propose a PAD evading attack by designing an adversarial patch. This generation process is illustrated in Figure 2. Furthermore, to defend against these adversarial attacks, we apply adversarial fine-tuning [22, 38] to the PAD models by incorporating the generated adversarial images in their training phase, resulting in improved robustness against adversarial examples.

3.1. Impersonation for Iris Recognition

We conduct an impersonation attack by generating an adversarial example on the normalized iris image and then warping it back to the original image (inverse normalization). To generate a robust adversarial example, we use a loss function based on the internal representation. This enables us to create effective perturbations resistant to image deformations and degradation introduced by PAIs.

3.1.1 Differentiable Inverse Iris Normalization

The normalization process[18] maps the 2D torus-shaped iris region to a rectangular image. A transformation from orthogonal coordinates $[x, y]$ to normalized image coordinates $[r, \theta]$, is given by

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} R\cos\theta' + x_c \\ R\sin\theta' + y_c \end{bmatrix}, \tag{1}$$

$$R = \left(\frac{r}{h}r_p + (1 - \frac{r}{h})r_i\right), \theta' = \frac{2\pi\theta}{w}, \tag{2}$$

$$\begin{bmatrix} x_c \\ y_c \end{bmatrix} = \begin{bmatrix} \frac{r}{h}x_p + (1 - \frac{r}{h})x_i \\ \frac{r}{h}y_p + (1 - \frac{r}{h})y_i \end{bmatrix}. \tag{3}$$

where $[x_p, y_p]$ denotes the center position of the pupil, and $[x_i, y_i]$ depicts the center position of the iris. The radius of the pupil and iris are denoted as r_p and r_i, respectively. The height and width of the rectangular output image are respectively represented by h and w.

Recently, Khan et al. [32] proposed DeformIrisNet for pupil-deliration simulations, a method that implicitly incorporates differentiable normalization using a spatial grid sampler [31]. We have implemented this differentiable approach in our study for inverse normalization. In this field of study, to the best of our knowledge, no papers providing an analytical exposition on this Inverse Normalization have been published. In light of this, we offer our explanation here.

The inverse transformation from normalized image coordinates $[r, \theta]$ to orthogonal coordinates $[x, y]$ is derived by squaring the sum of Eq. 1, with r representing the positive solutions of the resulting quadratic function as given by

$$x^2 + y^2 = R^2 + x_c^2 + y_c^2. \tag{4}$$

θ is described as

$$\theta = \arctan\left(\frac{y - y_c}{x - x_c}\right). \tag{5}$$

We use bilinear interpolation to smooth the mappings of pixels between the two coordinates. If the iris is cut at the edge of the image, we use border replication for padding. While our method allows for the application of perturbations to the iris area prior to normalization, we apply them after normalization. This is mainly because the perturbation area become excessive if manipulated before normalization.

3.1.2 Optimization Using Internal Representation

When generating adversarial examples in iris recognition, there are two critical issues. First, perturbations consisting of pixel changes without adjacency, such as Gaussian noise, may be lost during interpolation in deformation. Hence, perturbations that can be robust to deformation and interpolation are necessary. Second, iris-recognition systems adopt various matching methods (such as Hamming distance or cosine similarity), so it is necessary to be independent of the particular matching method used as possible. We use the internal representation to generate adversarial examples, which is known for its strong transferability [39, 52, 53]. This method also enables backpropagation on hand-crafted methods such as Gabor filter-based methods [43].

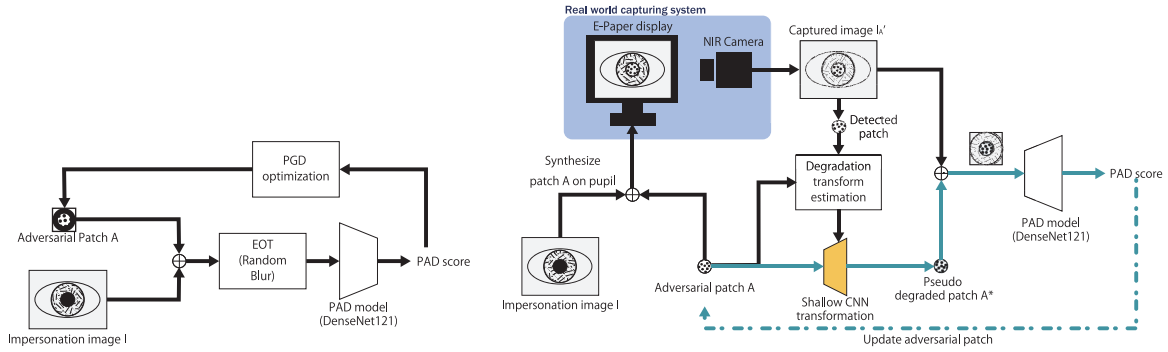
To optimize the generation of adversarial examples, we use the projected gradient descent (PGD) algorithm [38], which iteratively takes small steps in the direction that increases the loss function as given by

$$\mathbf{x}_{adv}^{t+1} = \text{Clip}_{\epsilon}(\mathbf{x}_{adv}^{t} + \alpha\text{sign}(\nabla L(\phi_k(\mathbf{x}_{adv}^{t}), \phi_k(\mathbf{y})))). \tag{6}$$

At each iteration $t + 1$, the equation generates the adversarial example $\mathbf{x}_{adv}^{t+1}$, where L is the loss function based on L2-norm, and ϕ_k describes the intermediate convolutional layer at the k-th level from the input. The hyperparameter α controls the step size of the iteration, while the perturbation is constrained within the ϵ-ball centered on the original image through the clipping operation.

3.2. Adversarial Patch for PAD evasion

We propose a circular-shaped adversarial patch on the pupil region. This method aids in preserving the adversarial perturbations for impersonation within the iris region. The digitally generated adversarial patch becomes ineffective due



(a) Initialization process (b) CIL optimization process

Figure 3. Our adversarial patch generation process. (a) Process of initializing a patch without capturing image, (b) Optimization process of PAD score using an e-Paper display and a camera.

to domain shifts from lower-dpi PAIs such as e-Papers and sensing systems. We consider it critical to establish a correspondence between digital and physical spaces. To bridge this gap, we propose using a capturing-in-the-loop optimization technique.

3.3. Capturing-in-the-loop optimization

Given an iris image $x \in \mathcal{X}$ and an adversarial patch $p \in \mathcal{P}$, we contemplate a PAD model $D : \mathcal{X} \rightarrow \mathbb{R}$. We formulate a synthesizing function $S : \mathcal{X} \times \mathcal{P} \rightarrow \mathcal{X}$ and a digitally synthesized image $x_p = S(x, p)$. The capturing function $C : \mathcal{X} \rightarrow \mathcal{X}$ signifies the display and capture process in the real world where it results in degradation of the input image and the adversarial patch becoming degraded and non-differentiable $\bar{p}$. In an attempt to acquire p in differentiable optimization, we propose an approximation function $f_\theta : \mathcal{P} \rightarrow \mathcal{P}$ with parameter θ to characterize the alignment between the digital and real-world environments. Our proposed approach involves resolving the equation below:

$$\min_{p,\theta} D(S(C(x_p), f_\theta(p))) + ||D_i(f_\theta(p)) - D_i(\bar{p})||_2 \quad (7)$$

In this equation, D_i represents the layers of D up to the i-th layer. We propose an iterative optimization of two parameters p and θ, an approach not yet explored in the context of physical adversarial attacks. Algorithm 1 and Figure 3 (b) outlines our CIL optimization process.

3.3.1 Adversarial Patch Initialization

For the initialization of CIL, we use Expectation Over Transformation (EOT) [2] to handle blur degradation during the patch initialization, as depicted in Figure 3 (a). Our initialization approach creates a black circular patch of radius t pixels and applies a mask of d pixels from the edge. We then incorporate the patch into the pupil area of the image and employ the Projected Gradient Descent (PGD) algorithm [38] to optimize the perturbation within a range of radius $t - d$ pixels. Using the random blur convolution as EOT enhances the patch's robustness against degradation. By optimizing with multiple images, we strengthen the patch's robustness, enabling it to handle potential variations in pupil dilation effectively. To address variations in pupil size among different irises, we use differentiable image sampling [31]. This performs an affine transformation on the circular patch to match with varying pupil sizes in the images, effectively acting as an EOT for scaling and enhancing the adaptability of the adversarial patch.

Algorithm 1 CIL algorithm

Require: initialized adversarial patch p_0,
impersonation image x
while $Score > threshold$ & $captureCont < maxCapture$ **do**
  $x_p \leftarrow S(x, p)$
  $x'_p \leftarrow C(x_p)$
  $Score \leftarrow PAD(x'_p)$
  $p' \leftarrow Extract from x'_p$
  for $i = 0; i < train_iterations; i++$ **do**
    $\min_\theta ||D_i(f_\theta(p)) - D_i(\hat{p})||_2$
    $i++$
  end for
  for $i = 0; i < max_iterations; i++$ **do**
    $Score \leftarrow D(S(x'_p, f_\theta(p))$
    $\min_p D(S(x'_p, f_\theta(p))$
  end for
  if $score < PADthreshold$ **then**
    break
  end if
end while

3.4. Enhancing PAD

We fine-tune PAD models using additional generated adversarial examples. This enables the PAD model to improve the accuracy of classifying generated adversarial iris images and suppresses the effectiveness of newly created adversar-

ial patches. Our proposed method involves real-time data generation and optimization in a real-world environment. Since carrying out iterative adversarial training involving repeated learning and generation is unrealistic, we assume that fine-tuning yields sufficient improvements in the performance of the PAD model.

4. Experiments

In our experiments, we created physical environments utilizing two distinct capturing settings and employed paper and e-Paper as PAIs. Firstly, we demonstrate that our proposed iris impersonation method outperforms conventional approaches by achieving higher success rates in both digital and physical environments. Additionally, we highlight the effectiveness of our adversarial patches against D-NetPAD [56] and establish that they maintain their efficiency in unknown physical settings. Furthermore, we show the effectiveness of defending against attacks by fine-tuning the images generated through attacks.

4.1. Iris Recognition System Environments

We explored the performance of our impersonation method with two iris recognition engines. These engines consist of the publicly available hand-crafted method OSIRIS [43] and the CNN-based approach T-Center [10]. Moreover, we assessed the transferability of impersonation using an additional iris recognition engine, the publicly available hand-crafted method USITv3 [49]. The recognition performance of each engine was evaluated on the ND-IRIS-0405 dataset [45, 46], consisting of 13,438 positive pairs and 1,985,562 negative pairs. The evaluation results for equal error rates (EERs) for T-Center, OSIRIS, and USITv3 were 4.46%, 2.06%, and 10.43%, respectively. For evaluations of impersonation attacks, we used the score at FAR $= 10^{-3}$ as matching thresholds.

We established two different imaging environments, each featuring a distinct camera setup. The first imaging environment (Camera-1) utilizes an IMX267 image sensor and with a 50 mm C-mount lens. We performed our capturing-in-the-loop optimization in this environment. The second imaging environment (Camera-2) involves an IMX178 image sensor with a 35 mm C-mount lens. We used near-infrared (NIR) light and a visible light cut filter in front of the lenses. These devices has enough optical resolutions and sensitivity for the iris recognition products. For the e-Paper display, we employed a 13.3-inch Boox Max Lumi screen (207 dpi). As for the paper, we printed the images on office recycled paper using a Canon iR-ADV C5560F printer (600 dpi). When generating the PA images, we adjusted the iris diameters to be between 10-12 mm for both the paper and electronic paper (e-Paper) display. We captured PAI with the diameter of the iris being at least 160 pixels.

4.2. Impersonation on Iris Region

To show the impersonation performance of internal representation, we compared five impersonation methods in digital and two camera settings:

- T-Center+PGD: This method is basic adversarial-example-generation on T-Center [10]. Matching scores (cosine similarity of features) are used for PGD loss.
- irisMorph [57]: This method involves morphing the query image's iris positions to match the target image and using linear blending to combine the images.
- Sur-OSIRIS [58]: This method trains a surrogate model of hand-crafted OSIRIS [43] feature extraction and generates adversarial examples using surrogate model (U-Net [51] structure). We trained the surrogate model using 10,000 images from ND-iris-0405 dataset. L2-norm between iris codes is used for PGD loss.
- T-Center-inter (proposed): Based on T-Center [10], this method employs L2-norm between internal representations from the Conv6 layer as PGD loss.
- OSIRIS-inter (proposed): This method is based on OSIRIS [43]. OSIRIS employs six different sizes of Gabor filters. We implemented these filters as convolution layers. The PGD loss is L2-norm between (6,h,w) sized internal representations.

We optimized these impersonation methods with a parameter of $\alpha = 1/255$ and evaluated three different ϵ values: $\{8/255, 16/255, 24/255\}$. To localize the iris circles for attacks, we used the iris localization network (ILN) [62]. In the evaluation phase, we used the detection methods of the three unique iris recognition engines. We randomly determined the source and target image pairs from the ND-IRIS-0405 dataset, ensuring that the images used for training T-Center were excluded. The class of the source and target images was also kept different. We assessed the impersonation success rates (ISRs) for both original images and e-Paper-displayed images. For the original images, we randomly selected 200 pairs.

The experimental results are shown in Table 1. As a result of all evaluations, our method achieves high ISRs in different engines and environments. The proposed OSIRIS-inter exhibits the highest ISRs in the digital environment using the OSIRIS engine, and the OSIRIS and USITv3 engines in multiple physical environments using multiple PAIs. While ISR of the proposed T-Center-inter achieves highest ISR in digital environment. Although the ISRs evaluated using CNN-based recognition T-Center result in that all approaches have low ISRs in physical environments. This result shows that the image domain which T-Center model trained differs from the image of PAIs.

These results confirm that our proposed methods demonstrate competitive performance compared to other methods. Especially in T-Center+PGD, perturbations on normalized images tend to be spatially high frequency and are canceled

Table 1. Impersonation attack success rate (ISR) at FAR $= 10^{-3}$. (w/o PAD)

	eps	Digital			e-Paper@Camera-1			Paper@Camera1			e-Paper@Camera-2			Paper@Camera-2		
		OSIRIS	T-Center	USITv3	OSIRIS	T-Center	USITv3	OSIRIS	T-Center	USITv3	OSIRIS	T-Center	USITv3	OSIRIS	T-Center	USITv3
T-Center + PGD [38]	08/255	0.000	0.010	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	16/255	0.000	0.010	0.000	0.005	0.000	0.010	0.010	0.000	0.000	0.000	0.000	0.005	0.010	0.000	0.000
	24/255	0.000	0.005	0.000	0.000	0.000	0.005	0.005	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.005
irisMorph [57]	–	0.940	0.235	**0.885**	0.105	0.000	0.110	0.765	0.025	0.350	0.070	0.000	0.115	0.555	0.005	0.040
Sur-OSIRIS [58]	08/255	0.670	0.030	0.495	0.000	0.000	0.010	0.620	0.000	0.330	0.095	0.000	0.045	0.470	0.000	0.110
	16/255	0.910	0.020	0.580	0.025	0.000	0.100	0.870	0.005	0.360	0.290	0.000	0.190	0.750	0.000	0.155
	24/255	0.935	0.040	0.585	0.095	0.000	0.105	**0.935**	0.015	0.350	**0.425**	0.000	0.250	**0.765**	0.005	0.150
TCenter-inter (ours)	08/255	0.335	0.060	0.340	0.020	0.000	0.030	0.275	0.000	0.165	0.020	0.000	0.030	0.185	0.000	0.060
	16/255	0.785	0.220	0.705	0.110	0.000	0.140	0.685	0.030	0.310	0.115	0.000	0.130	0.535	0.005	0.120
	24/255	0.895	**0.335**	0.755	0.165	0.000	0.215	0.740	**0.040**	0.330	0.205	0.000	0.190	0.630	0.020	0.140
OSIRIS-inter (ours)	08/255	0.865	0.025	0.820	0.070	0.000	0.100	0.640	0.005	0.380	0.075	0.000	0.105	0.470	0.000	0.120
	16/255	0.960	0.120	0.860	0.203	0.000	**0.305**	0.820	0.015	0.385	0.215	0.000	0.325	0.685	0.005	0.150
	24/255	**0.975**	0.120	0.860	**0.295**	0.000	0.300	0.860	0.035	**0.400**	0.305	0.000	**0.375**	0.690	0.005	**0.170**

Table 2. PAD evasion attack success rate (ESR).

	patch	Camera-1		Camera-2	
		e-Paper	Paper	e-Paper	Paper
D-NetPAD-org	without	0.000	0.000	0.000	0.000
	initial	0.330	**0.460**	0.505	**0.510**
	optimized	**0.880**	0.170	**0.520**	0.435
D-NetPAD-PA	without	0.000	0.000	0.000	0.000
	initial	0.020	**0.685**	0.170	**0.335**
	optimized	**0.715**	0.345	**0.535**	0.260
D-NetPAD-AF	without	0.000	0.000	0.000	0.000
	initial	0.000	**0.210**	0.000	**0.010**
	optimized	0.000	0.105	0.000	0.000

out by inverse normalization. This cancel-out occurs due to the lack of spatial information caused by the fully connected layers of T-Center feature extractor. On the other hand, Sur-OSIRIS extracts two-dimensional features from a U-Net structure consisting only of convolution and deconvolution layers. We considered that this structure also has the function of preserving spatial features, similar to our internal representation-based method.

4.3. Adversarial Patch and Adversarial Training

To initialize the adversarial patch, we used 20 randomly sampled iris images and optimize with parameters $\alpha = 2/255$ and $\epsilon = 255/255$. We conducted 100 iterations per image to initialize the patch. The size of the patch were set to 35×35 pixels, which means the radius of circular patch is $t = 17$ pixels. We also set the mask range d at 5 pixels. For the random blurring parameter, we defined a range of $1.5 \leq \sigma \leq 5$ with a kernel size of 21. To optimize and evaluate the patch, we used OSIRIS-inter images (with ϵ = 24/255) that exhibited high ISRs in previous experiments.

In the capturing optimization process, we set PGD optimization parameters as $(\alpha, \epsilon) = (5/255, 10/255)$, parameters in Algorithm 1 as $maxIteration = 2$ and $maxCapture = 50$. For the shallow CNN T described in Figure 3 (b), we used the following three distinct layers: conv1 expands the channels from 1 to 16 using an (11, 11) kernel size, conv2 maintains the 16 channels with a (5, 5) kernel size, and conv3 reduces the channels from 16 to 1 using a (5, 5) kernel size. To preserve the image size throughout the network, we apply zero padding. For evaluation, we captured randomly sampled 200 images. We used ILN [62] for detection during the attack and evaluation processes.

We focused on evaluating the evading performance of D-NetPAD [56], a publicly available training model with trained weights and threshold values. However, we assumed that our imaging environment and image domain differ from those in the original model. To address these issues, we evaluate multiple fine-tuned D-NetPAD models using PAI images captured in real-world settings and original images from the dataset. We define original and fine-tuned models as follows:

- D-NetPAD-org: Publicity available trained model of [56].
- D-NetPAD-PA: Fine-tuned model using randomly selected 400 live-labeled images from ND-IRIS-0405, 200 e-Paper-displayed images, and 200 paper-printed images for fine-tuning.
- D-NetPAD-AF: Fine-tuned model using 400 live (same as PA model), 100 paper, 100 e-Paper w/o patches, 100 paper with initial patches, and 100 e-Paper with optimized patches. The patches are generated using D-NetPAD-org.

We used 50% of the images for training and the other 50% of images for testing. In addition, the e-Paper and paper data for training and testing were captured in Camera-1. We fine-tuned 2 epochs because the PAD success rate on the test data reached one after two epochs. The threshold for PAD is determined based on the distribution of the test data according to the method described in the D-NetPAD paper and source code.

Table 2 shows the result of PAD evasion attacks using three PAD models when applied to adversarial patches under Camera-1 and Camera-2 environments using printed-paper and e-Paper displayed images. The results show that all PAD evasion success rates called Imposter Attack Presentation Match Rate (IAPMR) become 0 without adversarial patches in all environments. The initial patches on printed-paper images resulted in the highest IAPMR across both cameras. Furthermore, when applying the optimized

Table 3. Impersonation attack success rate (ISR) evading PAD at FAR $= 10^{-3}$.

	patch	e-Paper@Camera-1			Paper@Camera-1			e-Paper@Camera-2			Paper@Camera-2		
		OSIRIS	T-Center	USITv3	OSIRIS	T-Center	USITv3	OSIRIS	T-Center	USITv3	OSIRIS	T-Center	USITv3
D-NetPAD-org	without	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	initial	0.115	0.000	0.155	**0.395**	**0.010**	**0.190**	**0.160**	0.000	**0.195**	**0.355**	0.000	**0.120**
	optimized	**0.215**	0.000	**0.315**	0.155	0.005	0.060	0.155	0.000	0.180	0.340	0.000	0.090
D-NetPAD-PA	without	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	initial	0.010	0.000	0.010	**0.580**	**0.005**	**0.260**	**0.045**	0.000	**0.065**	**0.240**	**0.005**	**0.070**
	optimized	**0.185**	0.000	**0.275**	0.000	0.000	0.000	0.080	0.000	0.135	0.000	0.000	0.000
D-NetPAD-AF	without	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	initial	0.000	0.000	0.000	**0.195**	**0.005**	**0.065**	0.000	0.000	0.000	**0.005**	0.000	**0.005**
	optimized	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 4. Cross PAD evasion Imposter Attack Presentation Match Rate (IAPMR) and PAD performance evaluation @Camera-2.

	+PA patch IAPMR		+AF patch IAPMR		PAD performance	
	e-Paper	Paper	e-Paper	Paper	APCER	BPCER
PAD-PA	0.535	0.335	0.330	0.265	0.005	0.000
PAD-AF	**0.000**	**0.000**	**0.000**	**0.010**	**0.000**	0.000

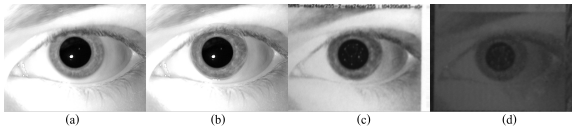


Figure 4. (a) query image (b) impersonation result (OSIRIS-inter) (c) printed-paper image with initial patch (D-NetPAD-org@Camera-1) (d) e-Paper-displayed image with optimized patch (D-NetPAD-org@Camera-1)

adversarial patch for the e-Paper@Camera-1 environment, the results achieved the highest IAPMRs across both camera environments with e-Paper. Comparatively, the performance of the proposed D-NetPAD-AF model is generally lower than that of D-NetPAD-org and D-NetPAD-PA. This result suggests that our fine-tuning enhances PAD against adversarial patches.

Table 3 presents the final ISR evading PAD protection, specifically focusing on those targeting OSIRIS-inter (eps=24/255). For cases where no fine-tuning was performed (D-NetPAD-org and D-NetPAD-PA), the e-Paper@Camera-1 experienced a maximum ISR of 31%. Further, in the Paper@Camera-1, this increased significantly to 58%. In the e-Paper@Camera-2 and Paper@Camera-2, ISRs were 19.5% and 35.5%, respectively. These results demonstrate the vulnerability of the PAD models without fine-tuning. In contrast, D-NetPAD-AF showed defensive capabilities against impersonation attacks, achieving a maximum ISR of 19.5%, considerably lower than the other models. This finding highlights the effectiveness of D-NetPAD-AF in improving robustness against adversarial attacks and the importance of fine-tuning the PAD model for enhanced security.

Furthermore, we compared the cross IAPMRs between the PA and AF models and benchmarked the PAD performances of each model in Table 4. We first evaluated data

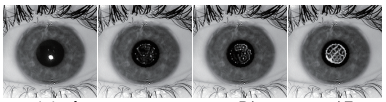


Figure 5. PAD evasion attack images with initialized patches

with patches generated using one model by the other model. As a result, the IAPMRs of the PA model are high, even using the patches generated by the AF model. On the other hand, the IAPMRs of the AF model result in almost zero. These results show that fine-tuning enhances the resilience of the model against adversarial patches. To evaluate the PAD performance itself, we arranged additional evaluation images. We selected 200 digital iris images from the ND-IRIS-0405 as live labels and obtained 400 PA images captured live images on e-Paper and paper in the Camera-2 environment. The selected live images are not used for patch generation and PAD fine-tuning. The right side of Table 4 shows the bonafide presentation classification error rate (BPCER) and attack presentation classification error rate (APCER) of each model. Both models achieve low error rates, showing that our fine-tuning method enhances PAD models while maintaining PAD performance.

5. Conclusion

We address the vulnerabilities of iris recognition systems to both image-based impersonation attacks and Presentation Attacks (PAs) in physical environments. Our contributions include physical impersonation attacks that target both the iris region and PAD methods, and a defense method using adversarial fine-tuning for enhancing the robustness of PAD against these attacks. The experimental results show the existence of physical adversarial examples for both iris recognition and PAD in real-world scenarios. Furthermore, we demonstrated that applying fine-tuning to D-NetPAD offers significant improvement in robustness against attacks while maintaining PAD performance. Our research paves the way for the development of more secure and robust iris recognition systems that are better equipped to counter sophisticated presentation attacks.

References

[1] Akshay Agarwal, Afzel Noore, Mayank Vatsa, and Richa Singh. Generalized contact lens iris presentation attack detection. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 4(3):373–385, 2022. 2, 3

[2] Anish Athalye, Logan Engstrom, Andrew Ilyas, and Kevin Kwok. Synthesizing robust adversarial examples. In *Proceedings of the 35th International Conference on Machine Learning*, pages 284–293. PMLR, 2018. 2, 3, 5

[3] Aidan Boyd, Jeremy Speth, Lucas Parzianello, Kevin W. Bowyer, and Adam Czajka. Comprehensive study in open-set iris presentation attack detection. *IEEE Transactions on Information Forensics and Security*, 18:3238–3250, 2023. 2

[4] Tom B Brown, Dandelion Man'e, Aurko Roy, Mart'ın Abadi, and Justin Gilmer. Adversarial patch. *arXiv preprint arXiv:1712.09665*, 2017. 2, 3

[5] Nicholas Carlini and David Wagner. Adversarial examples are not easily detected: Bypassing ten detection methods. In *Proceedings of the 10th ACM workshop on artificial intelligence and security*, pages 3–14, 2017. 3

[6] Nicholas Carlini and David Wagner. Towards evaluating the robustness of neural networks. In *2017 ieee symposium on security and privacy (sp)*, pages 39–57. IEEE, 2017. 2

[7] Cunjian Chen and Arun Ross. A multi-task convolutional neural network for joint iris detection and presentation attack detection. In *2018 IEEE Winter Applications of Computer Vision Workshops (WACVW)*, pages 44–51, 2018. 2, 3

[8] Cunjian Chen and Arun Ross. An explainable attention-guided iris presentation attack detector. In *2021 IEEE Winter Conference on Applications of Computer Vision Workshops (WACVW)*, pages 97–106, 2021. 2, 3

[9] Tianlong Chen, Sijia Liu, Shiyu Chang, Yu Cheng, Lisa Amini, and Zhangyang Wang. Adversarial robustness: From self-supervised pre-training to fine-tuning. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 696–705, 2020. 3

[10] Yifeng Chen, Cheng Wu, and Yiming Wang. T-Center: A novel feature extraction approach towards large-scale iris recognition. *IEEE Access*, 8:32365–32375, 2020. 2, 6

[11] Jeremy Cohen, Elan Rosenfeld, and Zico Kolter. Certified adversarial robustness via randomized smoothing. In *International Conference on Machine Learning*, pages 1310–1320. PMLR, 2019. 3

[12] Adam Czajka and Kevin W. Bowyer. Presentation attack detection for iris recognition: An assessment of the state-of-the-art. *ACM Comput. Surv.*, 51(4), 2018. 1, 2, 3

[13] Priyanka Das, Joseph Mcfiratht, Zhaoyuan Fang, Aidan Boyd, Ganghee Jang, Amir Mohammadi, Sandip Purnapatra, David Yambay, Sébastien Marcel, Mateusz Trokielewicz, Piotr Maciejewicz, Kevin Bowyer, Adam Czajka, Stephanie Schuckers, Juan Tapia, Sebastian Gonzalez, Meiling Fang, Naser Damer, Fadi Boutros, Arian Kuijper, Renu Sharma, Cunjian Chen, and Arun Ross. Iris Liveness Detection Competition (LivDet-Iris) - The 2020 Edition. In *2020 IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–9, 2020. 1, 2, 3

[14] John Daugman. Demodulation by complex-valued wavelets for stochastic pattern recognition. *International Journal of Wavelets, Multiresolution and Information Processing*, 1 (01):1–17, 2003. 2, 3

[15] Prithviraj Dhar, Amit Kumar, Kirsten Kaplan, Khushi Gupta, Rakesh Ranjan, and Rama Chellappa. EyePAD++: A distillation-based approach for joint eye authentication and presentation attack detection using periocular images. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 20186–20195, 2022. 2, 3

[16] Yinpeng Dong, Hang Su, Baoyuan Wu, Zhifeng Li, Wei Liu, Tong Zhang, and Jun Zhu. Efficient decision-based black-box adversarial attacks on face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7714–7722, 2019. 3

[17] James S. Doyle and Kevin W. Bowyer. Robust detection of textured contact lenses in iris recognition using BSIF. *IEEE Access*, 3:1672–1683, 2015. 2, 3

[18] J. Daugman et al. How iris recognition works. In *The essential guide to image processing*, pages 715–739. Elsevier, 2009. 1, 2, 4

[19] Abhishek Gangwar and Akanksha Joshi. DeepIrisNet: Deep iris representation with applications in iris recognition and cross-sensor iris recognition. In *2016 IEEE International Conference on Image Processing (ICIP)*, pages 2301–2305, 2016. 2

[20] T. Gittings, S. Schneider, and J. Collomosse. Vax-a-net: Training-time defence against adversarial patch attacks. In *Computer Vision – ACCV 2020: 15th Asian Conference on Computer Vision, Kyoto, Japan, November 30 – December 4, 2020, Revised Selected Papers, Part IV*, page 235–251, Berlin, Heidelberg, 2020. Springer-Verlag. 3

[21] Abhiram Gnanasambandam, Alex M. Sherman, and Stanley H. Chan. Optical adversarial attack. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, pages 92–101, 2021. 2, 3

[22] Ian Goodfellow, Jonathon Shlens, and Christian Szegedy. Explaining and harnessing adversarial examples. In *International Conference on Learning Representations*, 2015. 2, 3, 4

[23] Diego Gragnaniello, Giovanni Poggi, Carlo Sansone, and Luisa Verdoliva. An investigation of local descriptors for biometric spoofing detection. *IEEE Transactions on Information Forensics and Security*, 10(4):849–863, 2015. 2, 3

[24] Kathrin Grosse, Praveen Manoharan, Nicolas Papernot, Michael Backes, and Patrick McDaniel. On the (statistical) detection of adversarial examples, 2017. 3

[25] Andrej Hafner, Peter Peer, Žiga Emeršič, and Matej Vitek. Deep iris feature extraction. In *2021 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, pages 258–262, 2021. 2

[26] Lingxiao He, Haiqing Li, Fei Liu, Nianfeng Liu, Zhenan Sun, and Zhaofeng He. Multi-patch convolution neural network for iris liveness detection. In *2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pages 1–7, 2016. 2, 3

[27] Steven Hoffman, Renu Sharma, and Aran Ross. Convolutional neural networks for iris presentation attack detection: Toward cross-dataset and cross-sensor generalization. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1701–17018, 2018. 2, 3

[28] Bingyao Huang and Haibin Ling. Spaa: Stealthy projector-based adversarial attacks on deep image classifiers. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 534–542, 2022. 2, 3

[29] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q. Weinberger. Densely connected convolutional networks. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2261–2269, 2017. 2, 3

[30] ISO/IEC JTC 1/SC 37 Biometrics Technical Committee. ISO - ISO/IEC 39794-6:2021 - Information technology - Extensible biometric data interchange formats - Part 6: Iris image data. International Organization for Standardization, Geneva, CH, 2021. 2

[31] Max Jaderberg, Karen Simonyan, Andrew Zisserman, and koray kavukcuoglu. Spatial transformer networks. In *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2015. 4, 5

[32] Siamul Karim Khan, Patrick Tinsley, and Adam Czajka. Deformirisnet: An identity-preserving model of iris texture deformation. In *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 900–908, 2023. 4

[33] Naman Kohli, Daksha Yadav, Mayank Vatsa, Richa Singh, and Afzel Noore. Synthetic iris presentation attack using idcgan. In *2017 IEEE International Joint Conference on Biometrics (IJCB)*, pages 674–680, 2017. 3

[34] Alex Kurakin, Dan Boneh, Florian Tramèr, Ian Goodfellow, Nicolas Papernot, and Patrick McDaniel. Ensemble adversarial training: Attacks and defenses. 2018. 3

[35] Aishan Liu, Xianglong Liu, Jiaxin Fan, Yuqing Ma, Anlan Zhang, Huiyuan Xie, and Dacheng Tao. Perceptual-sensitive gan for generating adversarial patches. AAAI Press, 2019. 3

[36] Aishan Liu, Jiakai Wang, Xianglong Liu, Bowen Cao, Chongzhi Zhang, and Hang Yu. Bias-based universal adversarial patch attack for automatic check-out. In *Computer Vision – ECCV 2020*, pages 395–410. Springer International Publishing, 2020. 2, 3

[37] Ziquan Liu, Yi Xu, Xiangyang Ji, and Antoni B. Chan. Twins: A fine-tuning framework for improved transferability of adversarial robustness and generalization. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 16436–16446, 2023. 3

[38] Aleksander Madry, Aleksandar Makelov, Ludwig Schmidt, Dimitris Tsipras, and Adrian Vladu. Towards deep learning models resistant to adversarial attacks. *In International Conference on Learning Representations (ICLR)*, 2018. 2, 3, 4, 5, 7

[39] Muzammal Naseer, Salman H Khan, Shafin Rahman, and Fatih Porikli. Task-generalizable adversarial attack based on perceptual metric. *arXiv preprint arXiv:1811.09020*, 2018. 4

[40] Muzammal Naseer, Salman Khan, and Fatih Porikli. Local gradients smoothing: Defense against localized adversarial attacks. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1300–1307, 2019. 3

[41] Kien Nguyen, Clinton Fookes, Arun Ross, and Sridha Sridharan. Iris recognition with off-the-shelf cnn features: A deep learning perspective. *IEEE Access*, 6:18848–18855, 2018. 2

[42] Kien Nguyen, Clinton Fookes, Sridha Sridharan, and Arun Ross. Complex-valued iris recognition network. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1):182–196, 2023. 2

[43] Nadia Othman, Bernadette Dorizzi, and Sonia Garcia-Salicetti. OSIRIS: An open source iris recognition software. *Pattern Recognition Letters*, 82:124–131, 2016. 2, 4, 6

[44] Federico Pala and Bir Bhanu. Iris liveness detection by relative distance comparisons. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 664–671, 2017. 2, 3

[45] P. Jonathon Phillips, W. Todd Scruggs, Alice J. O'Toole, Patrick J. Flynn, Kevin W. Bowyer, Cathy L. Schott, and Matthew Sharpe. FRVT 2006 and ICE 2006 large-scale experimental results. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(5):831–846, 2010. 6

[46] P. Jonathon Phillips, W. Todd Scruggs, Alice J. O'Toole, Patrick J. Flynn, Kevin W. Bowyer, Cathy L. Schott, and Matthew Sharpe. The ND-IRIS-0405 iris image dataset. *arXiv preprint arXiv:1606.04853*, 2016. 6

[47] R. Raghavendra and Christoph Busch. Robust scheme for iris presentation attack detection using multiscale binarized statistical image features. *IEEE Transactions on Information Forensics and Security*, 10(4):703–715, 2015. 2, 3

[48] Sukrut Rao, David Stutz, and Bernt Schiele. Adversarial training against location-optimized adversarial patches. In *Computer Vision – ECCV 2020 Workshops*, pages 429–448. Springer International Publishing, 2020. 3

[49] Christian Rathgeb, Andreas Uhl, Peter Wild, and Heinz Hofbauer. Design decisions for an iris recognition sdk. *Handbook of iris recognition*, pages 359–396, 2016. 2, 6

[50] Min Ren, Yunlong Wang, Yuhao Zhu, Kunbo Zhang, and Zhenan Sun. Multiscale dynamic graph representation for biometric recognition with occlusions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(12): 15120–15136, 2023. 2

[51] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing. 3, 6

[52] Andras Rozsa, Manuel Günther, and Terranee E. Boult. Lots about attacking deep features. In *2017 IEEE International Joint Conference on Biometrics (IJCB)*, pages 168–176, 2017. 4

[53] Sara Sabour, Yanshuai Cao, Fartash Faghri, and David Fleet, J. Adversarial manipulation of deep representations. In *International Conference on Learning Representations (ICLR)*, 2016. 4

[54] Athena Sayles, Ashish Hooda, Mohit Gupta, Rahul Chatterjee, and Earlence Fernandes. Invisible perturbations: Physical adversarial examples exploiting the rolling shutter effect. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14666–14675, 2021. 2, 3

[55] Mahmood Sharif, Sruti Bhagavatula, Lujo Bauer, and Michael K Reiter. A general framework for adversarial examples with objectives. *ACM Transactions on Privacy and Security (TOPS)*, 22(3):1–30, 2019. 3

[56] Renu Sharma and Arun Ross. D-NetPAD: An explainable and interpretable iris presentation attack detector. In *International Joint Conference on Biometrics (IJCB)*, pages 1–10, 2020. 2, 3, 6, 7

[57] Renu Sharma and Arun Ross. Image-level iris morph attack. In *2021 IEEE International Conference on Image Processing (ICIP)*, pages 3013–3017, 2021. 1, 2, 6, 7

[58] Sobhan Soleymani, Ali Dabouei, Jeremy Dawson, and Nasser M Nasrabadi. Adversarial examples to fool iris recognition systems. In *2019 International Conference on Biometrics (ICB)*, pages 1–8. IEEE, 2019. 1, 2, 3, 6, 7

[59] Zhenan Sun, Hui Zhang, Tieniu Tan, and Jianyu Wang. Iris image classification based on hierarchical visual codebook. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(6):1120–1133, 2014. 2, 3

[60] Juan E. Tapia, Sebastian Gonzalez, and Christoph Busch. Iris liveness detection using a cascade of dedicated deep learning networks. *IEEE Transactions on Information Forensics and Security*, 17:42–52, 2022. 2, 3

[61] Simen Thys, Wiebe Van Ranst, and Toon Goedemé. Fooling automated surveillance cameras: Adversarial patches to attack person detection. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 49–55, 2019. 3

[62] Takahiro Toizumi, Koichi Takahashi, and Masato Tsukada. Segmentation-free direct iris localization networks. In *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 991–1000, 2023. 6, 7

[63] Kuo Wang and Ajay Kumar. Toward more accurate iris recognition using dilated residual features. *IEEE Transactions on Information Forensics and Security*, 14(12):3233–3245, 2019. 2

[64] Zhibo Wang, Siyan Zheng, Mengkai Song, Qian Wang, Alireza Rahimpour, and Hairong Qi. advpattern: Physical-world attacks on deep person re-identification via adversarially transformable patterns. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 8340–8349, 2019. 2, 3

[65] Koichiro Yamanaka, Ryutaroh Matsumoto, Keita Takahashi, and Toshiaki Fujii. Adversarial patch attacks on monocular depth estimation networks. *IEEE Access*, 8:179094–179104, 2020. 3

[66] David Yambay, James S. Doyle, Kevin W. Bowyer, Adam Czajka, and Stephanie Schuckers. LivDet-iris 2013 - Iris Liveness Detection Competition 2013. In *IEEE International Joint Conference on Biometrics*, pages 1–8, 2014. 3

[67] David Yambay, Benedict Becker, Naman Kohli, Daksha Yadav, Adam Czajka, Kevin W. Bowyer, Stephanie Schuckers, Richa Singh, Mayank Vatsa, Afzel Noore, Diego Gragnaniello, Carlo Sansone, Luisa Verdoliva, Lingxiao He, Yiwei Ru, Haiqing Li, Nianfeng Liu, Zhenan Sun, and Tieniu Tan. LivDet iris 2017 — Iris liveness detection competition 2017. In *2017 IEEE International Joint Conference on Biometrics (IJCB)*, pages 733–741, 2017.

[68] David Yambay, Brian Walczak, Stephanie Schuckers, and Adam Czajka. LivDet-Iris 2015 - Iris Liveness Detection Competition 2015. In *2017 IEEE International Conference on Identity, Security and Behavior Analysis (ISBA)*, pages 1–6, 2017. 3

[69] Zijing Zhao and Ajay Kumar. Towards more accurate iris recognition using deeply learned spatially corresponding features. In *International Conference on Computer Vision (ICCV)*, pages 3829–3838, 2017. 2