Deep Learning for Iris Recognition: A Survey

KIEN NGUYEN, Queensland University of Technology, Brisbane, Australia
HUGO PROENÇA, University of Beira Interior, IT: Instituto de Telecomunicações, Covilhã, Portugal
FERNANDO ALONSO-FERNANDEZ, Halmstad University, Halmstad, Sweden

In this survey, we provide a comprehensive review of more than 200 articles, technical reports, and GitHub repositories published over the last 10 years on the recent developments of deep learning techniques for iris recognition, covering broad topics on algorithm designs, open-source tools, open challenges, and emerging research. First, we conduct a comprehensive analysis of deep learning techniques developed for two main sub-tasks in iris biometrics: segmentation and recognition. Second, we focus on deep learning techniques for the robustness of iris recognition systems against presentation attacks and via human-machine pairing. Third, we delve deep into deep learning techniques for forensic application, especially in post-mortem iris recognition. Fourth, we review open-source resources and tools in deep learning techniques for iris recognition. Finally, we highlight the technical challenges, emerging research trends, and outlook for the future of deep learning in iris recognition.

CCS Concepts: • **Computing methodologies** → *Artificial intelligence*; *Computer vision*; *Computer vision tasks*; *Biometrics*;

Additional Key Words and Phrases: Iris recognition, deep learning, neural networks

1 INTRODUCTION

The human iris is a sight organ that controls the amount of light reaching the retina, by changing the size of the pupil. The texture of the iris is fully developed before birth, its minutiae do not depend on genotype, it stays relatively stable across a lifetime (except for disease- and normal aging-related biological changes), and it may even be used for forensic identification shortly after subject's death [36, 110, 170].

In terms of its information theory-related properties, the iris texture has an extremely high randotypic randomness, and is stable (permanent) over time, providing an exceptionally high entropy per mm.^{2} that justifies its higher discriminating power, when compared to other biometric modalities (e.g., face or fingerprint). The iris's collectability is another feature of interest and has

been the subject of discussion over the last decade: while it can be acquired using **commercial off-the-shelf** (**COTS**) hardware, either handheld or stationary, data can be even collected from at-a-distance, up to tens of meters away from the subjects [111]. Even though commercial visible-light (RGB) cameras can image the iris, **near infrared** (**NIR**)-based sensing dominates in most applications, due to better visibility of iris texture for darker eyes, rich in melanin pigment, which is characterized by lower light absorption in NIR spectrum compared to shorter wavelengths. In addition, NIR wavelengths are barely perceivable by the human eye, which augments users' comfort and avoids pupil contraction/dilation that would appear under visible light.

A seminal work by John Daugman brought to the community the Gabor filtering-based approach that became the dominant approach for iris recognition [34, 35, 37]. Even though subsequent solutions to iris image encoding and matching appeared, the IrisCodes approach is still dominant due to its ability to effectively search in massive databases with a minimal probability of false matches, at extreme time performance. By considering binary words, pairs of signatures are matched using XOR parallel-bit logic at lightning speed, enabling millions of comparisons/second per processing core. Also, most of the methods that outperformed the original techniques in terms of effectiveness do not work under the *one-shot learning* paradigm, assume multiple observations of each class to obtain appropriate decision boundaries, and—most importantly—have encoding/matching steps with time complexity that forbid their use in large environments (in particular, for *all-against-all* settings).

In short, Daugman's algorithm encodes the iris image into a binary sequence of 2,048 bits by filtering the iris image with a family of Gabor kernels. The varying pupil size is rectified by the Cartesian-to-polar coordinate system transformation, to end up with an image representation of canonical size, guaranteeing identical structure of the iris code independently of the iris and pupil size. This makes it possible to use the **Hamming Distance** (**HD**) to measure the similarity between two iris codes [37]. Its low false match rate at acceptable false non-match rates is the key factor behind the success of global-scale iris recognition installments, such as the national person identification and border security program Aadhaar program in India (with over 1.2 billion pairs of irises enrolled) [174], the **Homeland Advanced Recognition Technology** (**HART**) in the US (up to 500 million identities) [128], or the NEXUS system, designed to speed up border crossings for low-risk and pre-approved travelers moving between Canada and the US.

Deep learning-based methods, in particular using various **Convolutional Neural Network** (**CNN**) architectures, have been driving remarkable improvements in many computer vision applications over the last decade. In terms of biometrics technologies, it's not surprising that iris recognition has also seen an increasing adoption of purely data-driven approaches at all stages of the recognition pipeline: from preprocessing (such as off-axis gaze correction), segmentation, and encoding to matching. Interestingly, however, the impact of deep learning on the various stages of the iris recognition pipeline is uneven. One of the primary goals of this survey article is to assess where deep learning helped in achieving high performance and more secure systems, and which procedures did not benefit from more complex modeling.

The remainder of the article is structured as follows. Sections 2 and 3 review the application of deep learning in two main stages of the recognition pipeline: segmentation and recognition (encoding and comparison). Sections 4 and 5 analyze the state-of-the-art deep learning-based approaches in two applications: **Presentation Attack Detection** (**PAD**) and Forensic. Section 6 investigates how human and machine can pair to improve deep learning-based iris recognition. Section 7 focuses on approaches in less controlled environments of iris and periocular analysis. Section 8 reviews public resources and tools available in the deep learning-based iris recognition domain. Section 9 focuses on the future of deep learning for iris recognition with a discussion on emerging research directions in different aspects of iris analysis. The article is concluded in Section 10.
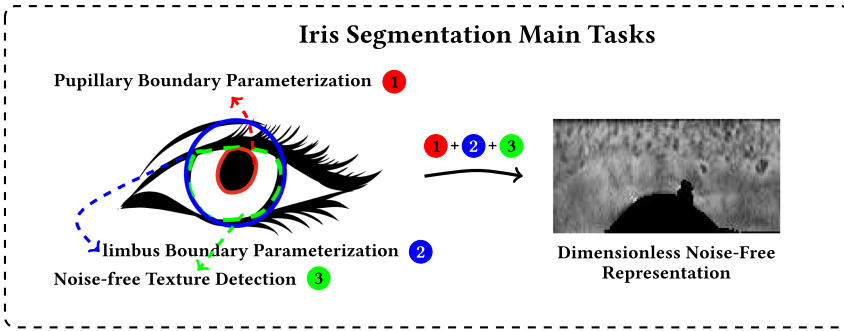
Fig. 1. Three main tasks typically associated to *iris segmentation*: (1) parameterization of the pupillary (inner) boundary; (2) parameterization of the limbus (outer) boundary; and (3) discrimination between the unoccluded (noise-free) and occluded (noisy) regions inside the iris ring. Such pieces of information are further used to obtain dimensionless polar representations of the iris texture, where feature extraction methods typically operate.

2 DEEP LEARNING-BASED IRIS SEGMENTATION

The segmentation of the iris is seen as an extremely challenging problem. As illustrated in Figure 1, segmenting the iris involves essentially three tasks: detect and parameterize the inner (pupillary) and outer (limbus) biological boundaries of the iris and also to locally discriminate between the noise-free/noisy regions inside the iris ring, which should be subsequently used in the feature encoding and matching processes.

This problem has motivated numerous research works for decades. From the pioneering integro-differential operator [34] up to subsequent handcrafted techniques based on active contours and neural networks (e.g., [68, 133, 147] and [175]) a long road has been traveled in this problem. Regardless of an obvious evolution in the effectiveness of such techniques, they all face particular difficulties in the case of heavily degraded data. Images are frequently motion-blurred, poorly focused, partially occluded, and off-angle. Additionally, in the case of visible light data, severe reflections from the environments surrounding the subjects are visible and even augment the difficulties of the segmentation task.

Recently, as in many other computer vision tasks, DL-based frameworks have been advocated as providing consistent advances over the state-of-the-art for the iris segmentation problem, with numerous models being proposed. A cohesive perspective of the most relevant recent DL-based methods is given in Table 1, with the techniques appearing in chronographic (and then alphabetical) order. The type of data each model aims at handling is given in the "*Data*" column, along with the datasets where the corresponding experiments were carried out and a summary of the main characteristics of each proposal ("*Features*" column). Here, considering that models were empirically validated in completely heterogeneous ways and using very different metrics, we decided not to include the summary performance of each model/solution.

Schlett et al. [144] provided a multi-spectral analysis to improve iris segmentation accuracy in visible wavelengths by preprocessing data before the actual segmentation phase, extracting multiple spectral components in the form of RGB color channels. Even though this approach does propose a DL-based framework, the different versions of the input could be easily used to feed DL-based models and augment the robustness to non-ideal data. Chen et al. [22] used CNNs that include dense blocks, referred to as a **dense-fully convolutional network** (**DFCN**), where the encoder part consists of dense blocks, and the decoder counterpart obtains the segmentation masks via transpose convolutions. Hofbauer et al. [72] parameterize the iris boundaries based on

Table 1. Cohesive Comparison of the Most Relevant DL-based Iris Segmentation Methods (NIR: *near-infrared*; VW: *visible wavelength*)

Method	Year	Data		Datasets	Features
		NIR	VW		
Schlett et al. [144]	2018	✗	✓	MobBIO	Preprocessing (combines different possibilities of the input RGB channels)
Trokielewicz and Czajka [166]	2018	✓	✓	Warsaw-Post-Mortem v1.0	Fine-tuned CNN (SegNet)
Chen et al. [22]	2019	✓	✓	CASIA-Irisv4-Interval, IITD, UBIRIS.v2	Dense CNN
Hofbauer et al. [72]	2019	✓	✗	IITD, CASIA-Irisv4-Interval, ND-Iris-0405	Cascaded architecture of four RefineNet, each connecting to one Residual net
Huynh et al. [76]	2019	✓	✗	OpenEDS	MobileNetV2 + heuristic filtering postproc.
Li et al. [7]	2019	✓	✗	CASIA-Iris-Thousand	Faster-R-CNN (ROI detection)
Kerrigan et al. [85]	2019	✓	✓	CASIA-Irisv4-Interval, BioSec, ND-Iris-0405, UBIRIS.v2, Warsaw-Post-Mortem v2.0, ND-TWINS-2009-2010	Resent + Segnet (with dilated convolutions)
Wu and Zhao [186]	2019	✓	✓	CASIA-Irisv4-Interval, UBIRIS.v2	Dense-U-Net (dense layers + U-Net)
Wei et al. [205]	2019	✓	✓	CASIA-Iris4-Interval, ND-IRIS-0405, UBIRIS.v2	U-Net with dilated convolutions
Fang and Czajka [50]	2020	✓	✓	ND-Iris-0405, CASIA, BATH, BioSec, UBIRIS, Warsaw-Post-Mortem v1.0 & v2.0	Fine-tuned CC-Net [106]
Ganeva and Myasnikov [55]	2020	✓	✗	MMU	U-Net, LinkNet, and FC-DenseNet (performance comparison)
Jalilian et al. [79]	2020	✓	✗		RefineNet + morphological postprocessing
Sardar et al. [142]	2020	✓	✓	CASIA-Irisv4-Interval, IITD, UBIRIS.v2	Squeeze-Expand module + active learning (interactive segmentation)
Trokielewicz et al. [172]	2020	✓	✓	ND-Iris-0405, CASIA, BATH, BioSec, UBIRIS, Warsaw-Post-Mortem v1.0 & v2.0	Fined-tuned SegNet [9]
Wang et al [178]	2020	✓	✓	CASIA-Iris-M1-S1/S2/S3, MICHE-I	Hourglass network
Wang et al. [176]	2020	✓	✓	CASIA-v4-Distance, UBIRIS.v2, MICHE-I	U-Net + multi-task attention net + postproc. (probabilistic masks priors & thresholding)
Li et al. [94]	2021	✓	✗	CASIA-Iris-Thousand	IRU-Net network
Wang et al. [184]	2021	✗	✓	Online Video Streams and Internet Videos	U-Net and Squeezenet to iris segmentation and detect eye closure
Kuehlkamp et al. [91]	2022	✓	✓	ND-Iris-0405, CASIA, BATH, BioSec, UBIRIS, Warsaw-Post-Mortem v2.0	Fined-tuning of Mask-RCNN architecture

Methods are listed in chronological (and then alphabetical) order.

segmentation maps yielding from a CNN, using a cascaded architecture with four RefineNet units, each directly connecting to one Residual net. Huynh et al. [76] discriminate between three distinct eye regions with a DL model and remove incorrect areas with heuristic filters. The proposed architecture is based on the encoder-decoder model, with depth-wise convolutions used to reduce the computational cost. Roughly at the same time, Li et al. [94] described the *Interleaved Residual U-Net* model for semantic segmentation and iris mask synthesis. In this work, unsupervised techniques (K-means clustering) were used to create intermediary pictorial representations of the ocular region, from where saliency points deemed to belong to the iris boundaries were found. Kerrigan et al. [85] assessed the performance of four different convolutional architectures designed for semantic segmentation. Two of these models were based on dilated convolutions, as proposed by Yu and Koltun [188]. Wu and Zhao [186] described the Dense U-Net model, which combines dense layers to the U-Net network. The idea is to take advantage of the reduced set of parameters

of the dense U-Net, while keeping the semantic segmentation capabilities of U-Net. The proposed model integrates dense connectivity into U-Net contraction and expansion paths. Compared with traditional CNNs, this model is claimed to reduce learning redundancy and enhance information flow, while keeping controlled the number of parameters of the model. Wei et al. [205] suggested to perform *dilated convolutions*, which is claimed to obtain more consistent global features. In this setting, convolutional kernels are not continuous, with zero values being artificially inserted between each non-zero position, increasing the receptive field without augmenting the number of parameters of the model.

More recently, Ganeva and Myasnikov [55] compared the effectiveness of three CNN architectures (U-Net, LinkNet, and FC- DenseNet), determining the optimal parameterization for each one. Jalilian et al. [79] introduced a scheme to compensate for texture deformations caused by the off-angle distortions, re-projecting the off-angle images back to the frontal view. The used architecture is a variant of RefineNet [96], which provides high-resolution prediction while preserving the boundary information (required for parameterization purposes).

The idea of interactive learning for iris segmentation was suggested by Sardar et al. [142], describing an interactive variant of U-Net that includes Squeeze Expand modules. Trokielewicz et al. [172] used DL-based iris segmentation models to extract highly irregular iris texture areas in post-mortem iris images. They used a pre-trained SegNet model, fine-tuned with a database of cadaver iris images. Wang et al. [178] (further extended in [179]) described a lightweight **deep convolutional neural network** (**DCNN**) specifically designed for iris segmentation of degraded images acquired by handheld devices. The proposed approach jointly obtains the segmentation mask and parameterized pupillary/limbic boundaries of the iris.

Observing that edge-based information is extremely sensitive to be obtained in degraded data, Li et al. [7] presented a hybrid method that combines edge-based information with deep learning frameworks. A compacted Faster R-CNN-like architecture was used to roughly detect the eye and define the initial region of interest, from where the pupil is further located using a Gaussian mixture model. Wang et al. [184] trained a DCNN that automatically extracts the iris and pupil pixels of each eye from input images. This work combines the power of U-Net and SqueezeNet to obtain a compact CNN suitable for real-time mobile applications. Finally, Wang et al. [176] parameterize both the iris mask and the inner/outer iris boundaries jointly, by actively modeling such information into a unified multi-task network.

A final word is given to *segmentation-less* techniques. Assuming that the accurate segmentation of the iris boundaries is one of the hardest phases of the whole recognition chain and the main source for recognition errors, some recent works have been proposing to perform biometrics recognition in non-segmented or roughly segmented data [132, 135]. Here, the idea is to use the remarkable discriminating power of DL frameworks to perceive the agreeing patterns between pairs of images, even on such *segmentation-less* representations.

3 DEEP LEARNING-BASED IRIS RECOGNITION

3.1 Deep Learning Models as a Feature Extractor

As illustrated in Figure 2, the idea here is to analyze a dimensionless representation of the iris data and produce a feature vector that lies in a hyperspace (embedding) where recognition is carried out.

In this context, Boyd et al. [15] explored five different sets of weights for the popular ResNet50 architecture to test if iris-specific feature extractors perform better than models trained for general tasks. Minaee et al. [105] studied the application of deep features extracted from VGG-Net
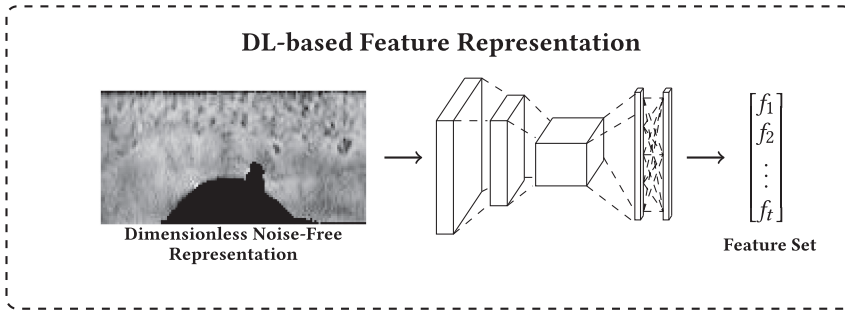
Fig. 2. The main task of DL-based iris feature extraction: given a dimensionless representation of the iris data, obtain its compact and representative representation—the feature set—that is further used in the classification phase.

for iris recognition, having authors observed that the resulting features can be well transferred to biometric recognition. Luo et al. [102] described a DL model with spatial attention and channel attention mechanisms, that are directly inserted into the feature extraction module. Also, a co-attention mechanism adaptively fuses features to obtain representative iris-periocular features. Hafner et al. [65] adapted the classical Daugman's pipeline, using CNNs to function as feature extractors. The DenseNet-201 architecture outperformed its competitors achieving state-of-the-art results both in the open and close world settings. Menotti et al. [104] assessed how DL-based feature representations can be used in spoofing detection, observing that spoofing detection systems based on CNNs can be robust to attacks already known and adapted, with little effort, to image-based attacks that are yet to come.

Yang et al. [196] generated multi-level spatially corresponding feature representations by an encoder-decoder structure. Also, a spatial attention feature fusion module was used to ensemble the resulting features more effectively. Chen et al. [23] addressed the large-scale recognition problem and described an optimized center loss function (tight center) to attenuate the insufficient discriminating power of the cross-entropy function. Nguyen et al. [112] explored the performance of state-of-the-art pre-trained CNNs on iris recognition, concluding that off-the-shelf CNN generic features are also extremely good at representing iris images, effectively extracting discriminative visual features and achieving promising results. Zhao et al. [207] proposed a method based on the capsule network architecture, where a modified routing algorithm based on the dynamic routing between two capsule layers was described, with three pre-trained models (VGG16, InceptionV3, and ResNet50) extracting the primary iris features. Next, a convolution capsule replaces the full connection capsule to reduce the number of parameters. Wang and Kumar [180] introduced the concept of *residual feature* for iris recognition. They described a residual network learning procedure with offline triplets selection and dilated convolutional kernels.

Other works have addressed the extraction of appropriate feature representations in multi-biometrics settings: Damer et al. [32] propose to jointly extract multi-biometric representations within a single DNN. Unlike previous solutions that create independent representations from each biometric modality, they create these representations from multi-modality (face and iris), multi-instance (iris left and right), and multi-presentation (two face samples), which can be seen as a fusion at the data level policy. Finally, concerned about the difficulty of performing reliable recognition in hand-held devices, Odinokikh et al. [121] combined the advantages of handcrafted feature extractors and advanced deep learning techniques. The model utilizes shallow and deep feature representations in combination with characteristics describing the environment, to reduce the intra-subject variations expected in this kind of environment.

Table 2. Cohesive Comparison of the Most Relevant DL-based Iris Recognition Methods (NIR: *near-infrared*; VW: *visible Wavelength*)

Category	Method	Year	Data		Datasets	Features
			NIR	VW		
Conventional classifiers	Menotti et at. [104]	2015	✓	✓	Biosec, LivDet-2013-Warsaw, MobBIOfake	Shallow CNNs + SVM for Spoofing Detection
	Minaee et al. [105]	2016	✓	✗	CASIA-Iris-Thousand, IITD	VGG + SVM
	Nguyen et al. [112]	2017	✓	✗	ND-CrossSensor-2013, CASIA-Iris-Thousand	AlexNet, VGG, Google Inception, ResNet, DenseNet + SVM
	Boyd et al. [15]	2019	✓	✓	CASIA-Irisv4-Interval, IITD, UBIRIS.v2	ResNet50 + SVM
	Boyd et al. [18]	2020	✓	✓	DCMEO1, Warsaw	AlexNet, ResNet, VGG, DenseNet + Cosine, Euclidean, MSE
	Hafner et al. [65]	2021	✓	✗	CASIA-Iris-Thousand	ResNet101 + DenseNet-201 + Cosine Similarity
Classification Networks	Gangwar et al. [56]	2016	✓	✗	ND-IRIS-0405, ND-CrossSensor-2013	DeepIrisNet
	Gangwar et al. [57]	2019	✓	✓	ND-IRIS-0405, UBIRIS.v2, MICHE-I, CASIA-Irisv4-Interval	DeepIrisNetV2
	Odinokikh et al. [121]	2019	✓	✗	CASIA-Iris-M1-S2, CASIA-Iris-M1-S3, Iris-Mobile	Feature Fusion + Softmax
	Zhao et al.[207]	2019	✓	✗	JluIrisV3.1, JluIrisV4, CASIA-Irisv4-Lamp	Capsule network + Softmax
	Chen et al. [23]	2020	✓	✗	ND-IRIS-0405, CASIA-Iris-Thousand, IITD cross sensor	T-Center loss
	Luo et al.[102]	2021	✓	✗	ND-IRIS-0405, CASIA-Iris-Thousand	Attention + Softmax Loss + Center Loss
Similarity Networks	Nianfeng et al. [97]	2016	✓	✗	Q-FIRE, CASIA-Cross-Sensor	DeepIris
	Zhao et al. [209]	2017	✓	✓	CASIA-Irisv4-Interval, IITD, UBIRIS.v2	UniNet (FeatNet+MaskNet) + Extended Triplet Loss
	Damer et al. [32]	2019	✓	✗	Biosecure, CASIA-Iris-Thousand/Lamp/Interval	Inception + Triplet Loss
	Wang et al. [180]	2019	✓	✓	CASIA-Irisv4-Interval, IITD, UBIRIS.v2	FeatNet + Dilated Convolution + Extended Triplet Loss
	Zhao et al. [211]	2019	✓	✗	ND-Iris-0405, Casia-Irisv4-Distance, IITD	FeatNet + Mask RCNN + Extended Triplet Loss
	Nguyen et al. [113]	2020	✓	✓	CASIA-v4-Distance, UBIRIS.v2, ND-CrossSensor-2013	Constrained Design Backbone + Extended Triplet Loss
	Yan et al. [195]	2021	✓	✗	CASIA-Iris-Thousand	Spatial Feature Reconstruction + Triplet Loss
	Yang et al. [196]	2021	✓	✗	CASIA-Irisv4-Thousand, CASIA-Irisv4-Distance, IITD	Dual Spatial Attention Network + Batch Hard
	Nguyen et al. [115]	2022	✓	✓	ND-CrossSensor-2013, CASIA-Iris-Thousand, UBIRIS.v2	Complex-valued Backbone + Extended Triplet Loss
	Kuehlkamp et al. [91]	2022	✓	✓	DCMEO1, DCMEO2, Warsaw-Post-Mortem v2.0	ResNet + Triplet Loss

Methods are listed in chronological (and then alphabetical) order.

3.2 Deep Learning-based Iris Matching Strategies

The existing matching strategies can be categorized into three categories: (1) using conventional classifiers, such as SVM, RF, and Sparse Representation; (2) softmax-based losses; and (3) pairwise-based losses. A cohesive perspective of the most relevant recent DL-based methods is given in Table 2, with the techniques appearing in a chronographic (and then alphabetically) order.

3.2.1 Conventional Classifiers. Various researchers have been using deep learning networks designed and pre-trained on the ImageNet dataset to extract iris feature representations, followed by a conventional classifier such as SVM, RF, Sparse Representation, and so on [15, 18, 112]. The key benefit of these approaches is the simplicity of "plug and play", where proven and pre-trained deep learning networks inherited from large-scale computer vision challenges are widely available and ready to be used [112]. Another benefit is that there is no need for large-scale iris image datasets to train these networks because they have already been trained on such large-scale datasets as ImageNet. Considering these networks usually contain hundreds of layers and millions of parameters and require millions of images to train, using pre-trained networks is extremely beneficial.

3.2.2   Iris Classification Networks. Iris classification networks couple deep learning architectures with a family of softmax-based losses to classify an iris image into a list of known identities. Coupling a softmax loss with a backbone network enables training the backbone network in an end-to-end manner via popular optimization strategies such as back-propagation and steepest gradient descent. Compared to the conventional classifier approaches, the DL-based backbones in this category are learnable directly from the iris data, allowing them to better represent the iris. The key benefit is that it is similar to a generic image classification task, hence all designs and algorithms in the generic image classification task can be trivially applied to the iris image data. Typical examples of these iris classification networks are [15, 56]. However, these softmax-based networks require the iris in the test image to be known in the identity classes in the training set, which means the networks must be re-trained whenever a new class (i.e., a new identity) is added. Gangwar et al. proposed two backbone networks (i.e., DeepIrisNet-A and DeepIrisNet-B) followed by a softmax loss for the iris recognition task [56]. Later, they proposed another backbone network but still followed by a softmax loss to classify one normalized iris image into a pre-defined list of identity [57].

Backbone Network Architectures: A wide range of backbone network architectures have been borrowed from generic image classification for the iris recognition task due to their similarity.

— AlexNet: AlexNet is the most primitive and has been shown as least accurate for iris recognition compared to others [16, 112].
— VGG: Boyd et al. [16], Nguyen et al. [112], and Minaee et al. [105] all experimented VGG16.
— ResNet: ResNet with its variants are the most popular backbone network architecture. Nguyen et al. experimented ResNet152 [112]. Boyd et al. experimented with three variants ResNet18, ResNet50, and ResNet152 in their post-mortem iris classification task [16].
— Inception: Zhao et al. employed a capsule network based on the InceptionV3 architecture [207].
— EfficientNet: Hsiao et al. [74] employed EfficientNet to extract iris features.

3.2.3   Iris Similarity Networks. Iris similarity networks couple deep learning architectures with a family of pairwise-based losses to learn a metric representing how similar or dissimilar two iris images are without knowing their identities. The pairwise loss aims at pulling images of the same iris closer and push images of different irises away in the similarity distance space. Different from the iris classification networks which only operate in an identification mode on a pre-defined identity list, iris similarity networks operate across both verification and identification modes with an open set of identities [209]. Typical examples of these iris similarity networks are [80, 97, 113, 180, 209]. There are three key benefits of these networks: (i) verification and identification: iris similarity networks operate across both verification and identification modes; (ii) open set of identities: iris similarity networks operate on an open set of identities; and (iii) explicit reflection: iris similarity networks directly and explicitly reflect what we want to achieve, i.e., small distances between irises of the same subject and larger distances between irises of different subjects.

Pairwise loss: Nianfeng et al. [97] proposed a pairwise network, which accepts two input images and directly outputs a similarity score. They designed a pairwise layer that accepts two input images and encodes their features via a backbone network. The backbone network is trained iteratively to minimize the dissimilarity distance between genuine pairs (pairs of the same identity) and maximize the dissimilarity distance between impostor pairs (pairs of different identities).

Triplet loss: Since the pairwise network is trained with separate genuine and impostor pairs, it may not converge well, which has been proven in face recognition [145]. Rather than using one

pair of two images to update the training as in the pairwise loss for each training iteration, the triplet loss employs a triplet of three images: an anchor image, a positive image with the same identity, and a negative image with a different identity [145]. The backbone network is trained to simultaneously minimize the similarity distance between the positive and the anchor images and maximize the distance between the negative and the anchor images. Tailored for iris images, Zhao et al. [180, 209, 211] proposed **Extended Triplet Loss** (**EPL**) to incorporate a bit-shifting operation to deal with rotation in the normalized iris images. Nguyen et al. also employed the ETL for their iris recognition network [113, 115]. Kuehlkamp et al. [91] proposed to improve the generic triplet loss function for iris recognition by forcing the distance to be positive (through the use of a sigmoid output layer), and adding a logarithmic penalty to the error. This modification allows the network to learn even when the difference between samples is negative and converges faster. Yan et al. [195] extended the generic triplet loss to batch triplet loss, in which the triplet loss is calculated over a batch of S subjects and K images for each subject. Performing batch triplet loss is usually expected to have a smooth loss function. Yang et al. [196] improved the triplet selection method for training by Batch Hard [197].

Backbone Network Architectures: Different from the classification iris networks, similarity iris networks are usually designed with their network architectures and are usually much "shallower" than the classification counterparts.

— FCN: All similarity iris networks employ **Fully Convolutional Networks** (**FCNs**) instead of CNNs. Compared to CNNs, FCNs [100] do not have a fully connected layer, allowing the output map to preserve the original spatial information. This is important to iris recognition since the output map can preserve spatial correspondence with the original input image [113, 209], thus enabling pixel-to-pixel matching. Zhao et al. [209] proposed an FCN architecture with 3 convolutional layers, followed by activation and pooling layers. Outputs of convolutional layers are up-sampled to the original input image size. The up-sample features are stacked and convolved by another convolutional layer to generate 2-dimension features with the same size as the input image. Later, they extended the backbone network with dilated convolutions [180]. Yan et al. [195] employed a ResNet architecture and fine-tuned it with the triplet loss. Kuehlkamp et al. only used a part of the ResNet architecture.
— NAS: Nguyen et al. [113] proposed to learn the network architecture directly from data rather than hand-designing it or using generic-image-classification architectures. They proposed a differential **Neural Architecture Search** (**NAS**) approach that models the architecture design process as a bi-level constrained optimization approach. This approach is not only able to search for the optimal network that achieves the best possible performance, but it can also impose constraints on resources such as model size or number of computational operations.
— Complex-valued: Observing that there is an intrinsic difference between the iris texture and generic object-based images where the iris texture is stochastic without consistent shapes, edges, or semantic structure, Nguyen et al. [115] argued the network architecture has to be better tailored to incorporate domain-specific knowledge to reach the full potential in the iris recognition setting. Another observation that they made is a majority of well-known handcrafted features such as IrisCode [35] transformed the iris texture image into a complex-valued representation first, then further encoded the complex-valued representation to get a final representation. They proposed to use fully complex-valued networks rather than popular real-valued networks. Complex-valued backbone networks better retain the phase, are more invariant to multi-scale, multi-resolution, and multi-orientation, have solid correspondence with the classic Gabor wavelets [173], hence are much better suited to iris recognition than their real-valued counterparts.

3.3 End-to-end Joint Iris Segmentation+Recognition Networks

Almost all existing approaches perform segmentation and normalization to transform an input image to a normalized rectangular 2D representation before recognition as this simplifies the representation learning. As segmentation and recognition may require a separate network themselves, this would cause redundancy in both computation and training, further slowing down a DL-based iris recognition approach. Several researchers have looked at approaches to perform end-to-end networks. One category is to perform segmentation-less recognition. Another category is to jointly learn segmentation and recognition using a unified network via multi-task learning.

Segmentation-less: These approaches feed the cropped iris images directly into a deep learning network to extract features. For example, Kuehlkamp et al. [91] used Mask R-CNN for semantic segmentation and fed the cropped iris region directly into a ResNet50 to extract features. Similarly, Chen et al. [24] also fed the cropped iris images directly into a DenseNet. Rather than feeding the cropped iris images directly, Proenca et al. transformed the cropped region (which is detected by SSD) into a polar representation first, then fed the polar representation into the VGG19 for extracting features [135].

Multi-task: Segmentation and recognition can be jointly learned with one unified network. This paves the way for multi-task learning. However, segmentation and recognition may require a different number of layers, hence research is required to perform using different intermediate layers for each task. To the best knowledge, there does not exist any approach to explore this direction.

4 DEEP LEARNING-BASED IRIS PRESENTATION ATTACK DETECTION

In parallel to the popularity of biometrics, the security of these systems against attacks has become of paramount importance. The most common attack is a **Presentation Attack** (**PA**), which refers to presenting a fake sample to the sensor. The goal can be either to *impersonate* somebody else identity (also known as *Impostor Attack Presentation*), or to *conceal* their own identity (also known as *Concealer Attack Presentation*). Via impostor attacks, a person could also enroll fraudulently, allowing a continuous manipulation of the system. The previous acronyms and terms in italics correspond to the vocabulary recommended in the series of ISO/IEC 30107 standards of the ISO/IEC **Subcommittee 37** (**SC37**) on Biometrics [163], which we will follow in the rest of this section. **Presentation Attack Instruments** (**PAIs**) used to carry out impostor attacks are typically generated from *bona fide* images of an iris from an individual who has legitimate access to the system. The iris is printed on a piece of paper (printout attack) or displayed on a screen (replay attack) and then presented to the sensor. The iris of deceased individuals can also be used as PAI since the texture remains intact for some hours [169]. Theoretically, it would be possible to print a genuine iris texture into a contact lens as well, although this has not been successfully demonstrated yet [16]. Concealer attacks, on the other hand, are commonly done via textured contact lenses that obscure or alter properties of the eye (such as color) to prevent the system from identifying the user. Synthetic iris images [191] not belonging to any specific identity could be used for similar purposes. Concealers can also present their legitimate iris, but in a way not expected by the system, e.g., closing eyelids as much as possible, looking to the sides (off-axis gaze), rotating the head, and so on.

Two challenges of PAs are that they happen outside the physical limits of the system, and they do not require specific knowledge of its inner workings, or any technical knowledge at all. Thus, if not properly tackled, they can derail public perception of even the most reliable biometric modality. It is even more critical if authentication is done without any supervision. *PAD* methods to counteract such attacks can be done [54]: (*i*) at the *hardware* (or sensor) level, using additional illuminators or

Table 3. Cohesive Comparison of the Most Relevant DL-based Iris Presentation Attack Detection Methods After the Surveys [16, 30] (NIR: *near-infrared*; VW: *visible wavelength*)

Category	Method	Year	Data NIR	Data VW	Datasets	Features
Feature Extraction	Fang et al. [44]	2020	✓	✗	LivDet-2017 (IIITD-WVU, ND-CLD)	VGG16, MobileNetv3-small (multi-layer features) + PCA + SVM
End-to-end Training	Arora and Bathia [8]	2020	✓	✗	LivDet-2017 (IIITD-WVU)	CNN with patch input
	Peng et al. [126]	2020	✓	✗	IPITRT, CASIA-Iris-v4, CASIA-Iris-Fake	LAILNet lightweight CNN
	Sharma and Ross [150]	2020	✓	✗	Proprietary, LivDet-2017 (IIITD-WVU, ND-CLD, Warsaw, Clarkson)	DenseNet121 pre-trained on ImageNet
	Chen and Ross [19]	2021	✓	✗	JHU-APL, LivDet-2017 (Warsaw, ND-CLD)	DenseNet121 pre-trained on ImageNet + AG-PAD channel and spatial attention
	Fang et al. [45]	2021	✓	✗	LivDet-2017 (IIITD-WVU, ND-CLD), ND-CLD-15,	MobileNetv3-small with micro-stripes
	Fang et al. [46]	2021	✓	✗	LivDet-2017 (IIITD-WVU, ND-CLD, Clarkson), ND-CLD-13, ND-CLD-15, IIITD-CLI	DenseNet + A-PBS spatial attention
	Fang et al. [47]	2021	✓	✗	LivDet-2017 (IIITD-WVU, ND-CLD, Clarkson)	ResNet50, VGG16, MobileNetv3-small
	Gupta et al. [63]	2021	✓	✗	MUIPA, UnMIPA, IIITD-CLI	CNN with multi-branch classification
	Sharma and Ross [151]	2021	✓	✓	OCT, NIR and VW images	VGG19, ResNet50, DenseNet121
	Zhang et al. [199]	2021	✓	✗	ND-CLD-13, CASIA-Iris-Fake, IF-VE	VGG16 + WRN local attention and global classification
	Agarwal et al. [1]	2022	✓	✗	MUIPA, UnMIPA, IIITD-CLI, LivDet-2017 (IIITD-WVU), ND-PSID	Siamese contraction-expansion CNN, feature fusion
	Agarwal et al. [2]	2022	✓	✗	MUIPA, UnMIPA, IIITD-CLI, LivDet-2017 (IIITD-WVU), ND-PSID, NDIris3D	Siamese CNN, score fusion
	Gautam et al. [59]	2022	✓	✗	SYN, IIITD-CLI, IIITD-IS	Autoencoder with reconstruction and classification loss
	Tapia et al. [162]	2022	✓	✓	LivDet-2020, Iris-CL1, Warsaw-Post-Mortem v3.0	MobileNetv2, data augmentation, class-weights
Hybrid Methods	Choudhary et al. [27]	2022	✓	✗	IIITD-CLI, ND-CLD-13, CASIA, LivDet-2017 (IIITD-WVU, ND-CLD, Clarkson)	MBISF (domain-specific filters), SIFT, Haralick, DenseNet, VGG8 + SVM classification
	Choudhary et al. [26]	2022	✓	✗	IIITD-CLI, ND-CLD-13, LivDet-2017 (Clarkson)	MBSIF (generic filters), MBSIF (domain-specific filters), SIFT, LBPV, DAISY, DenseNet121 + SVM classification
Adversarial Networks	Yadav and Ross [193]	2021	✓	✗	Casia-Iris-Fake, Berc-iris-fake, ND-CLD-15, LivDet-2017, MSU-IrisPA-01	BSIF, DESIST, VGG16, D-NetPAD, AlexNet

Methods are listed in chronological (and then alphabetical) order.

sensors that detect intrinsic properties of a living eye or responses to external stimuli (like pupil contraction or reflection), or (*ii*) at the *software* level, using only the footprint of the PA (if any) left in the same images captured with the standard sensor that will be employed for authentication.

Software-based techniques are in principle less expensive and intrusive since they do not demand extra hardware, and they will be the focus of this section.

Two comprehensive surveys on PAD are [30] (2018) and [16] (2020). While DL techniques were residual in the 2018 survey, they rose in popularity thereafter. We build this section upon the latest survey and summarize the most important developments in DL-PAD since it was published (Table 3). A descriptive summary of the datasets employed is given later in Section 8. The aim of PAD is to classify an image either as a *bona fide* or an *attack* presentation, so it is usually modeled as a two-class classification task. Typical strategies mimic the trend of the previous section when applying DL to iris recognition: either a CNN backbone is used to extract features that will feed a conventional classifier or the network is trained end-to-end to do the classification itself. Some hybrid methods also combine traditional hand-crafted with deep-learned features. In the same manner, the network may be initialized e.g., on the ImageNet dataset to take advantage of such a

large generic corpus, since available iris PAD data is more scarce. Another strategy also employed widely in the PAD literature is to use adversarial networks, where a GAN [60] is trained to generate synthetic iris images that the discriminator must use to detect attack samples.

4.1 CNNs for Feature Extraction

Since each layer of a CNN represents a different level of abstraction, Fang et al. [44] fused the features from the last four convolutional layers of two models (VGG16, MobileNetv3-small). The features are projected to a lower dimensional space by PCA and either concatenated for classification with SVM (feature fusion) or the classification scores of each level combined (score fusion). Using two databases of printouts and textured contact lenses, the method showed superiority over the use of the different layers individually, or the feature vector from the next-to-last layer of the networks.

4.2 End-to-end Classification Networks

Arora and Bathia [8] trained a CNN with 10 convolutional layers to detect contact lenses and printouts. Rather than using the entire image, the network is trained on patches from all parts of the iris image. The system showed superior performance compared to state-of-the-art methods which at that time, according to the article, were mostly based on hand-crafted features.

Focusing on embedded low-power devices, Peng et al. [126] adopted a **Lite Anti-attack Iris Location Network** (**LAILNet**) based on three dense blocks featuring depthwise separable convolutions to reduce the number of parameters. The algorithm demonstrated very good performance on three databases with printouts, synthetic irises, contact lenses, and artificial plastic eyes.

Also focusing on mobiles, Fang et al. [45, 48] used MobileNetv3-small. The contribution lies in the division of the normalized iris image into overlapped micro-stripes which are fed individually and a decision is reached by majority voting. The claimed advantages are that the classifier is forced to focus on the iris/sclera boundaries (given by their exact micro-stripes), the input dimensionality is lower and the amount of samples is higher (reducing overfitting), and the impact of imprecise segmentation is alleviated. Using three databases with contact lenses and printouts, the article featured extensive experimentation with cross-database, cross-sensor, and cross-attack settings.

Sharma and Ross [150] proposed D-NetPAD, based on DenseNet121, chosen due to benefits such as the maximum flow of information given by dense connections to all subsequent layers, or fewer parameters compared to counterparts like ResNet or VGG. The PAI included printouts, artificial eye, cosmetic contacts, kindle replay, and transparent dome on print, with experiments substantiating the effectiveness of the method on cross-PAI, cross-sensor, and cross-database scenarios.

Chen and Ross [19] proposed an explainable attention-guided detector (AG-PAD). To do so, the feature maps of a DenseNet121 were fed into two modules that independently capture inter-channel and inter-spatial feature dependencies. The outputs were then fused via element-wise sum to capture complementary attention features from both channel and spatial dimensions. With three datasets containing colored contact lenses, artificial eyes (Van Dyke/Doll fake eyes), printouts, and textured contact lenses, the attention modules are shown to improve accuracy over the baseline network. Using heatmap visualization, it is also shown that the attention modules force the network to attend to the annular iris textural region which, intuitively, plays a vital role for PAD.

Spatial attention was also explored by Fang et al. [46]. To find local regions that contribute the most to making accurate decisions and capturing pixel/patch-level cues, they proposed an **attention-based pixel-wise binary supervision** (**A-PBS**) method. To capture different levels of abstraction, they perform multi-scale fusion by adding spatial attention modules to feature maps

from three levels of a DenseNet backbone. Using six datasets with textured lenses and printouts, they outperformed previous state-of-the-art including scenarios with unknown attacks, sensors, and databases.

Given the difficulty of collecting iris PAD data, most databases contain, at most, a few hundred subjects. To address this, Fang et al. [47] studied data augmentation techniques that modify position, scale or illumination. Using three architectures (ResNet50, VGG16, and MobileNetv3-small) and three databases with printouts and textured contact lenses, they found that data augmentation improves PAD performance significantly, but each technique has a positive role on a particular dataset or CNN. They also explored the selection of augmentation techniques, finding, again, no consensus regarding the best combination, which was attributed to differences in capture environment, subject population, scale of the different datasets, or imbalance between bona fide and attack samples.

Gupta et al. [63] proposed MVANet, with five convolutional layers and three branches of fully connected layers. They addressed the challenge of unseen databases, sensors, and imaging environments on textured contact lens detection. The size of each layer of MVANet is different, thus capturing different features. They used three databases, each one captured in different settings (indoor/outdoor, different times of the day, varying weather, fixed/mobile sensors, etc.), with MVANET trained in one database at a time and tested on the other two. As a baseline, they fine-tuned three popular CNNs (VGG16, ResNet18, and DenseNet) initialized on ImageNet. The proposed network is shown to perform consistently better and more uniformly on the test databases than the baseline approaches.

Sharma and Ross [151] studied the viability of **Optical Coherence Tomography** (**OCT**). OCT provides a cross-sectional view of the eye, whereas traditional NIR or VW imaging provides 2D textural data. The PAIs considered are artificial eyes (Van Dyke eyes) and cosmetic lenses, evaluated on three different CNNs (VGG19, ResNet50, and DenseNet121). By both intra- (known PAs) and cross-attack (unknown PAs) scenarios, OCT is determined as a viable solution, although hardware cost is still a limiting factor. Indeed, OCT outperforms NIR and VW in the intra-attack scenario, while NIR generalizes better to unseen PAs. Cosmetic lenses also appear to be more difficult to detect than artificial eyes with any modality. Via heatmaps, it is seen as well that the fixation regions are different for each imaging modality and for each PAI, which could be a source of complementarity.

Zhang et al. [199] proposed a **Weighted Region Network** (**WRN**) to detect cosmetic lenses that include a local attention Weight Network (for evaluating the discriminating information of different regions) and a global classification Region Network (for characterizing global features). Such a strategy considers both the entire image and the attention effect by assigning different weights to regions. The mentioned networks are applied to a VGG16 backbone. The reported results showed improved performance compared to the state-of-the-art over three different databases.

The works by Agarwal et al. [1, 2] evaluated the detection of contact lenses. In [2], they trained a siamese CNN of five convolutional layers on two different inputs (the original image and its CLAHE version), which are then combined by weighted score fusion of the softmax layer. Adding a processed version of the raw image attempts to enhance the feature extraction capabilities of the CNN. A similar strategy is followed in [1], but here they used a siamese contraction-expansion CNN, and the processed image is an edge-enhanced image obtained via **Histogram of Oriented Gradients** (**HOG**). Another difference was the use of a feature-level fusion of the next-to-last CNN feature vectors, testing different strategies (vector addition, multiplication, concatenation, and distance). The articles employed several databases, with an extensive protocol including unseen subjects, environments (indoor vs. outdoor) and databases (sensors) that showcase the strength of the solutions against cross-domain changes. The methods also showed superiority

against popular CNN models (VGG16, ResNet18, and DenseNet) and the popular LBP and HOG hand-crafted features.

Gautam et al. [59] proposed a **Deep Supervised Class Encoding** (**DSCE**) approach consisting of an Autoencoder that exploits class information, and minimizes simultaneously the reconstruction and classification errors during training. Three datasets were used, containing textured lenses, printouts, and synthetic images, showing superiority over a variety of hand-crafted and deep-learned features.

Tapia et al.[162] used a two-stage serial architecture based on a modified MobiletNetv2. A first network was trained to only distinguish two classes (bona fide vs. attack). If it votes *bona fide*, the image is sent to a second network trained to classify it among three or four classes (bona fide or a different type of PAI: contact lenses, printout, or cadaver). Four databases were combined to obtain a super-set with the different PAIs, and class weights were also incorporated into the loss to compensate imbalance. The article applied contrast enhancement (CLAHE), and an aggressive data augmentation (rotation, blurring, contrast change, Gaussian noise, edge enhancement, image region dropout, etc.). They tested two image sizes, 224×224 and 448×448, observing that the extra detail of a higher resolution image results in more effective features. The article also carried out leave-one-out PAI tests for open-set evaluation, showing robustness in detecting unknown attacks.

4.3 Hybrid Methods

Choudhary et al. [26, 27] applied a Friedman test-based selection method to identify the best features of a set of hand-crafted and deep-learned ones. Each feature method feeds an SVM classifier, and the scores of the individual SVMs are fused via a weighted sum. A preliminary version of [26] without feature selection appeared in [25]. The databases of [27] include a medley of different PA (printouts, synthetic irises, artificial eyeballs, etc.), although the feature selection and classification methods are trained and evaluated separately on each database. The authors observed a saturation after a certain number of features are combined, and the superiority of the score-level fusion over other methods such as majority voting, feature-level fusion, and rank-level fusion. The work [26], on the other hand, concentrated on textured contact lenses attack, with an extensive set of evaluations including single sensor, cross-sensor, and combined sensor experiments. Apart from the generic live vs. attack scenario, it also reports binary and ternary classification across the different types of real (normal iris, soft lens) and fake (textured) classes. Naturally, the cross-sensor error is larger compared to the single-sensor, and the combined sensor error is also observed to be slightly larger. The latter is attributed to the larger intraclass variation created when images from different sensors are combined. In any case, an improvement in performance over previous works with the three datasets employed is observed after the proposed feature selection and score-level fusion method.

4.4 Adversarial Networks

Generative methods have been used by some approaches, either to use the trained discriminator for iris PAD, or to generate synthetic samples and augment under-represented classes. In this direction, Yadav and Ross [193] proposed **Cyclic Image Translation Generative Adversarial Network** (**CIT-GAN**) for multi-domain style transfer to generate synthetic samples of several PAIs (cosmetic contact lenses, printed eyes, artificial eyes, and kindle-display attacks). To do so, image translation is driven by a Styling Network that learns the style characteristics of each given domain. It also employs a Convolutional Autoencoder in the generator for image-to-image style translation, which takes a domain label as input along with an image. This is different than previous works of the same authors [191, 192] which employed the traditional generator/discriminator

approach driven by a noise vector. Different PAD methods using hand-crafted (BSIF, DESIST) and deep features (VGG16, D-NetPAD, and AlexNet) were evaluated, demonstrating that they can be improved by adding synthetically generated data. The quality of synthetic images is also superior to a competing generative method (Star-GAN v2), measured via FID score distributions.

4.5 Open Research Questions in Iris PAD

One of the open research issues is to design robust iris PAD methods with cross-sensor and cross-database capabilities, so they generalize to unseen imaging conditions. Attackers are constantly developing new attack methodologies to circumvent PAD systems, so an even more important issue is unseen PAIs (i.e., cross-PAI capabilities) [149]. Great results have been achieved in detecting known attack types (known as *closed-set* recognition), although cross-database evaluation (training in one database and testing in others) still appears as a difficult challenge due to changes in sensors, acquisition environments, or subjects. Moreover, generalizing to attacks that are unknown at the time of training (*open-set* recognition) is an even greater challenge for state-of-the-art methods [45]. Part of the problem lies in the limited size of existing databases, which is an issue for data-hungry DL approaches. Some solutions, as studied by some of the methods above, are data augmentation by geometric or illumination modifications [47], or creating additional synthetic data via generative methods [193]. Human-aided DL training is another promising avenue. Indeed, humans and machines cooperating in vision tasks are not new, and this strategy is finding its way into DL as well [14, 17]. For example, Boyd et al. [14] analyzed the utility of human judgment about salient regions of images to improve the generalization of DL models. Asked about regions that humans deem important for their decision about an image, the work proposed to transform the training data to incorporate such opinions, demonstrating an improvement in accuracy and generalization in leave-one-attack-type-out scenarios. In a similar work, Boyd et al. [17] incorporated annotated saliency maps into the loss function to penalize large differences with human judgment.

Recently, concerns have emerged about the observed bias of DL methods that leads to discriminatory performance differences based on the user's demographics, with face biometrics being the most talked-about and many companies and authorities banning its use [78]. This issue appears in iris PAD as well, as addressed by Fang et al. [49]. Using three baselines based on hand-crafted and DL approaches and a database of contact lenses, the authors showed a significant difference in the performance between male and female samples. In dealing with this phenomenon, examination of biases towards eye color or race is another direction worthwhile to consider.

Some elements considered as PAIs in this section, such as cosmetic lenses, may be worn normally by users without the purpose of fooling the biometric system, as it is the case of facial retouching via make-up, digital beautification, or augmented reality [69]. This poses the question of whether it is possible to use such images for authentication while diminishing the effect on the recognition performance. Suggested alternatives have been to detect and match portions of live iris tissue still visible [125] or incorporate ocular information of the surrounding area [4]. Unfortunately, in iris biometrics, recognition with textured contact lenses remains a hard problem to solve.

Another under-researched task is iris PAD in the visible spectrum. The majority of studies and datasets (Section 8) employ NIR illumination and specific iris close-up sensors. However, in some environments such as mobile or distant capture, such sensing is not guaranteed [119].

5 DEEP LEARNING-BASED FORENSIC IRIS RECOGNITION

Iris recognition has become the next biometric mode (in addition to the face, fingerprints, and palmprints) considered for large-scale forensic applications [52], and coincides in time with discoveries made in recent years about the possibility of employing iris in recognition of deceased subjects. This includes both matchings of iris patterns acquired a few hours after death with those
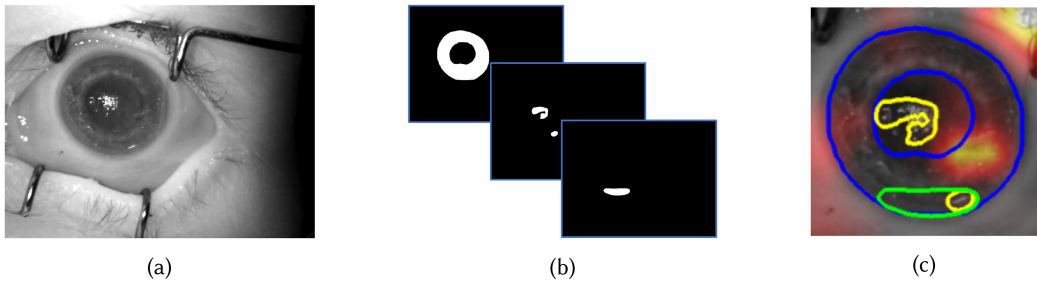
Fig. 3. Post-mortem iris recognition and visualization: (a) a good-quality post-mortem iris image; (b) top to bottom: deep learning-based detection of iris annulus, specular highlights, and decomposition-induced wrinkles; (c) segmentation results presented to a human examiner along with an overlaid heatmap visualizing regions judged as salient by the matching algorithm. Source: [91]

with longer PMIs (Post-Mortem Intervals), ranging from days [12, 143, 167, 168] to several weeks after demise [18, 170], as well as matching patterns acquired before death with those collected post-mortem [141].

Due to decomposition changes to the eye tissues, post-mortem iris images differ significantly from live iris images and rarely meet ISO/IEC 29794-6 quality requirements, as shown in Figure 3(a). The challenges are related to the appropriate detection of places when the cornea dries and generates irregular and large specular highlights, as well as regions where iris muscle furrows show up when the eyeball dehydrates. This is where DL-based methods may win over hand-crafted approaches, as the latter usually make strong assumptions about the anatomy of the iris appearance, making it possible to predict eyes undergoing random decomposition processes. Trokielewicz et al. proposed the first known to us iris recognition method designed specifically to cadaver irises [171, 172]. It incorporates a SegNet-based segmenter and Siamese networks-based feature extractor, both trained in a domain-specific way solely on post-mortem iris samples. An interesting element of this approach is that segmentation incorporates two models: one trained with "fine" ground truth masks, marking all details associated with eye decomposition, and the "coarse" model, aiming at detecting iris annulus and eyelids, as in classical iris recognition approaches. This allowed the application of a standard "rubber sheet" iris image normalization based on "coarse" masks, and at the same time excluded decomposition-driven artifacts from encoding, marked by the "fine" mask. Kuehlkamp et al. [91] in addition to detecting post-mortem deformations, as shown in Figure 3(c), they also proposed a human-interpretable visualization of a classification process. The visualization is based on the Class Activation Mapping mechanism [212] and highlights salient features used by the classifier in its judgment. This novelty in iris recognition algorithms may help human examiners locate iris regions that should be carefully inspected or to verify the algorithm's decision.

6 HUMAN-MACHINE PAIRING TO IMPROVE DEEP LEARNING-BASED IRIS RECOGNITION

Iris recognition is usually associated with automatic, solely machine-based, and rapid biometric means. It has been changing in the recent decade due to the constantly increasing ubiquitousness of iris recognition, especially owing to large governmental applications such as [174] or FBI's **Next Generation Identification** (**NGI**) System gradually replacing the **Integrated Automated Fingerprint Identification System** (**IAFIS**) [52]. This combined with the unique identification power of the iris whetted the appetite to apply this technique to identification problems normally

reserved for fingerprints and face: forensics, lost subjects search, or post-mortem identification. To have the legal power, however, the judgment about samples originating or not from the same eye conclusion must be confirmed by a trained human expert. And here is the place where DL-based iris image processing may play a useful role.

Trokielewicz et al. compared iris images in post-mortem iris recognition between humans and machines. They investigated which iris image regions humans and machines mainly attend to compare a pair of images. The machine-based attention maps are generated by Grad-CAM to highlight the regions that contribute the most to the deep learning model's prediction. The human-based attention maps are learned by tracking the gaze as the human is looking around the screen that displays iris image pairs and recording the regions where the human spends the most time. Interestingly while humans and machines tend to focus on a limited number of iris areas; however, the region, appearance, and density of these areas between humans and machines are different. As salient regions proposed by the deep learning model and identified from the human eye gaze do not overlap in general, the computer-added visual cues may potentially constitute a valuable addition to the forensic examiner's expertise, as it can highlight important discriminatory regions that the human expert might miss in their proceedings. This human-machine pairing is important as human subjects can make an incorrect decision even after spending quite some time observing many iris regions [120]. In addition, there has been a body of research showing that humans and machines do not perform similarly well under different conditions [20, 108, 154]. For example, Moreira et al. also showed that machines can outperform humans in healthy easy iris image pairs; however, humans outperform machines in disease-affected iris image pairs [108]. Human-machine pairing will improve deep learning-based iris recognition.

7 RECOGNITION IN LESS CONTROLLED ENVIRONMENTS: IRIS/PERIOCULAR ANALYSIS

Rooted in the seminal work due to Park et al. [124], efforts have been paid to the development of human recognition methods that—apart from the iris—also consider information in the vicinity of the eye to infer the identity. This is a relatively recent topic, termed as *periocular recognition*. The rationale is that the periocular region represents a tradeoff between the face and the iris. Periocular biometrics has been claimed to be particularly useful in environments that produce poor-quality data (e.g., visual surveillance). Recently, as in the case of iris, several DL-based solutions have been proposed.

Hernandez-Diaz et al. [70] tested the suitability of off-the-shelf CNN architectures to the periocular recognition task, observing that albeit such networks are optimized to classify generic objects, their features still can be effectively transferred to the periocular domain.

In the visual surveillance context, Kim et al. [88] infer subjects' identities based either on loose/tight regions of interest, depending on the perceived image quality. Hwang and Lee [77] prevent the loss of mid-level features and dynamically select the most important features for classification. Luo et al. [102] used self-attention channel and spatial mechanisms in the feature encoding module of a CNN, to obtain the most discriminative features of the iris and periocular regions.

Jung et al. [82]'s work is based on the concept of **label smoothing regularization** (**LSR**). Having as the main goal to reduce the intra-class variability, they described a so-called **Generalized LSR (GLSR)** by learning a pre-task network prediction that is claimed to improve the permanence of the obtained periocular features. Having similar purposes, Zanlorensi et al. [198] described a pre-processing step based on generative networks able to compensate for the typical data variations in visual surveillance environments. Nie et al. [118] applied convolutional restricted Boltzmann machines to the periocular recognition problem. Starting from a set of genuine pairs that are used as a constraint, a Mahalanobis distance metric is learned.

Obtaining auxiliary (e.g., soft biometrics) has been seen as an interesting direction for compensating for the lack of image quality. Zhao and Kumar [210] incorporate an attention model into a DL architecture to emphasize the most important regions in the periocular data. The same authors [208] described a semantics-assisted CNN framework to infer comprehensive periocular features. The whole model is composed of different networks, trained upon ID and semantic (e.g., gender, ethnicity) data, that are fused at the score and prediction levels. Similarly, Talreja et al. [158] described a multi-branch CNN framework that predicts simultaneously soft biometrics and ID labels, which are finally fused into the final response.

With regard to cross-spectral settings, Hernandez-Diaz et al. [71] used **conditional GANs (CGANs)** to convert periocular images between domains, that are further fed to intra-domain off-the-self frameworks. Sharma et al. [148] described a shallow neural architecture where each model learns the data features in each spectrum. Then, at a subsequent phase, all models are jointly fine-tuned, to learn the cross-spectral variability and correspondence features.

Finally, several works have attempted to faithfully fuse the scores/responses from iris and periocular data. Wang and Kumar [181] used periocular features to adaptively match iris data acquired in less constrained conditions. Their framework incorporates such discriminative information using a multilayer perceptron network. Zhang et al. [203] described a DL-model that exploits complementary information from the iris and the periocular regions, that applies *maxout* units to obtain compact representations for each modality and then fuses the discriminative features of the modalities through weighted concatenation. In the opposite direction, Proença and Neves [134] argued that the periocular recognition performance is optimized when the components inside the ocular globe (the iris and the sclera) are simply discarded.

8 OPEN-SOURCE DEEP LEARNING-BASED IRIS RECOGNITION TOOLS

Here we summarize the main properties of the datasets employed by the methods of the previous sections for DL-based iris segmentation, recognition, and PAD. We also describe available open-source software code for these tasks and other relevant tools.

8.1 Data Sources

Table 4 gives the technical details of the datasets used in the segmentation and recognition methods of Tables 1 and 2. Table 5 does the same for the iris PAD methods of Table 3. We show the main properties (spectrum, image size, identities, images, sessions) and relevant features. Only the datasets of the methods reported in the previous section are presented. Since we focus on the most recent developments, we consider that such an approach provides the most relevant datasets for each task. Of course, the list of available datasets after decades of iris research is much longer [122].

A first observation is the dominance of NIR over the visible (VW) spectrum, which should not be surprising since NIR is regarded as most suitable for iris analysis. However, research-wise, many segmentation and recognition studies (Tables 1, 2) use VW images, pushed by the success of challenging databases such as MICHE and UBIRIS. On the contrary, the VW modality in iris PAD research is residual (Table 3), a tendency also observed in pre-DL research [16, 30].

When it comes to the types of PAIs employed in iris PAD databases, they can be categorized into:

— PP: paper printout of a real iris image, i.e., from a live person.
— PPD: paper printout of a real iris image with a transparent 3D plastic eye dome on top.
— CLL: textured contact lenses worn by a live person.
— CLP: textured contact lenses on the printout (either a printout of a CLL image or a printout of a real iris image with a textured contact lens placed on top).

Table 4. Summary of Datasets Used in the DL-based Iris Segmentation and Recognition Methods of Tables 1 and 2 (NIR: *near-infrared*; VW: *visible wavelength*)

Name	Data	Size	# IDs	# Samples	# Sessions	Features
BATH [107]	NIR	1280 × 960	1600	16000	1	High quality images
BioSec [53]	NIR	640 × 480	400	3200	2	Office environment
Biosecure [123]	NIR	640 × 480	1334	2668	2	Office environment
CASIA-Cross-Sensor [187]	NIR	n/a	700	21000	1	Multi-sensor, multi-distance (12–30cm, 3–5m)
CASIA-Iris-Distance [40]	NIR	2352 × 1728	284	2567	1	Distant acquisition
CASIA-Iris-Interval [103]	NIR	320 × 280	395	2639	2	High quality images
CASIA-Iris-Lamp [185]	NIR	640 × 480	819	16212	1	Non-linear deformation
CASIA-Iris-M1-S1 [204]	NIR	1920 × 1080	140	1400	1	Mobile device
CASIA-Iris-M1-S2 [202]	NIR	1968 × 1024	400	6000	1	Mobile device, multi-distance (20, 25, 30cm)
CASIA-Iris-M1-S3 [203]	NIR	1920 × 1920	720	3600	1	Mobile device
CASIA-Iris-Thousand [200]	NIR	640 × 480	2000	20000	1	High quality images
DCME01 [18]	NIR, VW	n/a	254	621	1–9	–
DCME02 [91]	NIR	n/a	259	5770	1–53	–
IITD [92]	NIR	320 × 240	224	1120	1	Varying quality
Iris-Mobile [121]	NIR	n/a	750	22966	n/a	Mobile device, indoor & outdoor
JluIrisV3.1 [207]	NIR	640 × 480	120	1780	n/a	–
JluIrisV4 [207]	NIR	640 × 480	172	114904	n/a	–
LivDet-2013-Warsaw [28]	NIR	640 × 480	284	1667	1	High quality images
MICHE-I [38]	VW	var.	184	3732	2	Three mobile devices
MMU	NIR	320 × 240	92	460	1	High quality images
MobBIOfake [146]	VW	300 × 200	200	1600	1	With a handheld device
ND-CrossSensor-2013 [187]	NIR	640 × 480	1352	146550	27	Multi-sensor
ND-Iris-0405 [127]	NIR	640 × 480	712	64980	1	Varying quality
ND-TWINS-2009-2010	VW	n/a	435	24050	n/a	Facial pictures frontal, 3/4 and side views. Indoor & outdoor
OpenEDS [58]	NIR	640 × 400	304	356649	1	From head-mounted VR glasses
Q-FIRE [81]	NIR	var.	390	586560	2	Iris/face Videos, various distances and quality
UBIRIS.v1 [130]	VW	800 × 600	241	1877	2	Several noise factors
UBIRIS.v2 [131]	VW	400 × 300	522	11102	2	Distant acquisition, on the move
Warsaw [18]	NIR, VW	n/a	157	4866	1–13	–
Warsaw-Post-Mortem v1.0 [167]	NIR, VW	var.	34	1330	2–3	Deceased persons, 5–7h to 17 days post-mortem
Warsaw-Post-Mortem v2.0 [170]	NIR, VW	var.	73	2987	1–13	Deceased persons

— RA: replay attack, i.e., a real iris image shown on a display.
— AE: artificial eyeball (plastic eyes of two different types: Van Dyke Eyes, with higher iris quality details, and Scary eyes, plastic fake eyes with a simple pattern on the iris region).
— AEC: artificial eyeball with a textured contact lens on top.
— SY: synthetic iris, i.e., an image created via generative methods.
— PM: postmortem iris, i.e., an image acquired from cadaver eyes.

These PAIs mostly entail presenting the mentioned instrument to the iris sensor, which then captures an image of the artifact. An exception is "SY", which directly produces a synthetic digital image, although such an image could be used as a base to, for example, PP, PPD, RA, or AE attacks. In Table 5, it can be seen that CLL (textured lenses live) and PP (paper printouts) largely dominate as the most popular PAIs on the existing databases, and consequently, on the related research (Table 3). CLP (textured lenses on the printout) also appears in many studies, driven by the wide use of the LivDet-2017-IIITD-WVU set, which includes such PAI. CASIA-Iris-Fake, which contains AE (artificial eyes) and SY (synthetic irises) also appears in a few studies. Other attacks that one may expect in the digital era, such as RA (replay), however, are residual in datasets and recent studies.

Table 5. Summary of Datasets Used in the DL-based Iris Presentation Attack Detection Methods of Table 3 (NIR: *near-infrared*; VW: *visible wavelength*)

Name	PAIs	Data	Size	# IDs		# Samples			TTP	Features
				live	fake	live	fake	total		
CASIA-Iris-Fake [155]	PP, CLL, AE, SY	NIR	640 × 480	1000	815	6000	4120	10240		
IF-VE [199]	CLL	NIR	n/a	200	200	25000	25000	50000	✓	MS, ME
IPITRT [126]	PP	NIR	var.	58	n/a	1800	551	2351		ME
IIITD-CLI [89]	CLL	NIR	640 × 480	202	n/a	n/a	n/a	6570	✓	MS
IIITD-IS[3] [64]	PP, CLP	NIR	640 × 480	202	n/a	0	4848	4848		MS
LivDet-2017 [194] -Clarkson	PP, CLL	NIR	640 × 480	50	n/a	3954	4141	8095	✓	UPAI (additional patterned lenses)
-IIITD-WVU[1]	PP, CLL, CLP	NIR	640 × 480	n/a	n/a	2952	4507	7459	✓	MS, ME, UPAI (additional patterned lenses)
-ND-CLD[2]	CLL	NIR	640 × 480	n/a	n/a	2400	2400	4800	✓	UPAI (additional patterned lenses)
-Warsaw	PP	NIR	640 × 480	457	446	5168	6845	12013	✓	MS
LivDet-2020 [33]	PP, PPD, CLL, CLP, RA, AE, AEC, PM	NIR	640 × 480	n/a	n/a	5331	7101	12432		MS
Iris-CL1 [162]	PP	NIR	var.	n/a	n/a	n/a	1800	n/a		MS
JHU-APL [19]	CLL, AE	NIR	n/a	n/a	n/a	7191	7214	14405		ME
MSU-IrisPA-01 [191]	PP, CLL, RA, AE	NIR	640 × 480	n/a	n/a	1343	2523			
MUIPA [190]	PP, CLL	NIR	640 × 480	70	70	n/a	n/a	10296		ME
ND-CLD-13 [42]	CLL	NIR	640 × 480	330	n/a	3400	1700	5100	✓	MS
ND-CLD-15[2] [41]	CLL	NIR	640 × 480	n/a	n/a	4800	2500	7300	✓	MS
NDIris3D [51]	CLL	NIR	640 × 480	176	176	3458	3392	6850		MS
ND-PSID[4] [31]	CLL	NIR	640 × 480	238	238	3132	2664	5796		MS
UnMIPA [189]	CLL	NIR	640 × 480	162	162	9319	9387	18706		MS, ME
Warsaw-Post-Mortem v3.0 [172]	PM	NIR, VW	var.	0	79	0	1879	1879		MS

[1]Contains IIITD-CLI and IIITD-IS.
[2]Iris-LivDet-2017-ND-CLD is a subset of ND-CLD-15.
[3]IIITD-IS images are printouts of IIITD-CLI captured with an iris scanner and a flatbed scanner.
[4]ND-PSID is a subset of ND-CLD-15.
The type of PAIs (second column) is PP: paper printout, PPD: paper printout with a plastic dome, CLL: textured contact lenses (live), CLP: textured contact lenses (printout), RA: replay attack (display), AE: artificial eyeball, AEC: artificial eyeball with textured contact lens, SY: synthetic iris, PM: postmortem iris. TTP (next to the last column) indicates the existence of a training/test split. The features (last column) are MS: multi-sensor, ME: multi-environment (indoor/outdoor, light variability, mobile environment, etc.), and UPAI: unseen PAIs in the test set.

8.2 Software Tools

The availability of DL-based tools for iris biometrics has been scarce for years, especially for PAD [51]. In the following, we provide a short description of peer-reviewed references with associated source code (link included in the article, or easily found on the websites of the authors or dedicated sites such as www.paperswithcode.com). We describe (in this order) tools for segmentation, recognition, and PAD. For each type, the references are then presented in chronological order.

8.2.1 Segmentation. Lozej et al. [101] released their end-to-end DL model based on the U-Net architecture [139]. The model was trained and evaluated with a small set of 200 annotated iris images from the CASIA database. The authors also explored the impact of the model depth and the use of batch normalization layers.

Kerrigan et al. [85] released the code and models of Iris-recognition-OTS-DNN, a set of four architectures based on off-the-shelf CNNs trained for iris segmentation (two VGG-16 with dilated convolutions, one ResNet with dilated kernels, and one SegNet encoder/decoder). Training databases included CASIA-Irisv4-Interval, ND-Iris-0405, Warsaw-Post-Mortem v2.0, and

ND-TWINS-2009-2010, whereas testing data came from ND-Iris-0405 (disjoint subject), BioSec and UBIRIS.v2. Results showed that the DL solutions evaluated outperform traditional segmentation techniques, e.g., Hough transform or integro-differential operators. It was also seen that each test dataset had a method that performed best, with UBIRIS obtaining the worst performance. This should not come as a surprise, since it contains VW images with high variability taken distantly with a digital camera, whereas the other two are from close-up NIR iris sensors in controlled environments.

Wang et al. [176] released the code and models of their high-efficiency segmentation approach, IrisParseNet. A multi-task attention network was first applied to simultaneously predict the iris mask, pupil mask, and iris outer boundary. Then, from the predicted masks and outer boundary, a parameterization of the iris boundaries was calculated. The solution is complete, in the sense that the mask (including light reflections and occlusions) and the parameterized inner and outer iris boundaries are jointly achieved.

More recently, authors from the same group presented IrisSegBenchmark [177], an open iris segmentation evaluation benchmark where they implemented six different CNN architectures, including FCNs [100], Deeplab V1, V2, V3 [21], ParseNet [98], PSPNet [206], SegNet [9], and U-Net [139]. The methods were evaluated on CASIA-Irisv4-Distance, MICHE-I, and UBIRIS.v2. As in [85], results showed that the best method depends on the database, being: ParseNet for CASIA (NIR data), DeeplabV3 for MICHE (VW images from mobile devices), and U-Net for UBIRIS (VW images from a digital camera). In this case, however, the three test databases behaved approximately equal, since they all contain difficult distant data. CASIA showed a slightly better accuracy, suggesting that NIR data may be easier to segment. Traditional, non-DL methods were also evaluated, concluding that DL-based segmentation achieves superior accuracy.

Banerjee et al. [10] released the code of their V-Net architecture, designed to overcome some drawbacks of U-Net, such as instability to tackle iris segmentation or the tendency to overfit. A pre-processing stage on the YCrCb and HSV spaces was also added to detect salient regions and aid the detection of iris boundaries. The method was evaluated on the difficult UBIRIS.v2 VW dataset.

8.2.2 Recognition. The code of the DL method ThirdEye was released by Ahmad and Fuller [3], based on a ResNet-50 trained with triplet loss. Authors directly used segmented images without normalization to a rectangular 2D representation, arguing that such a step may be counterproductive in unconstrained images. The model was evaluated on the ND-0405, IITD, and UBIRIS.v2 datasets.

The models of Boyd et al. [15] for recognition have been also released, based on a ResNet-50 with different weight initialization techniques, comprising: from scratch (random), off-the-shelf ImageNet (general-purpose vision weights), off-the-shelf VGGFace2 (face recognition weights), fine-tuned ImageNet weights, and fine-tuned VGGFace2 weights. Both ImageNet and VGGFace2 are very large datasets with millions of images, and face images contain the iris region. Thus, using these datasets as initialization may be beneficial for iris recognition, where available training data is in the order of hundreds of thousands of images only. This strategy has been followed e.g., in ocular soft-biometrics as well [6]. The observed optimal strategy is indeed to fine-tune an off-the-shelf set of weights to the iris recognition domain, be general-purpose or face recognition weights.

8.2.3 Segmentation and Recognition Packages. A complete package comprising segmentation and feature encoding was provided by Tann et al.[161]. The segmentator is based on an FCN, but the encoding is based on hand-crafted Gabor filters [35]. The evaluation was done on CASIA-Irisv4-Interval and IITD.

In forensic investigation for diseased eyes and post-mortem samples, Czajka [29] also released a complete package combining segmentation and feature encoding. The models are based on

previous efforts of the author and co-workers, comprising a SegNet [172] and a CCNet [106] DL segmentators, but the feature encoder is based on hand-crafted BSIF filters.

Another complete segmentation and recognition package was released by Kuehlkamp et al. [91]. The segmentator is based on a fine-tuned Mask-RCNN architecture, with the cropped iris region fed directly into a ResNet50 pre-trained for face recognition on the very large VGGFace2 dataset, and fine-tuned for iris recognition using triplet loss. The article is oriented toward postmortem iris analysis, so the methods use a mixture of live and postmortem images for training and evaluation.

Parzianello and Czajka [125] also released the models and annotated data for their textured contact lens aware iris recognition method. The foundation is that such lenses may be used normally for cosmetic purposes, without the intention of fooling the biometric system. Therefore, they proposed to detect and match portions of live iris tissue still visible to enable recognition even when a person wears textured contact lenses. To do so, they applied a Mask R-CNN as a segmentation backbone, trained to detect authentically-looking parts of the iris using manually segmented samples from the NDIris3D dataset. Non-iris information is then removed from the training images by blurring it or replacing it with random noise to guide the subsequent recognition network (based on ResNet-18) to salient, non-occluded regions that should be used for matching.

8.2.4 Iris PAD. In the iris PAD arena, Gragnaniello et al. [61] proposed a CNN that incorporates domain-specific knowledge. Based on the assumption that PAD relies on residual artifacts left mostly in high-frequencies, a regularization term was added to the loss function which forces the first layer to behave as a high-pass filter. The method, which is available on the website of the first author, could be applied to PAD in multiple modalities, including iris and face.

The code and model of the method of Sharma and Ross [150] (D-NetPAD) is also available. It is based on DenseNet121 and trained for a variety of PAIs (printouts, artificial eye, cosmetic contacts, kindle replay, and transparent dome on print), with a script to retrain the method also available.

8.3 Other Tools: Iris Image Quality Assessment

Several image properties considered to potentially influence the accuracy of iris biometrics have been defined in support of the standard ISO/IEC 29794-6 [164]. They include the grayscale spread (dynamic range), iris size (pixels across the iris radius when the boundaries are modeled by a circle), dilation (ratio of the pupil to iris radius), usable iris area (percentage of non-occluded iris, either by eyelashes, eyelids, or reflections), the contrast of pupil and sclera boundaries, shape (irregularity) of pupil and sclera boundaries, margin (distance between the iris boundary and the closest image edge), sharpness (absence of defocus blur), motion blur, signal to noise ratio, gaze (deviation of the optical axis of the eye from the optical axis of the camera), and interlace of the acquisition device.

Low-quality iris images, which can potentially appear in uncontrolled or non-cooperative environments, are known to reduce the performance of iris location, segmentation, and recognition. Thus, an accurate quality assessment can be a valuable tool in support of the overall pipeline, either by dropping low-quality images, or invoking specialized processing [5]. One possibility might be to quantify the properties mentioned above, and place thresholds on each. A more elaborated alternative is to combine them according to some rule and produce an overall quality score. However, it is difficult to provide metrics that cover all types of quality distortions [157], and doing so for some indeed entails segmenting the iris.

Broadly, a biometric sample is of good quality if it is suitable for recognition, so quality should correlate with recognition performance [62]. As such, quality assessment can be viewed as a regression problem. Wang et al. [182] considered that a non-ideal eye image will pivot in the feature space around the embedding of an ideal image. They defined quality as the distance to the embedding of such an "ideal" image which, is regarded as a registration sample collected under a highly

controlled environment. They used a model to learn the mapping between images and **Distance in Feature Space** (**DFS**) directly from a given dataset. Quality is computed via attention-based pooling that combines a heatmap that comes from a coarse segmentation based on U-Net and the feature map of an extraction network based on MobileNetv2 pre-trained on CASIA-Iris-V4 and NDIRIS-0405.

9 EMERGING RESEARCH DIRECTIONS

In this section, we discuss the most relevant open challenges and hypothesize about emerging research directions that could become *hot-topics* in biometrics literature in the near future.

9.1 Resource-aware Designs of Iris Recognition Networks

Application-wise, iris recognition can be performed on a wide range of hardware, ranging from high-end computers to low-end embedded devices, or from large computer clusters to personal devices such as mobile phones. Performing recognition on resource-limited hardware could pose new challenges for deep learning-based iris networks, which usually contain hundreds of layers and millions of parameters. Therefore designing these deep learning networks necessarily need to be aware of the hardware platforms on which they will be run.

Lightweight models: Lightweight CNNs employ advanced techniques to efficiently tradeoff between resource and accuracy, minimizing their model size and computations in terms of the number of **floating point operations** (**FLOPs**), while retaining high accuracies. Specialized lightweight CNN architectures include MobileNets [73] and U-Net [139]. There are a few lightweight deep learning-based models for both segmentation and feature extraction. Fang et al. [50] adapted the lightweight CC-Net [106] for iris segmentation. CC-Net has a U-Net structure [139], able to retain up to 95% accuracy using only 0.1% of the trainable parameters. Boutros et al. [13] benchmarked MobileNet-V3 against deeper networks for iris recognition and showed that the MobileNet-based model can achieve similar EER with 85% less number of parameters and 80% less inference time.

Model compression: Studies have found that most of the large deep-learning models tend to be overparameterized, leading to lots of redundant parameters and operations in the network. This becomes more severe considering iris texture images are different from generic object-based images. This has motivated a hot trend looking to remove these redundancies from the models, including pruning, quantization, and low-rank factorization [95]. In our iris recognition literature, there are a few lightweight deep learning-based models for both segmentation and feature extraction. Tann et al. [161] quantized 64-bit floating points numbers of weights and activations of the full FCN-based iris segmentation model using an 8-bit **dynamic fixed-point** (**DFP**) format, which provides an 8× memory saving as well as speed enhancement due to reduced complexity of lower precision operations.

NAS: NAS automates the process of architecture design of neural networks by iteratively sampling a population of child networks, evaluating the child models' performance metrics as rewards, and learning to generate high-performance architecture candidates [43]. In our iris recognition literature, Nguyen et al. [113] showed that computation and memory can be incorporated into the NAS formulation to enable the resource-constrained design of deep iris networks.

9.2 Human-interpretable Methods and XAI

With hundreds of layers and millions of parameters, deep learning networks are usually opaque or "black box" where humans struggle to understand why a deep network predicts what it predicts.

This necessitates approaches to make deep learning methods more interpretable and understandable to humans. Interestingly, the need for human-interpretable methods has been raised even from the handcrafted era. For example, Shen et al. published a series of works [20, 152] on using iris crypts for iris matching. Iris crypts are clearly visible to humans in a similar way as finger minutiae. Another example is the macro-features [156] which use SIFT to detect key points and perform iris matching based on these keypoints [136]. Another notable work is by Proença et al. [132] where they proposed a deformation field to represent the correspondence between two iris images.

From a deep learning perspective, researchers have also attempted to visualize the matching. Kuehlkamp et al. [91] argued that existing iris recognition methods offer limited and non-standard methods of visualization to let human examiners interpret the model output. They applied **Class Activation Maps** (**CAMs**) [212] to visualize the level of contribution of each iris region to the overall matching score. Similarly, Nguyen et al. [115] also decomposed the final matching score into pixel-level to visualize the level of contribution of each pixel to the overall matching score.

9.3 Deep Learning-based Synthetic Iris Generation

Data synthesis provides an alternative to time- and resource-consuming database collection. One could create as many images as desired, with new textures that even do not match any existing identity, which would avoid privacy problems too. On the other hand, fake irises that are indistinguishable from real ones can be used for identity concealment attacks (if the image does not match any identity) or impersonation attacks (if the image resembles an existing identity) [30]. Indeed, synthetic irises are present in databases employed for iris PAD, such as CASIA-Iris-Fake (Table 5).

Regardless of the purpose or ability to detect if an image is synthetic, **Generative Adversarial Networks** (**GANs**) [60] have shown impressive photo-realistic generating capabilities in many domains. GANs learn to model image distributions by an adversarial process, where a discriminator assesses the realism of images synthesized by a generator. In the end, the generator has learned the distribution of the training data, being able to synthesize new images with the same characteristics.

For iris generation, some methods by Yadav et al. [191–193] were mentioned in iris PAD contexts (Section 4.4). RaSGAN [191, 192] followed the traditional approach of driving the generation/discrimination training by randomly sampling so-called latent vectors from a probabilistic distribution. As training progresses, the generator learns to associate features of the latent vectors with semantically meaningful attributes that naturally vary in the images. However, this does not impose any restriction on the relationship between features in latent space and factors of variation in the image domain, making it difficult to decode what the latent vectors represent. As a result, the image characteristics (eye color, eyelid shape, eyelashes, gender, age...) are generated randomly. Kohli et al. [90] presented iDCGAN for iris PAD, which also followed the latent vector sampling concept. To counteract such issues, researchers have tried to incorporate constraints or mechanisms that guide the generation process to a desired characteristic. For example, CIT-GAN [193] employed a Styling Network that learns style characteristics of each given domain, while taking as input a domain label that drives the network to embed a desired style into the generated data.

In a similar direction, Kaur and Manduchi [83, 84] proposed to synthesize eye images with a desired style (skin color, texture, iris color, identity) using an encoder-decoder ResNet. The method is aimed at manipulating gaze, so the generator receives a segmentation mask with the desired gaze and an image with the style that will see its gaze modified. To achieve cross-spectral recognition, Hernandez-Diaz et al. [71] used CGANs to convert ocular images between VW and NIR spectra

while keeping identity, so comparisons are done in the same spectrum. This allows the use of existing feature methods, which are typically optimized to operate in a single spectrum.

Despite great advances in DL-based synthetic image generation, one open problem is the possible identity *leakage* from the training set when creating data of non-existing identities, resulting in privacy issues. This has just been revealed recently in face generation [165]. Another issue in the opposite direction is the difficulty in preserving identity in the generation process when the target is precisely creating images of an existing identity with different properties. This is an issue being addressed in face-generation methods but is lacking in iris synthesis research.

9.4 Deep Learning-based Iris Super-resolution

One of the main constraints for existing iris recognition systems is the short distance of image acquisition, which usually requires a subject to stay still less than 60 cm from iris cameras. This is due to the requirement of a high-resolution iris region, e.g., 120 pixels across the iris diameter due to the European standard and NIST standard, despite the small physical size of an eye, i.e., 15×15 mm. The lack of resolution of imaging systems has critically adverse impacts on the recognition and performance of biometric systems, especially in less constrained conditions and long-range surveillance applications [116].

Super-resolution, as one of the core innovations in computer vision, has been an attractive but challenging solution to address the low-resolution problem in both general imaging systems and biometric systems. Deep learning-based super-resolution approaches have been across multiple works in iris recognition. Ribeiro et al. [137, 138] experimented with two deep learning single-image super-resolution approaches: **Stacked Auto-Encoders (SAEs)** and CNNs. Both approaches learn one encoder to map the high-resolution iris images to the low-resolution domain, and one decoder to learn to reconstruct the original high-resolution images from the low-resolution ones. Zhang et al. [201] learned a single CNN to learn a non-linear mapping function between LR images to HR images for mobile iris recognition. Wang et al. [183] extended the single CNN to two CNNs: one generator CNN and one discriminator CNN as in the GAN architecture. The generator functions similarly to the single LR - HR mapping CNN. Adding the discriminator CNN allows them to control the generator to generate HR images not just visually higher resolution but also preserve the identity of the iris. Mostofa et al. [109] incorporated a GAN-based photo-realistic super-resolution approach [93] to improve the resolution of LR iris images from the NIR domain before cross-matching the HR outputs with the HR images from the RGB domain. While these approaches showed improved performance, dealing with noisy data in such cases as iris at a distance and on the move could require the quality of an input iris image to be included in the super-resolution process [114]. In addition, Nguyen et al. argued that a fundamental difference exists between conventional super-resolution motivations and those required for biometrics, hence proposing to perform super-resolution at the feature level targeting explicitly the representation used by recognition [117].

9.5 Privacy in Deep Learning-based Iris Recognition

Privacy is becoming a key issue in computer vision and machine learning domains. In particular, it is accepted that the accuracy attained by deep learning models depends on the availability of large amounts of visual data, which stresses the need for privacy-preserving recognition solutions.

In short, the goal of privacy-preserving deep learning is to appropriately train models while preserving the privacy of the training datasets. While the utility of this kind of solution is obvious, there are certain concerns about the training data that supported the model creation, as the collection of images from a large number of individuals comes with significant privacy risks. In

particular, it should be considered that the subjects from whom the data were collected can neither delete nor control what actually will be learned from their data.

As with most of the existing biometric technologies, DL-based iris recognition poses challenges to privacy, which are even more concerning, considering the *data-driven* feature of such kind of systems. Particular attention should be paid to avoid function creep, guaranteeing that the system yielding from a set of data is not used for a different purpose than the originally communicated to the individual at the time of providing their information. Covert collection is another major concern, which is also particularly important for the iris trait, according to the possibility of being imaged from large distances and in a surreptitious way.

Particular attention has been paid to the development of fair recognition systems, in the sense that these kinds of systems should attain similar effectiveness in different subgroups of the population, regarding different features such as *gender*, *age*, *race*, or *ethnicity*. For data-driven systems, this might be a relevant challenge, considering that most of the existing datasets that support the learned systems have evident biases with regard to the subjects' characteristics above.

Lastly, in a more general machine learning perspective, potential attacks on the learned models have been concerning the research community and have been the scope of various recent works, attempting to provide defense mechanisms against: (i) model inversion attacks, that aim at reconstructing the training data from the model parameters (e.g., [87] and [67]); (ii) membership inference, that attempt to infer whether one individual was part of a training set (e.g., [75] and [153]); and (iii) training data extraction attacks, that aim at recovering individual training samples by querying the models (e.g, [86] and [39]).

9.6 Deep Learning-based Iris Segmentation

Being one of the earliest phases of the recognition process, segmentation is known as one of the most challenging, as it is at the front line for facing the dynamics of the data acquisition environments. This is particularly true, in the case of less constrained data acquisition protocols, where the resulting data have highly varying features and the particular conditions of each environment strongly determine the most likely data covariates.

In the segmentation context, the main challenge remains as the development of methods robust to *cross-domain* settings, i.e., able to segment the iris region for a broad range of image features, e.g., in terms of (1) illumination, (2) scale, (3) gaze, (4) occlusions, (5) rotation and (6) pose, corresponding to the acquisition in very different environments. Over the past decades, many research groups have been devoting their attention to improving the robustness of iris segmentation, which is known to be a primary factor for the final effectiveness of the recognition process. In this timeline, the proposed segmentation methods can be roughly grouped into three categories: (1) boundary-based methods (using the integro-differential operator or Hough transform); (2) based on handcrafted features (particularly suited for non-cooperative recognition, e.g., [160] and [159]); and (3) DL-based solutions.

For the latter family of methods, the emerging trends are closely related to the *general* challenges of DL-based segmentation frameworks, namely to obtain interpretable models that allow us to perceive what exactly these systems learning or the minimal neural architecture that guarantees a predefined level of accuracy. Also, the development of weakly supervised or even unsupervised frameworks is another *grand-challenge*, as it is accepted that such systems will likely adapt better to previously unseen data acquisition conditions. Finally, the computational cost of segmentation (both in terms of space and time) is another concern, with a special impact on the deployment of this kind of framework in mobile settings, and in the IoT setting [140].

9.7   Deep Learning-based Iris Recognition in Visible Wavelengths

Being a topic of study for over a decade (e.g., [99] and [129]), iris recognition in visible wavelengths remains essentially an interesting possibility for delivering biometric recognition from large distances (in conditions that are typically associated to visual surveillance settings) and in hand-held *commercial* devices, such as smartphones.

The emerging trends in this scope regard the development of alternate ways to analyze the multi-spectral information available in visible light data (typically RGB), i.e., by developing deep learning architectures optimized for fusion, either at the data, feature, score, or decision levels [11].

In the visual surveillance setting, the main challenge regards the development of optimized data acquisition settings, profiting from the advances in remote sensing technologies, that should be able to augment the quality (e.g., resolution and sharpness) of the obtained irises. In this scope, the research on active data acquisition technologies (based on PTZ devices, or similar) might also be an interesting emerging possibility [66].

10   CONCLUSIONS

Motivated by the tremendous success of DL-based solutions for many different solutions to every-day problems, machine learning is entering one of its golden eras, attracting growing interest from the research, commercial, and governmental communities. In short, deep learning uses multiple layers to represent the abstractions of data to build computational models that—even in a bit surprising way—typically surpass the previous generation of handcrafted-based automata. However, being extremely data-driven, the effectiveness of DL-based solutions is typically constrained by the existence of massive amounts of data, annotated consistently.

As in the generality of the computer-vision topics, a myriad of DL-based techniques has been proposed over the last years to perform biometric recognition, and—in particular—iris recognition. Nowadays, the existing methods cover the whole phases of the typical processing chain, from the preprocessing, segmentation, and feature extraction up to the matching and recognition steps.

Accordingly, this article provides the first comprehensive review of the historical and state-of-the-art approaches in DL-based techniques for iris recognition, followed by an in-depth analysis of pivoting and groundbreaking advances in each phase of the processing chain. We summarize and critically compare the most relevant methods for iris acquisition, segmentation, quality assessment, feature encoding, matching, and recognition problems, also presenting the most relevant open problems for each phase.

Finally, we review the typical issues faced in DL-based methods in this domain of expertise, such as unsupervised learning, black-box models, and online learning and illustrate how these challenges can be important to open prolific future research paths and solutions.

REFERENCES

[1] A. Agarwal, A. Noore, M. Vatsa, and R. Singh. 2022. Enhanced iris presentation attack detection via contraction-expansion CNN. *Pattern Recognition Letters* 159 (2022), 61–69.

[2] A. Agarwal, A. Noore, M. Vatsa, and R. Singh. 2022. Generalized contact lens iris presentation attack detection. *IEEE Transactions on Biometrics, Behavior, and Identity Science* 4 (2022), 1–1.

[3] S. Ahmad and B. Fuller. 2019. ThirdEye: Triplet based iris recognition without normalization. In *IEEE International Conference on Biometrics: Theory Applications and Systems (BTAS)*. 1–9.

[4] F. Alonso-Fernandez and J. Bigun. 2016. A survey on periocular biometrics research. *Pattern Recognition Letters* 82 (2016), 92–105.

[5] F. Alonso-Fernandez, J. Fierrez, and J. Ortega-Garcia. 2012. Quality measures in biometric systems. *IEEE Security and Privacy* 10, 6 (2012), 52–62.

[6] F. Alonso-Fernandez, K. Hernandez-Diaz, S. Ramis, F. J. Perales, and J. Bigun. 2021. Facial masks and soft-biometrics: Leveraging face recognition CNNs for age and gender prediction on mobile ocular images. *IET Biometrics* 10, 5 (2021), 562–580.

[7] M. Anisetti, Y. -H. Li, P. -J. Huang, and Y. Juan. 2019. An efficient and robust iris segmentation algorithm using deep learning. *Mobile Information Systems* 2019 (2019), 4568929.

[8] S. Arora and M. P. S. Bhatia. 2020. Presentation attack detection for iris recognition using deep learning. *International Journal of System Assurance Engineering* 11 (2020), 232238.

[9] V. Badrinarayanan, A. Kendall, and R. Cipolla. 2017. SegNet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39, 12 (2017), 2481–2495.

[10] A. Banerjee, C. Ghosh, and S. N. Mandal. 2022. Analysis of v-net architecture for iris segmentation in unconstrained scenarios. *SN Computer Science* 3 (2022), 208–232.

[11] B. Bigdeli, P. Pahlavani, and H. A. Amirkolaee. 2021. An ensemble deep learning method as data fusion system for remote sensing multisensor classification. *Applied Soft Computing* 110 (2021), 107563.

[12] D. S. Bolme, R. A. Tokola, C. B. Boehnen, T. B. Saul, K. A. Sauerwein, and D. W. Steadman. 2016. Impact of environmental factors on biometric matching during human decomposition. In *IEEE International Conference on Biometrics: Theory Applications and Systems (BTAS)*. IEEE, USA, 1–8.

[13] F. Boutros, N. Damer, K. Raja, R. Ramachandra, F. Kirchbuchner, and A. Kuijper. 2020. On benchmarking iris recognition within a head-mounted display for AR/VR applications. In *IEEE International Joint Conference on Biometrics (IJCB)*. USA.

[14] A. Boyd, K. Bowyer, and A. Czajka. 2022. Human-aided saliency maps improve generalization of deep learning. In *IEEE Winter Conference on Applications of Computer Vision (WACV)*.

[15] A. Boyd, A. Czajka, and K. Bowyer. 2019. DL-based feature extraction in iris recognition: Use existing models, fine-tune or train from scratch?. In *IEEE International Conference on Biometrics: Theory Applications and Systems (BTAS)*. 1–9.

[16] A. Boyd, Z. Fang, A. Czajka, and K. Bowyer. 2020. Iris presentation attack detection: Where are we now? *Pattern Recognition Letters* 138 (2020), 483–489.

[17] Aidan Boyd, Patrick Tinsley, Kevin W. Bowyer, and Adam Czajka. 2023. CYBORG: Blending human saliency into the loss improves deep learning-based synthetic face detection. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. 6108–6117.

[18] A. Boyd, S. Yadav, T. Swearingen, A. Kuehlkamp, M. Trokielewicz, E. Benjamin, P. Maciejewicz, D. Chute, A. Ross, P. Flynn, K. Bowyer, and A. Czajka. 2020. Post-mortem iris recognition—a survey and assessment of the state-of-the-art. *IEEE Access* 8 (2020), 136570–136593.

[19] C. Chen and A. Ross. 2021. An explainable attention-guided iris presentation attack detector. In *IEEE Workshop on Applications of Computer Vision (WACV)*. 97–106.

[20] J. Chen, F. Shen, D. Z. Chen, and P. Flynn. 2016. Iris recognition based on human-interpretable features. *IEEE Transactions on Information Forensics and Security* 11, 7 (2016), 1476–1485.

[21] L. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. Yuille. 2017. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 4 (2017), 834 –848.

[22] Y. Chen, W. Wang, Z. Zeng, and Y. Wang. 2019. An adaptive CNNs technology for robust iris segmentation. *IEEE Access* 7 (2019), 64517–64532.

[23] Y. Chen, C. Wu, and Y. Wang. 2020. T-center: A novel feature extraction approach towards large-scale iris recognition. *IEEE Access* 8 (2020), 32365–32375.

[24] Y. Chen, Z. Zeng, and F. Hu. 2019. End to end robust recognition method for iris using a dense deep convolutional neural network. In *Biometric Recognition*. Springer, Cham, 364–375.

[25] M. Choudhary, V. Tiwari, and Venkanna U. 2021. Ensuring secured iris authentication for mobile devices. In *IEEE International Conference on Consumer Electronics (ICCE)*. 1–5.

[26] M. Choudhary, V. Tiwari, and U. Venkanna. 2022. Iris liveness detection using fusion of domain-specific multiple BSIF and DenseNet features. *IEEE Transactions on Cybernetics* 52, 4 (2022), 2370–2381.

[27] M. Choudhary, V. Tiwari, and U. Venkanna. 2022. Identifying discriminatory feature-vectors for fusion-based iris liveness detection. *Journal of Ambient Intelligence and Humanized Computing* 14 (2022), 10605–10616.

[28] A. Czajka. 2013. Database of iris printouts and its application: Development of liveness detection method for iris recognition. In *International Conference on Methods and Models in Automation and Robotics (MMAR)*. 28–33.

[29] A. Czajka. 2021. Iris recognition designed for post-mortem and diseased eyes. Retrieved 10 october 2022 from https://github.com/aczajka/iris-recognition---pm-diseased-human-driven-bsif

[30] A. Czajka and K. Bowyer. 2018. Presentation attack detection for iris recognition: An assessment of the state-of-the-art. *ACM Computing Surveys* 51, 4, Article 86 (2018), 35 pages.

[31] A. Czajka, Z. Fang, and K. Bowyer. 2019. Iris presentation attack detection based on photometric stereo features. In *IEEE Winter Conference on Applications of Computer Vision (WACV)*. 877–885.

[32] N. Damer, K. Dimitrov, A. Braun, and A. Kuijper. 2019. On learning joint multi-biometric representations by deep fusion. In *IEEE International Conference on Biometrics: Theory Applications and Systems (BTAS)*. 1–8.

[33] P. Das, J. Mcfiratht, Z. Fang, A. Boyd, G. Jang, A. Mohammadi, S. Purnapatra, D. Yambay, S. Marcel, M. Trokielewicz, P. Maciejewicz, K. Bowyer, A. Czajka, S. Schuckers, J. Tapia, S. Gonzalez, M. Fang, N. Damer, F. Boutros, A. Kuijper, R. Sharma, C. Chen, and A. Ross. 2020. Iris liveness detection competition (LivDet-Iris) - the 2020 edition. In *IEEE International Joint Conference on Biometrics (IJCB)*. 1–9.

[34] J. Daugman. 1993. High confidence visual recognition of persons by a test of statistical independence. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 15, 11 (1993), 1148–1161.

[35] J. Daugman. 2007. New methods in iris recognition. *IEEE Transactions on Systems, Man and Cybernetics* 37 (2007), 1167–1175.

[36] J. Daugman. 2016. Information theory and the IrisCode. *IEEE Transactions on Information Forensics and Security* 11 (2016), 400–409.

[37] J. Daugman. 2021. Collision avoidance on national and global scales: Understanding and using big biometric entropy. *TechRxiv* 2021 (2021), 1–4.

[38] M. De Marsico, M. Nappi, D. Riccio, and H. Wechsler. 2015. Mobile iris challenge evaluation (MICHE)-I, biometric iris dataset and protocols. *Pattern Recognition Letters* 57 (2015), 17–23.

[39] X. Ding, H. Fang, Z. Zhang, K. -K. R. Choo, and H. Jin. 2022. Privacy-preserving feature extraction via adversarial training. *IEEE Transactions on Knowledge and Data Engineering* 34, 4 (2022), 1967–1979.

[40] W. Dong, Z. Sun, and T. Tan. 2009. A design of iris recognition system at a distance. In *Chinese Conference on Pattern Recognition*. 1–5.

[41] J. S. Doyle and K. Bowyer. 2015. Robust detection of textured contact lenses in iris recognition using BSIF. *IEEE Access* 3 (2015), 1672–1683.

[42] J. S. Doyle, K. Bowyer, and P. Flynn. 2013. Variation in accuracy of textured contact lens detection based on sensor and lens pattern. In *IEEE International Conference on Biometrics: Theory Applications and Systems (BTAS)*. 1–7.

[43] T. Elsken, J. H. Metzen, and F. Hutter. 2019. Neural architecture search: A survey. *Journal of Machine Learning Research* 20, 55 (2019), 1997–2017.

[44] M. Fang, N. Damer, Fadi Boutros, F. Kirchbuchner, and A. Kuijper. 2020. Deep learning multi-layer fusion for an accurate iris presentation attack detection. In *IEEE International Conference on Information Fusion (FUSION)*. 1–8.

[45] M. Fang, N. Damer, F. Boutros, F. Kirchbuchner, and A. Kuijper. 2021. Cross-database and cross-attack iris presentation attack detection using micro stripes analyses. *Image and Vision Computing* 105 (2021), 104057.

[46] M. Fang, N. Damer, F. Boutros, F. Kirchbuchner, and A. Kuijper. 2021. Iris presentation attack detection by attention-based and deep pixel-wise binary supervision network. In *IEEE International Joint Conference on Biometrics (IJCB)*. 1–8.

[47] M. Fang, N. Damer, F. Boutros, F. Kirchbuchner, and A. Kuijper. 2021. The overlapping effect and fusion protocols of data augmentation techniques in iris PAD. *Machine Vision and Applications* 33 (2021), 8–29.

[48] M. Fang, N. Damer, F. Kirchbuchner, and A. Kuijper. 2020. Micro stripes analyses for iris presentation attack detection. In *IEEE International Joint Conference on Biometrics (IJCB)*. 1–10.

[49] M. Fang, N. Damer, F. Kirchbuchner, and A. Kuijper. 2021. Demographic bias in presentation attack detection of iris recognition systems. In *28th European Signal Processing Conference (EUSIPCO)*. 835–839.

[50] Z. Fang and A. Czajka. 2020. Open source iris recognition hardware and software with presentation attack detection. In *IEEE International Joint Conference on Biometrics (IJCB)*. 1–8.

[51] Z. Fang, A. Czajka, and K. Bowyer. 2021. Robust iris presentation attack detection fusing 2D and 3D information. *IEEE Transactions on Information Forensics and Security* 16 (2021), 510–520.

[52] FBI Criminal Justice Information Services (CJIS) Division. 2021. Next Generation Identification (NGI). Retrieved 2021 from https://www.fbi.gov/services/cjis/fingerprints-and-other-biometrics/ngi

[53] J. Fierrez, J. Ortega-Garcia, D. Torre Toledano, and J. Gonzalez-Rodriguez. 2007. Biosec baseline corpus: A multimodal biometric database. *Pattern Recognition* 40, 4 (2007), 1389–1392.

[54] J. Galbally and M. Gomez-Barrero. 2016. A review of iris anti-spoofing. In *4th International Conference on Biometrics and Forensics (IWBF)*. 1–6.

[55] Y. Ganeeva and E. Myasnikov. 2020. Using convolutional neural networks for segmentation of iris images. In *International Multi-Conference on Industrial Engineering and Modern Technologies (FarEastCon)*. 1–4.

[56] A. Gangwar and A. Joshi. 2016. DeepIrisNet: Deep iris representation with applications in iris recognition and cross-sensor iris recognition. In *IEEE International Conference on Image Processing*. 2301–2305.

[57] A. Gangwar, A. Joshi, P. Joshi, and R. Raghavendra. 2019. DeepIrisNet2: Learning deep-iriscodes from scratch for segmentation-robust visible wavelength and near infrared iris recognition. arXiv:1902.05390. Retrieved from https://arxiv.org/abs/1902.05390

[58] S. J. Garbin, Y. Shen, I. Schuetz, R. Cavin, G. Hughes, and S. S. Talathi. 2019. OpenEDS: Open eye dataset. arXiv:1905.03702. Retrieved from https://arxiv.org/abs/1905.03702

[59] G. Gautam, A. Raj, and S. Mukhopadhyay. 2022. Deep supervised class encoding for iris presentation attack detection. *Digital Signal Processing* 121 (2022), 103329.

[60] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. 2014. Generative adversarial nets. In *International Conference on Neural Information Processing Systems (NIPS)*. USA, 9.

[61] D. Gragnaniello, C. Sansone, G. Poggi, and L. Verdoliva. 2016. Biometric spoofing detection by a domain-aware convolutional neural network. In *International Conference on Signal-Image Technology and Internet-Based Systems (SITIS)*.

[62] P. Grother and E. Tabassi. 2007. Performance of biometric quality measures. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29 (2007), 531–543.

[63] M. Gupta, S. Singh, A. Agarwal, M. Vatsa, and R. Singh. 2021. Generalized iris presentation attack detection algorithm under cross-database settings. In *International Conference on Pattern Recognition (ICPR)*. 5318–5325.

[64] P. Gupta, S. Behera, M. Vatsa, and R. Singh. 2014. On iris spoofing using print attack. In *International Conference on Pattern Recognition (ICPR)*. 1681–1686.

[65] A. Hafner, P. Peer, Ž. Emeršič, and M. Vitek. 2021. Deep iris feature extraction. In *International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*. 258–262.

[66] Y. Han. 2021. Design of an active infrared iris recognition device. In *IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC)*. 435–437.

[67] Y. He, G. Meng, K. Chen, X. Hu, and J. He. 2022. Towards security threats of deep learning systems: A survey. *IEEE Transactions on Software Engineering* 48, 5 (2022), 1743–1770.

[68] Z. He, T. Tan, Z. Sun, and X. Qiu. 2009. Toward accurate and fast iris segmentation for iris biometrics. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31, 9 (2009), 1670–1684.

[69] Pontus Hedman, Vasilios Skepetzis, Kevin Hernandez-Diaz, Josef Bigun, and Fernando Alonso-Fernandez. 2022. On the effect of selfie beautification filters on face detection and recognition. *Pattern Recognition Letters* 163, (2022), 104–111.

[70] K. Hernandez-Diaz, F. Alonso-Fernandez, and J. Bigun. 2018. Periocular recognition using CNN features off-the-shelf. In *International Conference of the Biometrics Special Interest Group (BIOSIG)*. 1–5.

[71] K. Hernandez-Diaz, F. Alonso-Fernandez, and J. Bigun. 2020. Cross-spectral periocular recognition with conditional adversarial networks. In *IEEE International Joint Conference on Biometrics (IJCB)*. 1–9.

[72] H. Hofbauer, E. Jalilian, and A. Uhl. 2019. Exploiting superior CNN-based iris segmentation for better recognition accuracy. *Pattern Recognition Letters* 120 (2019), 17–23.

[73] A. Howard, M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan, Q. V. Le, and H. Adam. 2019. Searching for MobileNetV3. In *IEEE International Conference on Computer Vision (ICCV)*.

[74] C. -S. Hsiao and C. -P. Fan. 2021. EfficientNet based iris biometric recognition methods with pupil positioning by u-net. In *3rd International Conference on Computer Communication and the Internet (ICCCI)*. 1–5.

[75] L. Hu, J. Li, G. Lin, S. Peng, Z. Zhang, Y. Zhang, and C. Dong. 2022. Defending against membership inference attacks with high utility by GAN. *IEEE Transactions on Dependable and Secure Computing* 20 (2022), 1–1.

[76] V. T. Huynh, S. -H. Kim, G. -S. Lee, and H. -J. Yang. 2019. Eye semantic segmentation with a lightweight model. In *IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*. 3694–3697.

[77] H. Hwang and E. C. Lee. 2020. Near-infrared image-based periocular biometric method using convolutional neural network. *IEEE Access* 8 (2020), 158612–158621.

[78] A. Jain, D. Deb, and J. Engelsma. 2021. Biometrics: Trust, but verify. *IEEE Transactions on Biometrics, Behavior, and Identity Science* 4 (2021), 1–1.

[79] E. Jalilian, M. Karakaya, and A. Uhl. 2020. End-to-end off-angle iris recognition using CNN based iris segmentation. In *International Conference of the Biometrics Special Interest Group (BIOSIG)*. 1–7.

[80] E. Jalilian, G. Wimmer, A. Uhl, and M. Karakaya. 2022. Deep learning based off-angle iris recognition. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. 4048–4052.

[81] P. A. Johnson, P. Lopez-Meyer, N. Sazonova, F. Hua, and S. Schuckers. 2010. Quality in face and iris research ensemble (Q-FIRE). In *IEEE International Conference on Biometrics: Theory Applications and Systems (BTAS)*. 1–6.

[82] Y. G. Jung, C. Y. Low, J. Park, and A. B. J. Teoh. 2020. Periocular recognition in the wild with generalized label smoothing regularization. *IEEE Signal Processing Letters* 27 (2020), 1455–1459.

[83] H. Kaur and R. Manduchi. 2020. EyeGAN: Gaze-preserving, mask-mediated eye image synthesis. In *IEEE Winter Conference on Applications of Computer Vision (WACV)*. 299–308.

[84] H. Kaur and R. Manduchi. 2021. Subject guided eye image synthesis with application to gaze redirection. In *IEEE Winter Conference on Applications of Computer Vision (WACV)*. 11–20.

[85] D. Kerrigan, M. Trokielewicz, A. Czajka, and K. Bowyer. 2019. Iris recognition with image segmentation employing retrained off-the-shelf deep neural networks. In *IEEE International Conference on Biometrics (ICB)*. 1–7.

[86] F. Khalid, M. A. Hanif, S. Rehman, R. Ahmed, and M. Shafique. 2019. TrISec: Training data-unaware imperceptible security attacks on deep neural networks. In *IEEE 25th International Symposium on On-Line Testing and Robust System Design (IOLTS)*. 188–193.

[87] M. Khosravy, K. Nakamura, Y. Hirose, N. Nitta, and N. Babaguchi. 2022. Model inversion attack by integration of deep generative models: Privacy-sensitive face generation from a face recognition system. *IEEE Transactions on Information Forensics and Security* 17 (2022), 357–372.

[88] M. C. Kim, J. H. Koo, S. W. Cho, N. R. Baek, and K. R. Park. 2018. Convolutional neural network-based periocular recognition in surveillance environments. *IEEE Access* 6 (2018), 57291–57310.

[89] N. Kohli, D. Yadav, M. Vatsa, and R. Singh. 2013. Revisiting iris recognition with color cosmetic contact lenses. In *IEEE International Conference on Biometrics (ICB)*. 1–7.

[90] N. Kohli, D. Yadav, M. Vatsa, R. Singh, and A. Noore. 2017. Synthetic iris presentation attack using iDCGAN. In *IEEE International Joint Conference on Biometrics (IJCB)*. 674–680.

[91] A. Kuehlkamp, A. Boyd, A. Czajka, K. Bowyer, P. Flynn, D. Chute, and E. Benjamin. 2022. Interpretable deep learning-based forensic iris segmentation and recognition. In *2nd WACV Workshop on Explainable and Interpretable Artificial Intelligence for Biometrics (XAI4B)*. IEEE, USA, 1–8.

[92] A. Kumar and A. Passi. 2010. Comparison and combination of iris matchers for reliable personal authentication. *Pattern Recognition* 43, 3 (2010), 1016–1026.

[93] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi. 2017. Photo-realistic single image super-resolution using a generative adversarial network. In *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*. 105–114.

[94] Y. -H. Li, W. R. Putri, M. S. Aslam, and C. -C. Chang. 2021. Robust iris segmentation algorithm in non-cooperative environments using interleaved residual u-net. *Sensors* 21, 4 (2021), 1–20.

[95] T. Liang, J. Glossner, L. Wang, S. Shi, and X. Zhang. 2021. Pruning and quantization for deep neural network acceleration: A survey. *Neurocomputing* 461 (2021), 370–403.

[96] Guosheng Lin, Anton Milan, Chunhua Shen, and Ian Reid. 2017. RefineNet: multi-path refinement networks for high-resolution semantic segmentation. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 5168–5177.

[97] N. Liu, M. Zhang, H. Li, Z. Sun, and T. Tan. 2016. DeepIris: Learning pairwise filter bank for heterogeneous iris verification. *Pattern Recognition Letters* 82 (2016), 154–161.

[98] W. Liu, A. Rabinovich, and A. Berg. 2016. ParseNet: Looking wider to see better. In *International Conference on Learning Representations (ICLR)*.

[99] X. Liu, Y. Bai, Y. Luo, Z. Yang, and Y. Liu. 2019. Iris recognition in visible spectrum based on multi-layer analogous convolution and collaborative representation. *Pattern Recognition Letters* 117 (2019), 66–73.

[100] J. Long, E. Shelhamer, and T. Darrell. 2015. Fully convolutional networks for semantic segmentation. In *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*. 3431–3440.

[101] J. Lozej, B. Meden, V. Struc, and P. Peer. 2018. End-to-end iris segmentation using u-net. In *IEEE International Work Conference on Bioinspired Intelligence (IWOBI)*. 1–6.

[102] Z. Luo, J. Li, and Y. Zhu. 2021. A deep feature fusion network based on multiple attention mechanisms for joint iris-periocular biometric recognition. *IEEE Signal Processing Letters* 28 (2021), 1060–1064.

[103] L. Ma, T. Tan, Y. Wang, and D. Zhang. 2003. Personal identification based on iris texture analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25, 12 (2003), 1519–1533.

[104] D. Menotti, G. Chiachia, A. Pinto, W. R. Schwartz, H. Pedrini, A. X. Falcão, and A. Rocha. 2015. Deep representations for iris, face, and fingerprint spoofing detection. *IEEE Transactions on Information Forensics and Security* 10, 4 (2015), 864–879.

[105] S. Minaee, A. Abdolrashidiy, and Y. Wang. 2016. An experimental study of deep convolutional features for iris recognition. *IEEE Signal Processing in Medicine and Biology Symposium*, 1–6.

[106] S. Mishra, P. Liang, A. Czajka, Danny Z. Chen, and X. Hu. 2019. CC-NET: Image complexity guided network compression for biomedical image segmentation. In *IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. 57–60.

[107] D. Monro, S. Rakshit, and D. Zhang. 2007. DCT-based iris recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29, 4 (2007), 586–595.

[108] D. Moreira, M. Trokielewicz, A. Czajka, K. Bowyer, and P. Flynn. 2019. Performance of humans in iris recognition: The impact of iris condition and annotation-driven verification. In *IEEE Winter Conference on Applications of Computer Vision (WACV)*. 941–949.

[109] M. Mostofa, S. Mohamadi, J. Dawson, and N. Nasrabadi. 2021. Deep GAN-based cross-spectral cross-resolution iris recognition. *IEEE Transactions on Biometrics, Behavior, and Identity Science* 3, 4 (2021), 443–463.

[110] A. Muron and J. Pospisil. 2000. The human iris structure and its usages. In *Acta Univ Plalcki Physica*. Acta Universitatis, 87–95.

[111] K. Nguyen, C. Fookes, R. Jillela, S. Sridharan, and A. Ross. 2017. Long range iris recognition: A survey. *Pattern Recognition* 72 (2017), 123–143.

[112] K. Nguyen, C. Fookes, A. Ross, and S. Sridharan. 2017. Iris recognition with off-the-shelf CNN features: A deep learning perspective. *IEEE Access* 6 (2017), 18848–18855. Invited Paper.

[113] K. Nguyen, C. Fookes, and S. Sridharan. 2020. Constrained design of deep iris networks. *IEEE Transactions on Image Processing* 29 (2020), 7166–7175.

[114] K. Nguyen, C. Fookes, S. Sridharan, and S. Denman. 2011. Quality-driven super-resolution for less constrained iris recognition at a distance and on the move. *IEEE Transactions on Information Forensics and Security* 6, 4 (2011), 1248–1258.

[115] K. Nguyen, C. Fookes, S. Sridharan, and A. Ross. 2022. Complex-valued iris recognition network. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45 (2022), 182–196.

[116] K. Nguyen, C. Fookes, S. Sridharan, M. Tistarelli, and M. Nixon. 2018. Super-resolution for biometrics: A comprehensive survey. *Pattern Recognition* 78 (2018), 23–42.

[117] K. Nguyen, S. Sridharan, S. Denman, and C. Fookes. 2012. Feature-domain super-resolution framework for gabor-based face and iris recognition. In *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*. 2642–2649.

[118] L. Nie, A. Kumar, and S. Zhan. 2014. Periocular recognition using unsupervised convolutional RBM feature learning. In *International Conference on Pattern Recognition (ICPR)*. 399–404.

[119] I. Nigam, M. Vatsa, and R. Singh. 2015. Ocular biometrics: A survey of modalities and fusion approaches. *Information Fusion* 26 (2015), 1–35.

[120] NIST. 2021. IEG: Iris Examiner Training Discussion. Retrieved 10 october 2022 from https://www.nist.gov/itl/iad/image-group/ieg-iris-examiner-training-discussion

[121] G. Odinokikh, M. Korobkin, I. Solomatin, I. Efimov, and A. Fartukov. 2019. Iris feature extraction and matching method for mobile biometric applications. In *IEEE International Conference on Biometrics (ICB)*. 1–6.

[122] L. Omelina, J. Goga, J. Pavlovicova, M. Oravec, and B. Jansen. 2021. A survey of iris datasets. *Image and Vision Computing* 108 (2021), 104109.

[123] J. Ortega, J. Fierrez, F. Alonso, J. Galbally, M. R. Freire, J. Gonzalez, C. Garcia, J. Alba, E. Gonzalez-Agulla, E. Otero, S. Garcia-Salicetti, L. Allano, B. Ly-Van, B. Dorizzi, J. Kittler, T. Bourlai, N. Poh, F. Deravi, Ming N. R. Ng, M. Fairhurst, J. Hennebert, A. Humm, M. Tistarelli, L. Brodo, J. Richiardi, A. Drygajlo, H. Ganster, F. M. Sukno, S. Pavani, A. Frangi, L. Akarun, and A. Savran. 2010. The multiscenario multienvironment biosecure multimodal database (BMDB). *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32, 6 (2010), 1097–1111.

[124] U. Park, R. Jillela, A. Ross, and A. Jain. 2011. Periocular biometrics in the visible spectrum. *IEEE Transactions on Information Forensics and Security* 6, 1 (2011), 96–106.

[125] L. Parzianello and A. Czajka. 2022. Saliency-guided textured contact lens-aware iris recognition. In *IEEE Workshop on Applications of Computer Vision (WACV)*. 330–337.

[126] H. Peng, B. Li, D. He, and J. Wang. 2020. End-to-end anti-attack iris location based on lightweight network. In *IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA)*. 821–827.

[127] P. J. Phillips, W. T. Scruggs, A. J. O'Toole, P. Flynn, K. Bowyer, C. L. Schott, and M. Sharpe. 2010. FRVT 2006 and ICE 2006 large-scale experimental results. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32, 5 (2010), 831–846.

[128] Planet Biometrics. 2017. Homeland Advanced Recognition Technology (HART). Retrieved 10 october 2022 from http://www.planetbiometrics.com/article-details/i/5598/desc/dhs-launches-rfp-for-hart/

[129] H. Proença. 2013. *Iris Recognition in the Visible Wavelength*. Springer, UK.

[130] H. Proença and L. Alexandre. 2005. UBIRIS: A noisy iris image database. In *Image Analysis and Processing – ICIAP 2005*. Springer, Germany, 970–977.

[131] H. Proenca, S. Filipe, R. Santos, J. Oliveira, and L. Alexandre. 2010. The UBIRIS.v2: A database of visible wavelength iris images captured on-the-move and at-a-distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32, 8 (2010), 1529–1535.

[132] H. Proença and J. C. Neves. 2017. IRINA: Iris recognition (even) in inaccurately segmented data. In *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*. 6747–6756.

[133] H. Proença. 2010. Iris recognition: On the segmentation of degraded images acquired in the visible wavelength. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32, 8 (2010), 1502–1516.

[134] H. Proença and J. Neves. 2018. Deep-PRWIS: Periocular recognition without the iris and sclera using deep learning frameworks. *IEEE Transactions on Information Forensics and Security* 13, 4 (2018), 888–896.

[135] H. Proença and J. C. Neves. 2019. Segmentation-less and non-holistic deep-learning frameworks for iris recognition. In *IEEE International Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 2296–2305.

[136] G. Quinn, J. Matey, P. Grother, and E. Watters. 2021. Statistics of visual features in the human Iris. *NIST Interagency/Internal Report (NISTIR), National Institute of Standards and Technology*, Gaithersburg, MD, [online], https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=932452 (Accessed March 19, 2024).

[137] E. Ribeiro, A. Uhl, and F. Alonso-Fernandez. 2019. Iris super-resolution using CNNs: Is photo-realism important to iris recognition? *IET Biometrics* 8 (2019), 69–78.

[138] E. Ribeiro, A. Uhl, F. Alonso-Fernandez, and R. A. Farrugia. 2017. Exploring deep learning image super-resolution for iris recognition. In *25th European Signal Processing Conference (EUSIPCO)*. 2176–2180.

[139] O. Ronneberger, P.Fischer, and T. Brox. 2015. U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Springer, 234–241.

[140] I. Saleh. 2018. *Internet of Things (IoT): Concepts, Issues, Challenges, and Perspectives*. Wiley Data and Cybersecurity.

[141] A. Sansola. 2015. *Postmortem iris Recognition and its Application in Human Identification*. Master's thesis. Boston University, USA.

[142] M. Sardar, S. Banerjee, and S. Mitra. 2020. Iris segmentation using interactive deep learning. *IEEE Access* 8 (2020), 219322–219330.

[143] K. Sauerwein, T. B. Saul, D. W. Steadman, and C. B. Boehnen. 2017. The effect of decomposition on the efficacy of biometrics for positive identification. *Journal of Forensic Sciences* 62, 6 (2017), 1599–1602.

[144] T. Schlett, C. Rathgeb, and C. Busch. 2018. Multi-spectral iris segmentation in visible wavelengths. In *IEEE International Conference on Biometrics (ICB)*. 190–194.

[145] F. Schroff, D. Kalenichenko, and J. Philbin. 2015. FaceNet: A unified embedding for face recognition and clustering. In *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*. 815–823.

[146] A. Sequeira, H. Oliveira, J. Monteiro, J. Monteiro, and J. Cardoso. 2014. MobILive 2014 - mobile iris liveness detection competition. In *IEEE International Joint Conference on Biometrics (IJCB)*. 1–6.

[147] S. Shah and A. Ross. 2009. Iris segmentation using geodesic active contours. *IEEE Transactions on Information Forensics and Security* 4, 4 (2009), 824–836.

[148] A. Sharma, S. Verma, M. Vatsa, and R. Singh. 2014. On cross spectral periocular recognition. In *IEEE International Conference on Image Processing*. 5007–5011.

[149] D. Sharma and A. Selwal. 2021. On data-driven approaches for presentation attack detection in iris recognition systems. In *Recent Innovations in Computing*. Springer, Singapore, 463–473.

[150] R. Sharma and A. Ross. 2020. D-NetPAD: An explainable and interpretable iris presentation attack detector. In *IEEE International Joint Conference on Biometrics (IJCB)*. 1–10.

[151] R. Sharma and A. Ross. 2021. Viability of optical coherence tomography for iris presentation attack detection. In *International Conference on Pattern Recognition (ICPR)*. 6165–6172.

[152] F. Shen and P. Flynn. 2013. Using crypts as iris minutiae. In *Biometric and Surveillance Technology for Human and Activity Identification X*. 87120B.

[153] L. Song, R. Shokri, and P. Mittal. 2019. Membership inference attacks against adversarially robust deep learning models. In *IEEE Security and Privacy Workshops (SPW)*. 50–56.

[154] L. Stark, K. Bowyer, and S. Siena. 2010. Human perceptual categorization of iris texture patterns. In *IEEE International Conference on Biometrics: Theory Applications and Systems (BTAS)*. 1–7.

[155] Z. Sun, H. Zhang, T. Tan, and J. Wang. 2014. Iris image classification based on hierarchical visual codebook. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 36, 6 (2014), 1120–1133.

[156] M. S. Sunder and A. Ross. 2010. Iris image retrieval based on macro-features. In *International Conference on Pattern Recognition (ICPR)*. 1318–1321.

[157] E. Tabassi, P. Grother, and W. Salamon. 2011. IREX II - IQCE ::iris quality calibration and evaluation: performance of iris image quality assessment algorithms. *National Institute of Standards and Technology*, Gaithersburg, MD, [online], https://doi.org/10.6028/NIST.IR.7820 (Accessed March 19, 2024).

[158] V. Talreja, N. M. Nasrabadi, and M. C. Valenti. 2022. Attribute-based deep periocular recognition: Leveraging soft biometrics to improve periocular recognition. In *IEEE Winter Conference on Applications of Computer Vision (WACV)*.

[159] C.-W. Tan and A. Kumar. 2012. Unified framework for automated iris segmentation using distantly acquired face images. *IEEE Transactions on Image Processing* 21, 9 (2012), 4068–4079.

[160] T. Tan, Zh. He, and Z. Sun. 2010. Efficient and robust segmentation of noisy iris images for non-cooperative iris recognition. *Image and Vision Computing* 28, 2 (2010), 223–230.

[161] H. Tann, H. Zhao, and S. Reda. 2019. A resource-efficient embedded iris recognition system using fully convolutional networks. *ACM Journal on Emerging Technologies in Computing Systems* 16, 1 (2019), 1–23.

[162] J. Tapia, S. Gonzalez, and C. Busch. 2022. Iris liveness detection using a cascade of dedicated deep learning networks. *IEEE Transactions on Information Forensics and Security* 17 (2022), 42−52.

[163] ISO/IEC 30107-1. Information technology Biometric presentation attack detection Part 1: Framework. 2016. https://www.iso.org/standard/53227.html

[164] ISO/IEC 29794-6. Information technology Biometric sample quality Part 6: Iris image data. 2015. https://www.iso.org/standard/54066.html

[165] P. Tinsley, A. Czajka, and P. Flynn. 2021. This face does not exist... but it might be yours! identity leakage in generative models. In *IEEE Winter Conference on Applications of Computer Vision (WACV)*. 1319−1327.

[166] M. Trokielewicz and A. Czajka. 2018. Data-driven segmentation of post-mortem iris images. In *International Workshop on Biometrics and Forensics (IWBF)*. IEEE, Italy, 1−7.

[167] M. Trokielewicz, A. Czajka, and P. Maciejewicz. 2016. Human iris recognition in post-mortem subjects: Study and database. In *IEEE International Conference on Biometrics: Theory Applications and Systems (BTAS)*. IEEE, USA, 1−6.

[168] M. Trokielewicz, A. Czajka, and P. Maciejewicz. 2016. Post-mortem human iris recognition. In *IEEE International Conference on Biometrics (ICB)*. IEEE, Sweden, 1−6.

[169] M. Trokielewicz, A. Czajka, and P. Maciejewicz. 2018. Presentation attack detection for cadaver iris. In *IEEE International Conference on Biometrics: Theory Applications and Systems (BTAS)*. 1−10.

[170] M. Trokielewicz, A. Czajka, and P. Maciejewicz. 2019. Iris recognition after death. *IEEE Transactions on Information Forensics and Security* 14, 6 (2019), 1501−1514.

[171] M. Trokielewicz, A. Czajka, and P. Maciejewicz. 2020. Post-mortem iris recognition resistant to biological eye decay processes. In *IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, USA, 1−8.

[172] M. Trokielewicz, A. Czajka, and P. Maciejewicz. 2020. Post-mortem iris recognition with deep-learning-based image segmentation. *Image and Vision Computing* 94 (2020), 103866.

[173] M. Tygert, J. Bruna, S. Chintala, Y. LeCun, S. Piantino, and A. Szlam. 2016. A mathematical motivation for complex-valued convolutional networks. *Neural Computation* 28 (2016), 815−825.

[174] Unique Identification Authority of India. 2021. AADHAAR: https://uidai.gov.in/en/

[175] M. Vatsa, R. Singh, and A. Noore. 2008. Improving iris recognition performance using segmentation, quality enhancement, match score fusion, and indexing. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 38, 4 (2008), 1021−1035.

[176] C. Wang, J. Muhammad, Y. Wang, Z. He, and Z. Sun. 2020. Towards complete and accurate iris segmentation using deep multi-task attention network for non-cooperative iris recognition. *IEEE Transactions on Information Forensics and Security* 15 (2020), 2944−2959.

[177] C. Wang and Z. Sun. 2020. A benchmark for iris segmentation. *Journal of Computer Research and Development* 57, 2 (2020), 395−412.

[178] C. Wang, Y. Wang, B. Xu, Y. He, Z. Dong, and Z. Sun. 2020. A lightweight multi-label segmentation network for mobile iris biometrics. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. 1006−1010.

[179] Caiyong Wang, Jawad Muhammad, Yunlong Wang, Zhaofeng He, and Zhenan Sun. 2020. Towards complete and accurate iris segmentation using deep multi-task attention network for non-cooperative iris recognition. *IEEE Transactions on Information Forensics and Security* 15, (2020), 2944−2959.

[180] K. Wang and A. Kumar. 2019. Toward more accurate iris recognition using dilated residual features. *IEEE Transactions on Information Forensics and Security* 14, 12 (2019), 3233−3245.

[181] K. Wang and A. Kumar. 2021. Periocular-assisted multi-feature collaboration for dynamic iris recognition. *IEEE Transactions on Information Forensics and Security* 16 (2021), 866−879.

[182] L. Wang, K. Zhang, M. Ren, Y. Wang, and Z. Sun. 2020. Recognition oriented iris image quality assessment in the feature space. In *IEEE International Joint Conference on Biometrics (IJCB)*. 1−9.

[183] X. Wang, H. Zhang, J. Liu, L. Xiao, Z. He, L. Liu, and P. Duan. 2019. Iris image super resolution based on GANs with adversarial triplets. In *Chinese Conference on Biometric Recognition (LNCS)*. Switzerland, 346 − 53.

[184] Z. Wang, J. Chai, and S. Xia. 2021. Realtime and accurate 3D eye gaze capture with DCNN-based iris and pupil segmentation. *IEEE Transactions on Visualization and Computer Graphics* 27, 1 (2021), 190−203.

[185] Z. Wei, T. Tan, and Z. Sun. 2007. Nonlinear iris deformation correction based on gaussian model. In *Advances in Biometrics*. Springer, Germany, 780−789.

[186] X. Wu and L. Zhao. 2019. Study on iris segmentation algorithm based on dense u-net. *IEEE Access* 7 (2019), 123959−123968.

[187] L. Xiao, Z. Sun, R. He, and T. Tan. 2013. Coupled feature selection for cross-sensor iris recognition. In *IEEE International Conference on Biometrics: Theory Applications and Systems (BTAS)*. 1−6.

[188] Fisher Yu and Vladlen Koltun. 2016. Multi-scale context aggregation by dilated convolutions. In *International Conference on Learning Representations*.

[189] D. Yadav, N. Kohli, M. Vatsa, R. Singh, and A. Noore. 2019. Detecting textured contact lens in uncontrolled environment using DensePAD. In *IEEE International Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*.

[190] D. Yadav, N. Kohli, S. Yadav, M. Vatsa, R. Singh, and A. Noore. 2018. Iris presentation attack via textured contact lens in unconstrained environment. In *IEEE Winter Conference on Applications of Computer Vision (WACV)*. 503–511.

[191] S. Yadav, C. Chen, and A. Ross. 2019. Synthesizing iris images using RaSGAN with application in presentation attack detection. In *IEEE International Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 2422–2430.

[192] S. Yadav, C. Chen, and A. Ross. 2020. Relativistic discriminator: A one-class classifier for generalized iris presentation attack detection. In *IEEE Winter Conference on Applications of Computer Vision (WACV)*. 2624–2633.

[193] S. Yadav and A. Ross. 2021. CIT-GAN: Cyclic image translation generative adversarial network with application in iris presentation attack detection. In *IEEE Winter Conference on Applications of Computer Vision (WACV)*. 2411–2420.

[194] D. Yambay, B. Becker, N. Kohli, D. Yadav, A. Czajka, K. Bowyer, S. Schuckers, R. Singh, M. Vatsa, A. Noore, D. Gragnaniello, C. Sansone, L. Verdoliva, L. He, Y. Ru, H. Li, N. Liu, Z. Sun, and T. Tan. 2017. LivDet iris 2017 iris liveness detection competition 2017. In *IEEE International Joint Conference on Biometrics (IJCB)*. 733–741.

[195] Z. Yan, L. He, Y. Wang, Z. Sun, and T. Tan. 2021. Flexible iris matching based on spatial feature reconstruction. *IEEE Transactions on Biometrics, Behavior, and Identity Science* (2021), 1–1.

[196] K. Yang, Z. Xu, and J. Fei. 2021. DualSANet: Dual spatial attention network for iris recognition. In *IEEE Winter Conference on Applications of Computer Vision (WACV)*. 888–896.

[197] Y. Yuan, W. Chen, Y. Yang, and Z. Wang. 2020. In defense of the triplet loss again: Learning robust person re-identification with fast approximated triplet loss and label distillation. In *IEEE International Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 1454–1463.

[198] L. A. Zanlorensi, H. Proença, and D. Menotti. 2020. Unconstrained periocular recognition: Using generative deep learning frameworks for attribute normalization. In *Int. IEEE International Conference on Image Processing*. 1361–1365.

[199] H. Zhang, Y. Bai, H. Zhang, J. Liu, X. Li, and Z. He. 2021. Local attention and global representation collaborating for fine-grained classification. In *International Conference on Pattern Recognition (ICPR)*. 10658–10665.

[200] H. Zhang, Z. Sun, and T. Tan. 2010. Contact lens detection based on weighted LBP. In *International Conference on Pattern Recognition (ICPR)*. 4279–4282.

[201] Q. Zhang, H. Li, Z. He, and Z. Sun. 2016. Image super-resolution for mobile iris recognition. In *Biometric Recognition*. Springer, Cham, 399–406.

[202] Q. Zhang, H. Li, Z. Sun, Z. He, and T. Tan. 2016. Exploring complementary features for iris recognition on mobile devices. In *IEEE International Conference on Biometrics (ICB)*. 1–8.

[203] Q. Zhang, H. Li, Z. Sun, and T. Tan. 2018. Deep feature fusion for iris and periocular biometrics on mobile devices. *IEEE Transactions on Information Forensics and Security* 13, 11 (2018), 2897–2912.

[204] Q. Zhang, H. Li, M. Zhang, Z. He, Z. Sun, and T. Tan. 2015. Fusion of face and iris biometrics on mobile devices using near-infrared images. In *Biometric Recognition*. Springer, Cham, 569–578.

[205] W. Zhang, X. Lu, Y. Gu, Y. Liu, X. Meng, and J. Li. 2019. A robust iris segmentation scheme based on improved u-net. *IEEE Access* 7 (2019), 85082–85089.

[206] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia. 2017. Pyramid scene parsing network. In *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*. 6230–6239.

[207] T. Zhao, Y. Liu, G. Huo, and X. Zhu. 2019. A deep learning iris recognition method based on capsule network architecture. *IEEE Access* 7 (2019), 49691–49701.

[208] Z. Zhao and A. Kumar. 2017. Accurate periocular recognition under less constrained environment using semantics-assisted CNN. *IEEE Transactions on Information Forensics and Security* 12, 5 (2017), 1017–1030.

[209] Z. Zhao and A. Kumar. 2017. Towards more accurate iris recognition using deeply learned spatially corresponding features. In *IEEE International Conference on Computer Vision (ICCV)*. 3829–3838.

[210] Z. Zhao and A. Kumar. 2018. Improving periocular recognition by explicit attention to critical regions in deep neural network. *IEEE Transactions on Information Forensics and Security* 13, 12 (2018), 2937–2952.

[211] Z. Zhao and A. Kumar. 2019. A deep learning based unified framework to detect, segment and recognize irises using spatially corresponding features. *Pattern Recognition* 93 (2019), 546 – 557.

[212] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba. 2016. Learning deep features for discriminative localization. In *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*. 2921–2929.