

# Complex-Valued Iris Recognition Network

Kien Nguyen<sup>ID</sup>, *Member, IEEE*, Clinton Fookes<sup>ID</sup>, *Senior Member, IEEE*,  
Sridha Sridharan<sup>ID</sup>, *Life Senior Member, IEEE*, and Arun Ross<sup>ID</sup>, *Senior Member, IEEE*

**Abstract**—In this work, we design a fully complex-valued neural network for the task of iris recognition. Unlike the problem of general object recognition, where real-valued neural networks can be used to extract pertinent features, iris recognition depends on the extraction of both phase and magnitude information from the input iris texture in order to better represent its biometric content. This necessitates the extraction and processing of phase information that cannot be effectively handled by a real-valued neural network. In this regard, we design a fully complex-valued neural network that can better capture the multi-scale, multi-resolution, and multi-orientation phase and amplitude features of the iris texture. We show a strong correspondence of the proposed complex-valued iris recognition network with Gabor wavelets that are used to generate the classical IrisCode; however, the proposed method enables a new capability of automatic complex-valued feature learning that is tailored for iris recognition. We conduct experiments on three benchmark datasets - ND-CrossSensor-2013, CASIA-Iris-Thousand and UBIRIS.v2 - and show the benefit of the proposed network for the task of iris recognition. We exploit visualization schemes to convey how the complex-valued network, when compared to standard real-valued networks, extracts fundamentally different features from the iris texture.

**Index Terms**—Automatic complex-valued iris feature learning, data-driven iris recognition, complex-valued networks

## 1 INTRODUCTION

THE human iris is a powerful biometric pattern that has the potential to deliver high recognition accuracy at low false match rates. This is due to (i) the complex textural pattern of the iris that is believed to be unique to each eye, and (ii) the limited genetic penetrance of the iris texture [8], [29], [40]. The success of iris recognition - besides its attractive physical characteristics - is rooted in the development of efficient feature descriptors, especially the IrisCode introduced in Daugman's pioneering work [6], [8], [11] and many other descriptors that have subsequently evolved [20], [22], [26], [35], [46], [48].

The last few years have seen a transition in the iris recognition community to deep neural networks to take advantage of their *automatic feature learning* capability [24], [31], [52], [55]. This provides us with an alternative approach in feature design by automatically learning and discovering feature representations directly from data, eliminating some of the pitfalls in developing handcrafted features [1]. Despite the promise and a number of initial efforts in this direction, deep neural networks have not been exactly revolutionary in iris recognition. This could be because existing deep iris recognition networks in the literature are directly derived from

general deep learning theory for natural images. These approaches do not take into account some of the unique properties of the iris texture. Compared to the statistics of natural images as defined in [14], the iris texture itself is stochastic [10] without consistent shapes, edges, or stromal morphology. This makes it significantly different from non-stochastic and structured patterns occurring in object-based natural images. The *intrinsic differences between iris texture images and object-based natural images* requires automatic feature learning to be tailored with domain-specific knowledge in order for it to reach its full potential in the iris recognition setting.

The contribution of this paper stems from our investigation of the classical IrisCode. Technically speaking, we can split the encoding process of the IrisCode into two steps: (i) representing the iris texture image in the complex space by applying Gabor wavelets, and (ii) quantizing the phase of the complex representation to generate the final descriptor. While the first step emphasizes the importance of the complex-valued representation, the second accentuates the importance of phase information. The importance of both complex-valued representation and phase information has been reinforced in many other handcrafted approaches. For example, Kong *et al.* represented the iris texture image in the Gabor complex space and employed precise phase to encode the iris [22]. Monro *et al.* [27] and Miyazawa *et al.* [26] represented the iris texture image in the Cosine and Fourier spaces and then used the phase to encode the iris.

All existing deep networks proposed for iris recognition in the literature have not been able to incorporate this domain knowledge. First, they all operate in the real-valued space with real-valued operations and feature maps. Compared to a complex-valued representation, existing real-valued iris recognition networks may be limited in the space in which they can learn feature representations. The fact that they completely ignore the *complex-valued nature* of the iris feature representation may discard many of the benefits of

- Kien Nguyen, Clinton Fookes, and Sridha Sridharan are with the Image and Video Research Laboratory, SAIVT, School of Electrical Engineering and Computer Science, Queensland University of Technology, Brisbane, QLD 4000, Australia. E-mail: {k.nguyenthanh, c.fookes, s.sridharan}@qut.edu.au.
- Arun Ross is with the Department of Computer Science and Engineering, Michigan State University, East Lansing, MI 48824 USA. E-mail: rossarun@cse.msu.edu.

Manuscript received 21 May 2021; revised 9 January 2022; accepted 13 February 2022. Date of publication 24 February 2022; date of current version 5 December 2022.

(Corresponding author: Kien Nguyen.)

Recommended for acceptance by R. P. Wildes.

Digital Object Identifier no. 10.1109/TPAMI.2022.3152857

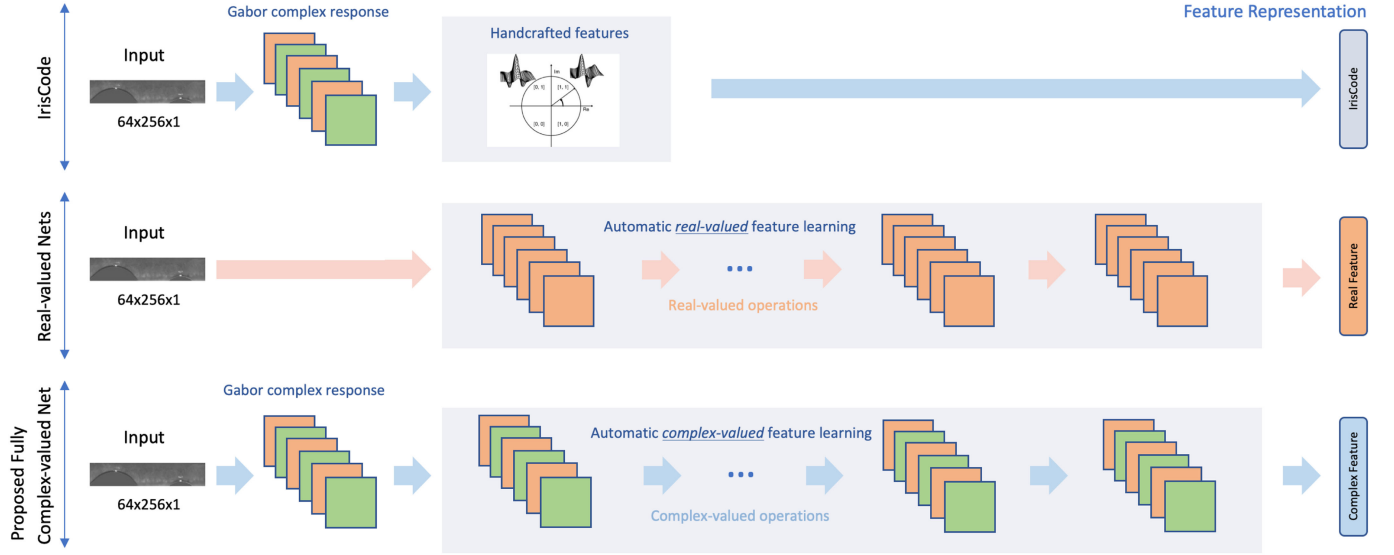


Fig. 1. The classic IrisCode depends on the complex-valued Gabor response followed by phase quantization, which is not learned from the data. Existing deep iris recognition networks perform automatic feature learning in the real-valued domain, which cannot adequately capture and retain phase information along the pipeline. This paper proposes to shift processing of the iris texture to a fully complex-valued space, which has a richer representation capacity, better leverages the unique characteristics of the iris texture, and has a strong correspondence with complex Gabor wavelets.

the complex-valued representation. Second, all existing deep iris recognition networks have no mechanism to explicitly retain the phase along the network pipeline. Explicitly modeling and processing phase information in the automatic feature learning process is expected to better deal with the unique properties of the iris texture.

Motivated by these insights, we propose to shift from standard automatic real-valued feature learning as in current deep iris networks to *automatic complex-valued feature learning*. On the one hand, the shift from a real-valued space to a complex-valued space will enrich the search space for the feature learning process. On the other hand, the space shift allows us to incorporate domain knowledge of iris recognition to tailor modern deep learning to take into account the unique properties of the iris texture.

Applying complex-valued networks in the iris recognition setting is not obvious, since the iris images themselves are real numbers, which explains why all existing deep iris networks have been designed and operated only in the real domain. We re-purpose the complex response of Gabor wavelets as the complex input to be further processed by a fully complex-valued network. This enables us to retain the complex-valued nature of the feature representation through the entire pipeline. When the complex-valued nature of the feature representation is retained, it is trivial to recover the phase information from the real and imaginary components. Compared to the classical IrisCode and its phase-based derivatives, the proposed fully complex-valued iris recognition network retains the benefits of operating in the complex space, but learns a feature representation in a data-driven manner as opposed to being human-designed. This is illustrated in Fig. 1.

The core contributions of this work can be divided into three parts.

- We propose to shift processing of the iris texture to the complex-valued space in deep neural networks. To the best of our knowledge, this is the first attempt

of its kind. The complex-valued concepts allow the network to better cater to the unique properties of iris features, explicitly retaining the phase information in the feature representation. However, it also requires fundamental changes to the network design and training process.

- We propose a novel fully complex-valued network to enable this shift. It enables, for the first time, automatic complex-valued feature learning in the iris recognition setting. Compared to standard automatic feature learning in standard real-valued networks, automatic complex-valued feature learning discovers more discriminative representations for the iris texture.
- We provide an insight that fully complex-valued networks have a solid mathematical and theoretical foundation to better suit iris recognition than standard real-valued networks. In addition, complex-valued networks show a strong correspondence with the classic IrisCode and its phase-based derivatives. This allows us to validate the optimality of the existing handcrafted feature representations.

The remainder of the paper is organized in 5 sections. Section 2 discusses the evolution of representation in iris recognition, from classical handcrafted approaches to modern deep learning and the need for tailored deep learning for iris recognition. Section 3 justifies why complex-valued networks better suit iris recognition than standard real-valued networks. Section 4 presents our proposed complex-valued iris recognition network. Section 5 describes our experimental results and the paper is concluded in Section 6.

## 2 REPRESENTATION EVOLUTION IN IRIS RECOGNITION

Over the last 20 years, representation of the iris texture for iris recognition has evolved from classic handcrafted features to modern representation learning using deep learning.

*Classic Handcrafted Features.* The richness and stochasticity of the iris texture lends itself to the application of Gabor wavelets with Hamming distance for comparing iris images [8]. Complex-valued Gabor responses capture the rich details of the iris texture in terms of both spectral and spatially localized properties [7]. Both coarsely-quantized phase [11] and precise phase [22] of the Gabor response can create stable bit streams for a given eye, which are, nevertheless, sufficiently different across different eyes. The efficiency of phase encoding in a complex-valued space is reinforced through many other descriptors that have subsequently evolved [20], [22], [26], [27], [48]. Miyazawa *et al.* and Monro *et al.* encoded phase components in 2D Discrete Fourier Transforms (DFTs) and 2D Discrete Cosine Transforms (DCTs) to extract a phase code for representing iris information [26], [27]. Similarly, Tan *et al.* encoded Zernike moments-based phase features which are computed from partially overlapping regions to more effectively accommodate local pixel region variations in the normalized iris images [48]. Departing from the complex-valued phase-based encoding, some notable works harness ordinal measures [46], compressive sensing theory [35] and deformation field [37] to represent irises.

The use of handcrafted features has certain advantages. One of the biggest advantages is the use of domain knowledge gained over a period of time to focus on the biological and physical processes constituting iris recognition. For example, algorithms such as IrisCode [5] are based on a deep understanding of the iris structure. Modern representation learning approaches using deep networks provide another perspective in designing iris features by automatically learning representations directly from data [3].

*Modern Representation Learning.* Deep networks have been employed to learn feature representations for iris images automatically and directly from data. There are two categories of deep networks proposed for iris recognition: classification networks and similarity networks. Classification iris networks employ deep architectures with a softmax loss to classify an input iris image into a list of known identities. Typical examples of these networks are [13], [30]. The main requirement of the softmax-based networks is that the test image has to belong to one of the classes in the training set, which means the networks will have to be re-trained whenever a new class is added. In contrast, similarity iris networks employ deep architectures with a pairwise loss to learn a metric representing how similar or dissimilar two iris images are without knowing their identities. Zhao *et al.* argued that classification networks may not be optimal for iris recognition since the iris texture is inherently stochastic and does not exhibit structural information or meaningful hierarchies [55]. Compared to the classification networks, similarity iris networks directly reflect what we want to achieve, i.e., to train the representation to correspond to iris (dis)similarity. This results in irises of the same subject having small distances and irises of different subjects having larger distances. Typical examples of these networks are [24], [31], [52], [55].

To fully exploit the capacity of deep learning in the iris recognition setting, it is required to tailor networks to suit the domain specific properties of the iris texture and for iris

recognition. This paper will propose a tailored deep learning approach for iris recognition using complex-valued networks.

### 3 WHY COMPLEX-VALUED NETWORKS FOR IRIS RECOGNITION?

Complex-valued networks originated in application domains where the input is complex-valued such as remote sensing [2], [54] and MRI fingerprinting [50]. Compared to standard real-valued networks, complex-valued networks offer three key distinctive advantages for iris recognition that can not be directly achieved by their standard real-valued counterparts.

*Richer Representational Capacity for Automatic Feature Learning.* One key advantage of neural networks is automatic feature learning, which employs hierarchical multi-layer networks to learn a feature representation directly from data [1]. The complex-valued space of complex-valued networks allows the learning algorithm to explore a richer and more versatile search space than the real-valued space of real-valued networks, potentially leading to the capability to learn more discriminative and informative representations for iris recognition.

*Better Leverage the Unique Characteristics of the Iris Texture.* Standard real-valued networks are designed to learn the appearance of objects through consistent shapes, edges, or other semantic structures [53]. However, the fact that the iris texture is stochastic [10] with no consistent shapes, edges or stromal morphology would make real-valued networks struggle to learn any meaningful semantic structures from the iris texture and unable to realize the full potential of automatic feature learning.

From a mathematical perspective, there are two key distinctive properties of complex-valued networks that are highly desirable in iris recognition.

- Sensitive to phase: Complex-valued networks retain and process complex-valued features, hence they are able to retain and recover phase information. This makes them sensitive to phase structure [17]. In iris recognition, phase is more important than magnitude [8]. Phase encoding has been the key to many classical handcrafted features [11], [22], [26], [27], [48]; however, standard real-valued networks have no direct mechanism to retain this. Complex-valued networks, in contrast, naturally allow us to directly and explicitly retain phase components due to its operation in a complex space.
- Locally stationary stochastic processes: The authors of [49] have provided a solid mathematical framework to prove that complex-valued networks learn representations that are invariant to scale, resolution and orientation variations. All three properties, i.e., invariant to multi-scale, multi-resolution and multi-orientation, are highly desirable to deal with variations in the iris texture.

*Correspondence With Gabor Wavelets.* In [49], the authors have provided a mathematical framework to prove that complex-valued networks can be viewed as data-driven nonlinear multiwavelet packets. Gabor wavelets, as the core of the IrisCode, can be easily approximated by a complex-valued network. As depicted in Fig. 1, the automatic complex feature

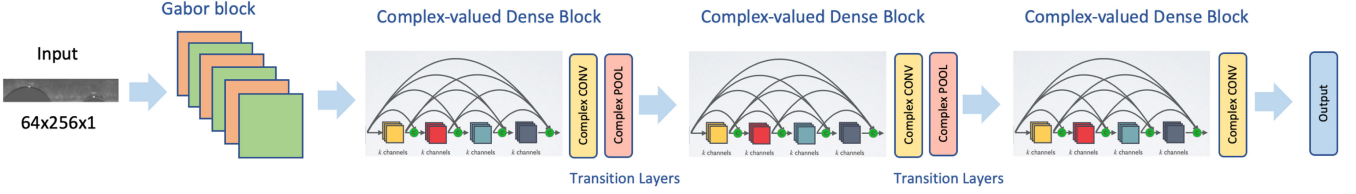


Fig. 2. ComplexIrisNet architecture. The normalized iris texture image is first fed through a Gabor Block, which generates the complex response to be subsequently processed by complex-valued operations. The following Dense Blocks and Transition Blocks are equipped with fully complex-valued operations, allowing them to process iris features in the complex-valued domain. The feature maps of the last CONV layer in the last Transition layer are used for feature representation.

learning block in the proposed complex-valued network can converge to the handcrafted features in the IrisCode and shrink a complex-valued network to the IrisCode or its other phase-based derivatives. This generalization capability, interestingly, allows us to validate the optimality of the handcrafted IrisCode. During the training process, the network weights are learned to best represent the iris features. If the IrisCode features are truly the best representation for the iris problem, the training should converge the proposed network to emulate those features and we should achieve accuracy on par with the IrisCode. However, as will be shown in Section 4, the better recognition results of the proposed network prove that handcrafted IrisCode representations are not always the optimal answer and a new discriminative and more informative feature representation can be deduced directly from the data by the proposed network. Standard real-valued networks do not obviously exhibit the same exact correspondence with data-driven wavelets [49], making them less effective in the iris feature learning.

In summary, our contribution to develop complex-valued networks for iris recognition is driven by multiple factors. First, our approach will provide a richer representational capacity through a complex representation which is data driven through automatic feature learning. Second, our approach is driven by the biological characteristics of the iris texture, which is stochastic in nature and lacking in consistent shapes, edges or semantic structures. Through provision of a network which is able to retain and recover phase information in the iris textural features, we can better exploit its intrinsic features for greater accuracy while making it invariant to scale, resolution and orientation variations. Third, by explicitly encoding phase information, we can directly compare our approach with both phase-based handcrafted approaches (such as the IrisCode) along with standard real-valued networks to validate their optimality for the iris recognition setting.

## 4 THE PROPOSED FULLY COMPLEX-VALUED IRIS RECOGNITION NETWORK

We propose ComplexIrisNet - a densely-connected fully-convolutional and complex-valued network architecture - for the iris recognition task. The proposed network is illustrated in Fig. 2. There are three major characteristics to highlight for the proposed network:

- **Fully complex-valued:** All network operations and feature maps are complex-valued. This brings it in line with the nature of the iris feature representation since the interaction between real and imaginary

components, such as phase, has long been acknowledged as important.

- **Fully convolutional:** The fully-connected layers at the end of a deep neural network typically do not preserve the spatial adjacency information present in the input image, unlike the earlier convolutional layers [44]. Therefore, we do not employ fully-connected layers in our network architecture.
- **Densely connected:** Each layer takes all preceding feature-maps as input. The compelling benefits of the densely connected mechanism are: it alleviates the vanishing-gradient problem, strengthens feature propagation, encourages feature re-use and substantially reduces the number of parameters [21]. This architecture is the state-of-the-art in large-scale ImageNet visual challenges, with the dual benefit of multiple skip connections of the ResNet architecture [16] and network-in-network connections of the inception architecture [47].

The architecture of the proposed network and complex-valued operations are discussed in Sections 4.1 and 4.2. Section 4.3 discusses the changes required to train and infer a complex-valued network.

### 4.1 Network Architecture

A normalized iris image is the input for the proposed network. The input is first fed into a Gabor Block, which embeds the input iris images into the complex space. The complex representations are subsequently processed by multiple Dense Blocks, which are densely connected and fully convolutional. Each Dense Block is followed by a Transition Block, which further encourages interaction within the representation and promotes compactness of the feature map. Output from the last CONV layer is used as a feature representation for the input iris.

*Stage 1: Input.* A normalized iris image with a size of  $64 \times 256$  is used as input for the network. This normalized iris image is generated in the pre-processing step. The pre-processing step is described in Section 5.2.

*Stage 2: Gabor Block.* Consists of a group of complex-valued convolutional filters. We explicitly initialize these filters with Gabor-like complex kernels to simulate a family of Gabor wavelet transform in iris recognition

$$g(x, y; \lambda, \theta, \psi, \delta, \gamma) = \exp\left(-\frac{x'^2 + y'^2}{2\delta^2}\right) \exp\left(i\left(2\pi\frac{x'}{\lambda} + \psi\right)\right), \quad (1)$$

where  $x' = x \cos(\theta) + y \sin(\theta)$  and  $y' = -x \sin(\theta) + y \cos(\theta)$ . Here,  $\theta$  is the orientation of the normal to the parallel stripes



of a Gabor function,  $\lambda$  is the wavelength of the sinusoidal factor,  $\delta$  is the sigma/standard deviation of the Gaussian envelop and  $\gamma$  is the spatial aspect ratio and specifies the ellipticity of the support of the Gabor function [11], [12]. The filters in this layer are initialized with different values of these five parameters to cover the family of the Gabor wavelet transform [8], [11]. With  $M$  filters, the output will be  $H \times W \times (2M)$  with  $M$  real and  $M$  imaginary components. These  $M$  Gabor-like filters will transform the input iris image into the complex space.

Some researchers have also tried to combine Gabor filters with CNNs. For example, Kwolek[23] extracted intrinsic facial Gabor features to feed to the subsequent CNNs. Sawar *et al.* replaced a number of certain weight kernels of a CNN with Gabor filters to reduce the number of parameters and the model complexity [41]. Luan *et al.* modulated convolutional filters via Gabor filter banks to improve the robustness to geometric transformation of CNNs [25]. Compared to [23], [25], [41], we take a different approach by employing the entirety of the complex-valued outputs after applying the Gabor filters instead of merely the real-valued components. More importantly, the complex-valued Gabor response is subsequently processed by a fully complex-valued network to retain the complex-valued nature of the representation through the entire network.

This Gabor Block has several interesting characteristics worth mentioning. Visualization of deep learning networks has shown that most networks learn edge detectors like the effect of Gabor filters in the first layer [1]. Interestingly, Gabor filters have also been very effective in the early stages of the pipeline for encoding the iris textures before other encoding operations such as the phase-quadrant quantization are engaged in the generation of IrisCode [6]. Hence, this block, besides its main purpose to embed the input into the complex domain for further processing, also models the prior in the classic handcrafted features. Even though the network may be arguably able to learn the first layer itself, explicitly steering the network architecture with a Gabor Block not only benefits in naturally embedding the input into the complex domain, but it also reduces the number of parameters to learn, which subsequently results in reducing the required amount of training data.

**Stage 3: Dense Block.** Consists of a number of complex-valued convolutional layers with dense connectivity. The dense connectivity within layers in these blocks enables the modeling of complicated relationships within the complex-valued feature maps. The real-valued dense blocks have been shown to be very effective in encoding visual details in the large-scale visual tasks [21]. Our dense blocks are specifically equipped with the complex-valued operations which will be discussed in Section 4.2.

A dense block is illustrated in Fig. 3. There are  $N$  layers in each dense block. Each layer  $x_i$  is fed through a combination of network operations as a function,  $H_{i+1}$ , to generate the next layer  $x_{i+1}$ . Six network operations are performed between layers:  $BN-ReLU-CONV(1 \times 1)-BN-ReLU-CONV(3 \times 3)$ , where, BN stands for complex-valued Batch Normalization, ReLU stands for complex-valued Rectified Linear Unit, and CONV stands for complex-valued Convolution. These operations will be discussed in Section 4.2. The dense connectivity is shown as the  $l^{th}$  layer receives the

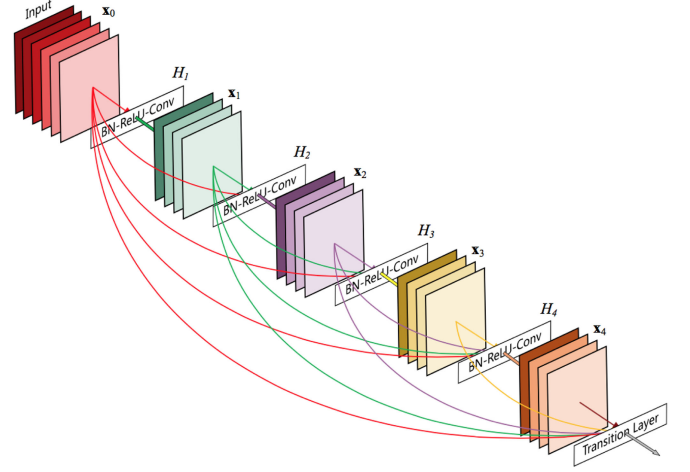


Fig. 3. Dense Block architecture. Each layer ( $x_l$ ) in the block receives the feature maps of all preceding layers ( $x_0, \dots, x_{l-1}$ ) as inputs. Six complex-valued network operations,  $BN-ReLU-CONV(1 \times 1)-BN-ReLU-CONV(3 \times 3)$ , are performed between two successive layers.

feature maps of all preceding layers,  $x_0, \dots, x_{l-1}$ , as input, i.e.,

$$x_l = H_l([x_0, x_1, \dots, x_{l-1}]). \quad (2)$$

The key benefit of the dense connectivity is allowing the signal to flow smoothly and mitigating the gradient vanishing and exploding effect, encouraging feature re-use, substantially reducing the number of parameters and thus enabling a deep visual representation [21].

Each Dense Block is followed by a Transition Block. A Transition Block contains one BN layer, one CONV layer, one ReLU layer and one  $2 \times 2$  Pooling layer. The purpose of the Transition Blocks is to improve the compactness of the feature maps and to encourage interaction within the representation.

**Stage 4: Output.** The feature maps from the last CONV layer of the last Transition Block are used as the feature representation for the input iris. Depending on the number of Dense Blocks, the size of the output will vary. One example of the architecture of the proposed complex-valued deep iris network, ComplexIrisNet, with three Dense Blocks and three Transition Blocks is presented in Table 1.

## 4.2 Complex-Valued Network Operations

The complex-valued network can be considered a generalization of its real-valued counterparts. While many concepts from real-valued networks can be generalized trivially, the lack of ordering of the complex field makes generalization of some concepts tricky [15]. Without total ordering, two general complex numbers are not comparable; specifically, the  $\min$  and  $\max$  operations are not defined. The ReLU layer, max pooling layer and the optimization problem itself all rely on these operators. This section discusses these non-trivial changes.

A complex number,  $z \in \mathbb{C}$ , consisting of a real component,  $x \in \mathbb{R}$ , and an imaginary component,  $y \in \mathbb{R}$ , is defined as follows:

$$z = x + iy \in \mathbb{C}, \quad \text{where } \sqrt{-1} = i. \quad (3)$$

TABLE 1  
Layer Configuration of the ComplexIrisNet

Blocks	Kernel Size	Output Size	#Layers	#Parameters
Input		64 × 256 × 1		
Gabor Block	(7×7) × 40	32 × 128 × 40	4	1,960
Dense Block1	1×1 3×3	32 × 128 × 168	36	205,824
Transition Block1	(1×1) × 32	16 × 64 × 32	4	5,376
Dense Block2	1×1 3×3	16 × 64 × 160	36	204,800
Transition Block2	(1×1) × 70	8 × 32 × 20	4	3,200
Total			84	421,160

The complex number can also be represented with a magnitude,  $r \in \mathbb{R}$ , and a phase,  $\theta \in \mathbb{R}$ , as  $z = re^{i\theta}$ . A complex function is defined as

$$f: \mathbb{C} \rightarrow \mathbb{C} \quad (4)$$

$$f(z) = u(z) + iv(z), \quad \text{where } u, v: \mathbb{R}^2 \rightarrow \mathbb{R}. \quad (5)$$

**Complex CONV.** A convolution operation of a feature map,  $W$ , with a convolutional filter,  $k$ , is presented as  $W * k$ , where  $*$  stands for the convolution operation. A complex-valued convolution operation of a complex-valued feature map,  $W = A + iB$ , with a complex-valued convolutional filter,  $k = x + iy$ , is also a complex-valued feature map,  $W * k = (A + iB) * (x + iy) = (A * x - B * y) + i(B * x + A * y)$ , since the convolution is distributive.

Mathematically, for one complex-valued convolutional layer  $l$ , the complex-valued input,  $I^{(l)}$ , which is the feature map output from the previous layer, is convolved with a set of complex-valued kernels,  $K^{(l)}$ , to generate a complex-valued output,  $O^{(l)}$ .

- Input:  $I^{(l)} \in \mathbb{C}^{W_1 \times H_1 \times (2C_1)}$
- Complex-valued convolutional kernels:  $K^{(l)} \in \mathbb{C}^{W_3 \times H_3 \times (2C_1) \times (2C_2)}$
- Output:  $O^{(l)} = I^{(l)} * K^{(l)} \in \mathbb{C}^{W_2 \times H_2 \times (2C_2)}$

The kernel set has  $C_2$  complex-valued kernels, each with a size of  $W_3 \times H_3 \times (2C_1)$ . The convolution of the input with one kernel is illustrated in Fig. 5. The output from convolving the input with  $C_2$  kernels are concatenated to generate the final output  $O^{(l)}$ . There are two more parameters in convolution, stride,  $S$ , and zero-padding size,  $P$ . These parameters are similar to their real-valued counterpart.

**Complex Activation.** Rectified Linear Unit (ReLU) is the most common activation function used. We generalize it to the complex domain as follows:

$$\text{ReLU}(z) = \begin{cases} z, & \text{if } \arg(z) \in [0, \frac{\pi}{2}], \\ 0, & \text{otherwise.} \end{cases} \quad (6)$$

**Complex Pooling.** Spectral Pooling is the best candidate in comparison with others (Average Pooling, Max Pooling) since it can handle complex numbers [39]. In addition, the pooling is performed by truncating the representation in the frequency domain. This approach preserves considerably more information per parameter than other pooling strategies and enables flexibility in the choice of dimensionality of the pooling output. This representation also enables a new form of stochastic regularization by randomized modification of resolution. This has been shown to achieve

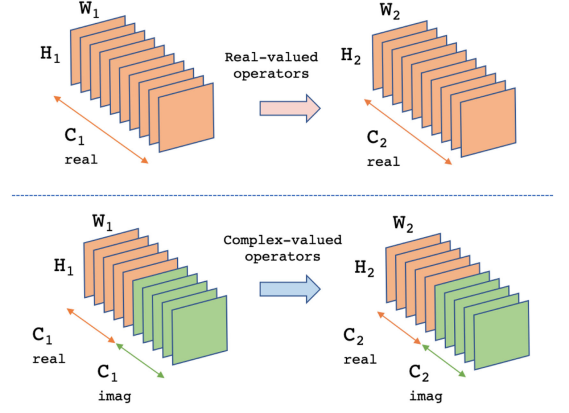


Fig. 4. Real-valued versus Complex-valued operators and feature maps. Orange and green blocks denote real-valued and complex-valued feature maps, respectively. Pink and light blue arrows denote real- and complex-valued operations, respectively. While the volumetric shapes of the real- and complex-valued feature maps look similar, the ways by which they are interpreted and calculated differ in nature.

competitive results on classification and approximation tasks, without using any dropout or max-pooling [39].

**Complex BN.** Batch Normalization of a complex-valued input,  $z$ , is calculated as

$$\text{BN}(z) = \gamma \tilde{z} + \beta, \quad (7)$$

where  $\tilde{z}$  is the normalization of  $z$

$$\tilde{z} = (V)^{-1/2}(z - E[z]). \quad (8)$$

$V$  is the covariance matrix,  $E$  is the mean of the input  $z$ .

### 4.3 Training and Inference in a Complex-Valued Network

The designed network is first trained to learn the weights that best encode the input image. Once trained, it can function as a feature extractor, which infers the feature representation from the normalized iris input image by performing a forward pass. This section highlights how the complex-valued domain shift changes the inference and training process.

**Inference.** The shift to the complex-valued domain requires to store both real-valued and imaginary-valued components in the feature maps.

- A conventional real-valued layer performs a real-valued operation (i.e., real-valued conv, pooling, activation, etc.), to transform an input feature map of size  $W_1 \times H_1 \times C_1$  to an output feature map of size  $W_2 \times H_2 \times C_2$ .
- A complex-valued layer performs a complex-valued operation (i.e., complex-valued conv, pooling, activation, etc. as discussed in Section 4.2) to transform an input feature map of size  $W_1 \times H_1 \times (2C_1)$  to an output feature map of size  $W_2 \times H_2 \times (2C_2)$ . Half of the complex-valued input feature map,  $W_1 \times H_1 \times (C_1)$ , stores the real components and the other half,  $W_1 \times H_1 \times (C_1)$ , stores the imaginary components.

Operations performed on each layer decide the relationship between the input and output feature maps as discussed in Section 4.2. Visualization of the feature map difference is illustrated in Fig. 4.

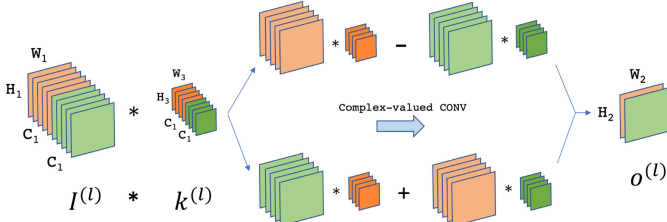


Fig. 5. Complex-valued convolution: the complex-valued input  $I^{(l)}$  is convolved with a complex-valued kernel  $k^{(l)}$  to output a complex-valued output  $o^{(l)}$ . For each convolutional layer, multiple kernels,  $K^{(l)}$ , are employed to generate the final output  $O^{(l)}$ .

**Training.** The common way to train a network, i.e., to minimize the loss function, is by using an iterative steepest gradient descent algorithm (SGD). The SGD algorithm minimizes the loss function by iteratively updating the weights by moving step-wise in the direction of the loss function's steepest descent, which is opposite to its gradient. It iteratively performs: (1) a forward pass using the currently estimated weights to calculate the loss, then subsequently (2) a backward pass to update the weights based on how well the current weights have performed as assessed by the current loss value. The forward pass in the complex-valued domain has been discussed in the previous paragraph. The backward pass is the process to back-propagate the gradients of the loss function,  $C$ , through the network layers and using the chain rule to update the weights.

In the conventional real-valued layers, the gradient is back-propagated from the layer  $l_{i+1}$  to the layer  $l_i$  as

$$\frac{\partial C}{\partial l_i} = \frac{\partial C}{\partial l_{i+1}} \frac{\partial l_{i+1}}{\partial l_i}, \quad (9)$$

where,  $l_i$  is real-valued the feature map input at layer  $i$ th and  $l_{i+1}$  is real-valued the feature map input at layer  $(i+1)$ th.

In the complex-valued layers, the gradient of the loss at the layer  $(i+1)$ th is calculated as

$$\nabla_C(l_{i+1}) = \frac{\partial C}{\partial \Re l_{i+1}} + i \frac{\partial C}{\partial \Im l_{i+1}}, \quad (10)$$

where,  $\Re l_i$  and  $\Im l_i$  are the real and imaginary components of the complex-valued feature map at layer  $i$ . The chain rule is slightly different from the real-valued counterpart as

$$\nabla_C(l_i) = \frac{\partial C}{\partial \Re l_i} + i \frac{\partial C}{\partial \Im l_i} \quad (11)$$

$$= \left( \frac{\partial C}{\partial \Re l_{i+1}} \frac{\partial \Re l_{i+1}}{\partial \Re l_i} + \frac{\partial C}{\partial \Im l_{i+1}} \frac{\partial \Im l_{i+1}}{\partial \Re l_i} \right) \quad (12)$$

$$+ i \left( \frac{\partial C}{\partial \Re l_{i+1}} \frac{\partial \Re l_{i+1}}{\partial \Im l_i} + \frac{\partial C}{\partial \Im l_{i+1}} \frac{\partial \Im l_{i+1}}{\partial \Im l_i} \right) \\ = \frac{\partial C}{\partial \Re l_{i+1}} \left( \frac{\partial \Re l_{i+1}}{\partial \Re l_i} + i \frac{\partial \Re l_{i+1}}{\partial \Im l_i} \right) \quad (13)$$

$$+ \frac{\partial C}{\partial \Im l_{i+1}} \left( \frac{\partial \Im l_{i+1}}{\partial \Re l_i} + i \frac{\partial \Im l_{i+1}}{\partial \Im l_i} \right) \\ = \Re(\nabla_C(l_{i+1})) \left( \frac{\partial \Re l_{i+1}}{\partial \Re l_i} + i \frac{\partial \Re l_{i+1}}{\partial \Im l_i} \right) \\ + \Im(\nabla_C(l_{i+1})) \left( \frac{\partial \Im l_{i+1}}{\partial \Re l_i} + i \frac{\partial \Im l_{i+1}}{\partial \Im l_i} \right). \quad (14)$$

This chain rule allows the gradient of the loss to be back-propagated from the last layer through each layer in the network. The weights are updated accordingly.

**Loss Function.** There are two types of loss functions to train a network: classification losses and pairwise losses. The majority of previous iris recognition approaches based on deep networks use classification losses trained over a set of known iris identities and then utilize the intermediate bottleneck layer as a representation scheme that extends beyond the set of identities used in training. The downsides of this approach are its indirectness and inefficiency [42] while relying on the assumption of the generalizability of the bottleneck representation to new irises. Compared to classification losses, pairwise losses *directly reflect what we want to achieve, i.e., train the representation to correspond to iris similarity*: images of the same iris have small distances and images of different irises have larger distances. We leverage the recent success of a pairwise loss function called Extended Triplet loss as investigated in [55] to train the proposed network. Compared to [55] and [42], there is a fundamental change in optimizing the loss function due to shifting the weights from the real domain to the complex domain.

To form a triplet, we need an anchor image, a positive image and a negative image. The positive image belongs to the same class with the anchor while the negative image belongs to a different class. Denoting the output vector of the network as  $f$  (in our setting this would be the last convolutional layer), we can represent the output features for a particular triplet  $i$  as  $f_i^a, f_i^p, f_i^n$ , denoting the output features for the anchor, positive and negative images, respectively. The goal of a triplet loss function is to make the distance between  $f_i^a$  and  $f_i^n$  (i.e., images from different classes) larger than the distance between  $f_i^a$  and  $f_i^p$  (i.e., images from the same class) by a minimum margin  $\alpha$ . The extended triple loss function is defined as

$$ETL = \frac{1}{N} \sum_{i=1}^N [D(f_i^a, f_i^p) - D(f_i^a, f_i^n) + \alpha], \quad (15)$$

where,  $D$  is the distance between two vectors. In the iris recognition case,  $D$  between two feature maps,  $f^1$  and  $f^2$ , has been designed to deal with eye rotations and segmentation masks as

$$D(f^1, f^2) = \min_{-B \leq b \leq B} \{FD(f_b^1, f^2)\}, \quad (16)$$

where,  $b$  denotes that the feature map has been shifted horizontally by  $b$  pixels;  $B$  is the maximum of the shift allowed, i.e., 4 pixels in this work; and  $FD$  is the Fractional Distance which takes the masks into consideration

$$FD(f^1, f^2) = \frac{1}{|M|} \sum_{(x,y) \in M} (f_{x,y}^1 - f_{x,y}^2)^2. \quad (17)$$

## 5 EXPERIMENTAL RESULTS

We conducted our experiments on three public datasets:

- ND-CrossSensor-Iris-2013 dataset<sup>1</sup>: is the largest publicly available iris dataset in the literature in terms of the number of images [34]. It contains 116,564 iris images captured by the LG2200 iris camera from

1. <https://sites.google.com/a/nd.edu/public-cvrl/data-sets>



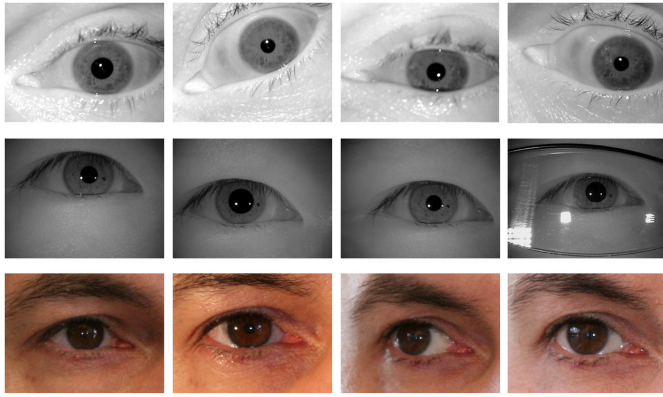


Fig. 6. Sample images from the ND-CrossSensor-2013 (first row), CASIA-Iris-Thousand (second row) and UBIRIS.v2 datasets (last row).

676 subjects. This is a super set of the ND-IRIS-0405 dataset with double the number of images (116,564 versus 64,980) and the number of subjects (676 versus 356). The number of images for each subject in the test set is between 11 and 312. After removing some falsely segmented samples, the test set contains 57,253 images in total, with 3,838,981 genuine pairs and 3,270,285,300 impostor pairs.

- CASIA-Iris-Thousand dataset<sup>2</sup>: contains 20,000 iris images from 1,000 subjects, which were collected using the IKEMB-100 camera from IrisKing [33]. The test set, after removing some falsely segmented samples, contains 81,799 genuine pairs and 76,530,208 impostor pairs.
- UBIRIS.v2 iris dataset<sup>3</sup>: contains 11,102 iris images from 261 subjects with 10 images each subject. The images were captured under unconstrained conditions (at-a-distance, on-the-move and in the visible wavelength), with corresponding more realistic noise factors [36]. The test set, after removing some falsely segmented samples, contains 9,629 genuine pairs and 4,317,548 impostor pairs.

Experiments with these three datasets allow us to validate the performance of the ComplexIrisNet in diverse iris recognition scenarios, ranging from close-up distances in ND-CrossSensor-Iris-2013 and CASIA-Iris-Thousand to long distances in UBIRIS.v2; and from conventional near-infrared imaging in ND-CrossSensor-Iris-2013 and CASIA-Iris-Thousand to visible spectrum imaging in UBIRIS.v2. Sample images and statistics of the three datasets are depicted in Fig. 6 and summarized in Table 2.

## 5.1 Performance Metrics

To report the performance of iris recognition approaches, we rely on Decision Error Trade-off (DET) curves and False Rejection Rates (FRRs). The DET curve is a graphical plot that illustrates the diagnostic ability of a classifier by calculating False Rejection Rate (FRR) against False Acceptance Rate (FAR) at various threshold settings. In this work, FRRs at FAR = 0.1% are reported due to its popular adoption in

the field. Equal Error Rate (EER) is the operating point where FAR is equal to FRR.

Due to the differences in the number of images, the ND-CrossSensor-Iris-2013 dataset is chosen for our intra-dataset experiments and the other two (CASIA and UBIRIS) are used for cross-dataset experiments.

*Intra-Dataset Performance.* Similar to [55], we train and test on disjoint identities. We use 80% of the left eye images for training, 20% of the left eye images for validation and all right eye images for testing. This guarantees that none of the classes in the test subset is in the train subset to avoid experimental bias. The training subset is used to train the ComplexIrisNet in an end-to-end manner to find the best weights. The validation subset is used to tune hyperparameters for the ablation study. The testing subset is used to report the intra-dataset performance.

*Cross-Dataset Performance.* The proposed network is further investigated for generalization capability through training in one dataset and transferring the learned model to others. The pre-trained ComplexIrisNet, which has been pre-trained on the ND-CrossSensor-Iris-2013 dataset in the intra-dataset experiments, is subsequently cross-validated on two other datasets, CASIA-Iris-Thousand and UBIRIS.v2. We investigate two cross-dataset configurations: with and without fine-tuning. With fine-tuning, the pre-trained ComplexIrisNet is fine-tuned using all left eye images and is tested using all right eye images on the two datasets. Without fine-tuning, the pre-trained ComplexIrisNet is applied directly to all right eye images in the two datasets.

## 5.2 Experimental Setup

We first pre-process the iris images via segmentation and normalization. The iris image is first segmented using two circles for the inner and outer boundaries of the iris region corresponding to the pupillary and limbus boundaries, respectively. We adopt a deep multi-task learning framework for joint iris segmentation and localization as in [51]. The iris region from the raw Cartesian coordinates is then re-mapped to a dimensionless pseudo-polar coordinate, where the iris region is normalized to a fixed and rectangular size of  $64 \times 256$  pixels [6]. The corresponding noise mask is also normalized to facilitate matching in the later stages.

It has to be stressed here that all algorithms discussed in the experiments were performed on the same test sets. As the preprocessing step may affect the overall performance, we ran the handcrafted, the state of the art and the proposed complex-valued algorithms on the same test sets for a fair comparison.

The ComplexIrisNet is trained using the back-propagation algorithm with Stochastic Gradient Descent and with the Nesterov momentum set at 0.9. The norm of gradients is clipped to 1. The learning rate is initialized at 0.01 for the first 10 epochs, then set to 0.1 from epoch 10 to 100, and then annealed by a factor of 10 at epochs 130 and 160. The training error of the ComplexIrisNet is summarized in Fig. 7.

## 5.3 Ablation Study

### 5.3.1 Impact of Dense Architecture Variants

We first investigate how the architecture in terms of the number of dense blocks affects the performance. We vary

2. <http://biometrics.idealtest.org>

3. <http://iris.di.ubi.pt/ubiris2.html>



TABLE 2  
Statistics of Three Datasets, ND-CrossSensor-2013, CASIA-Iris-Thousand and UBIRIS.v2, Used in This Research

	# Subjects	# Images	Distance	Imager	Image Resolution	Iris Diameter	Wavelength	Subject Cooperation
ND-CrossSensor-2013	676	111,564	Close-up	LG2200	640x480	200	NIR	Highly
CASIA-Iris-Thousand	1,000	20,000	Close-up	IKEMB-100	640x480	180	NIR	Highly
UBIRIS.v2	261	11,102	4-8 meters	CanonEOS 5D	800x600	180-80	Visible	Less

this number from 1 to 3, since in the natural image classification setting, 3 dense blocks have shown to achieve state-of-the-art performance [21]. Each dense block is followed by a transition block as explained in Section 4.1. Considering that the iris images are not as diverse as the natural images, a smaller-sized network may be able to achieve good results and avoid overfitting. Our experimental results show that a network with two dense blocks achieves the best performance during validation with a 1.31% FRR (False Reject Rate) at a FAR (False Accept Rate) of 0.1% on the ND-CrossSensor-Iris-2013 validation subset as shown in Fig. 8. Even a network with one dense block exhibits very good performance - 2.84% FRR at FAR = 0.1% - proving the high modeling capacity of the proposed complex-valued architecture in the iris recognition setting. From this point on, two dense blocks were used in our ComplexIrisNet unless otherwise indicated.

### 5.3.2 Impact of Modeling Complex-Valued Feature Maps

We further investigate the benefits of performing complex-valued operations within a deep network by comparing with variants of real-valued counterparts. We consider four real-valued representations. As shown in Equation (3),  $z = x + iy = re^{j\theta} \in \mathbb{C}$ , where  $x, y, r, \theta \in \mathbb{R}$  are the real part, the imaginary part, the magnitude and the phase of  $z$ , respectively. Four real-valued baselines are:

- 1.  $r$  - Pure real network: Treat the input image purely with a real-valued network, there is no Gabor Block to generate the complex-valued response. This is similar to all other deep iris recognition networks in the literature.
- 2.  $(x, y)$  - Double-channel real network: Treat the complex-valued Gabor response as one real-valued network.
- 3.  $(r), (\theta)$  - 2 separate networks (Real and Imaginary): Treat the complex-valued Gabor response as

two real-valued networks: one processes the real response  $x$ , and the other processes the imaginary response  $y$ .

- 4.  $(r), (\theta)$  - 2 separate networks (Magnitude and Phase): Treat the complex-valued Gabor response as two real-valued networks: one processes the magnitude response  $r$ , and the other processes the phase response  $\theta$ .

Fig. 9 explains the architectural difference between the four real-valued baselines and the proposed complex-valued iris network. It is clear that the real-valued networks, even with the same configurations (the number of layers, the number of filters per layer and the size of feature maps), would struggle to capture the phase and the intrinsic geometry of complex-valued data. This has been reinforced in terms of performance in Fig. 10.

The comparison illustrates the benefit of shifting the network operations to the complex domain. We hypothesize that the standard real-valued networks fail to fully capture the intrinsic geometric property of complex-valued data. Real-valued operations discard some of the geometric details and do not necessarily maintain the phase and spectral information of the data as it moves through the network. The complex-valued operations in our complex-valued network better capture the geometry and explicitly retain the phase information, which justifies its superior accuracy compared to the real-valued counterparts.

*Visualization.* We visualize and analyze the network's predictions to understand the learning process underlying the complex network. GradCAM [43] and CNN Fixation [28] are two popular methods to visualize CNNs; however, they are not designed for such pairwise networks as our ComplexIrisNet. The work in [45], [57] have proposed to decompose the last convolution activation to highlight the image regions that contribute the most to the overall matching score. We employ these frameworks to visualize and

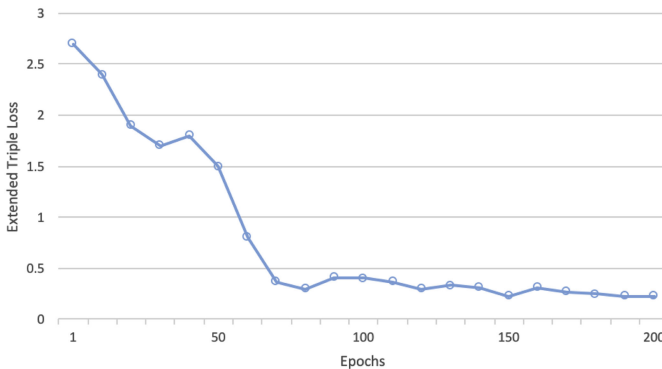


Fig. 7. Training losses of the ComplexIrisNet on the ND-CrossSensor-Iris-2013 training subset.

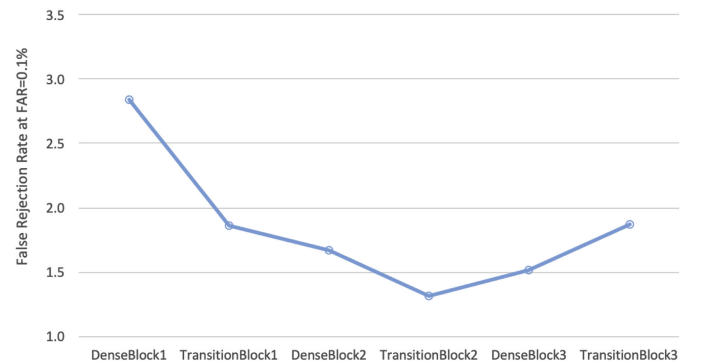


Fig. 8. Performance comparison of different layers of the ComplexIrisNet in terms of False Rejection Rate at 0.1% False Acceptance Rate on the ND-CrossSensor-Iris-2013 validation subset.

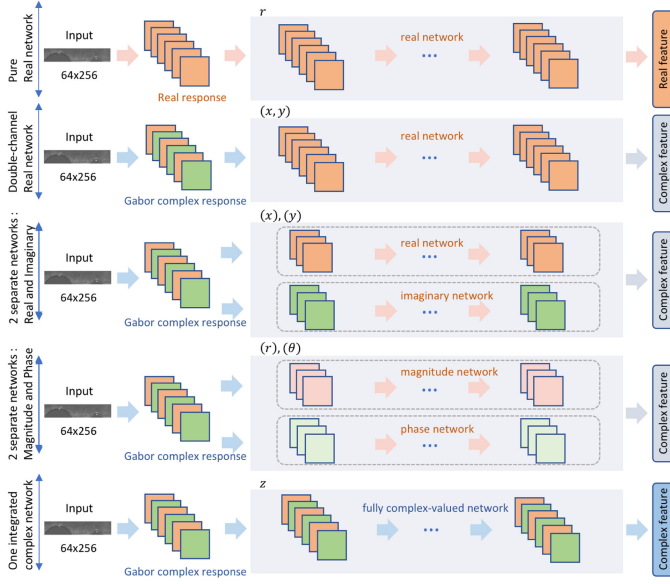


Fig. 9. Architecture comparison of four real-valued baselines and the complex-valued iris network. The complex-valued iris network enables complex-valued feature maps and complex-valued operations through the whole network.

compare the last convolution activation of both complex-valued and real-valued iris networks.

Fig. 11 visualizes the activation decomposition of three genuine pairs and one impostor pair for the subject ID 02463 from the ND-CrossSensor-2013 dataset. The first two columns, respectively, depict the gallery and probe image pairs. The third and fourth columns, respectively, illustrate the heatmap overlays from the complex iris network and the real iris network. The heatmap overlays show the relative spatial contribution of each image region to the overall matching score. Hot colors denote high similarity while cold colors denote lower similarity. There are two observations from this figure to reflect on:

- For real-valued iris network: The real networks tend to focus on the regions where the texture shows significant intensity change, which is demonstrated when many hot color regions in the real heatmap overlays

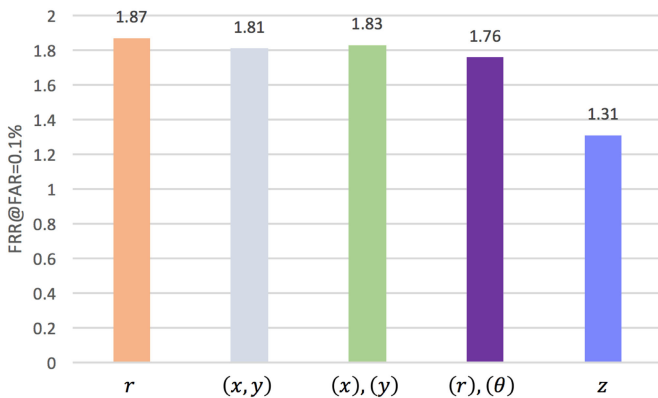


Fig. 10. Performance comparison of four real-valued network baselines and the proposed ComplexIrisNet on the ND-CrossSensor-Iris-2013 validation subset. This illustrates the benefits of performing complex-valued operations, which better capture the geometry and explicitly retain the phase information, which justifies its superior accuracy compared to the real-valued counterparts.

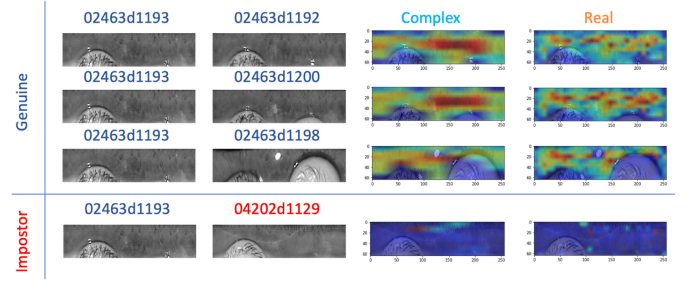


Fig. 11. Visualization of the activation decomposition of three genuine iris image pairs and one impostor iris image pair from the ND-CrossSensor-2013 dataset. The real iris networks tend to focus more on highly textured regions with significant spatial change, such as crypts. In contrast, the complex iris networks have a more diverse coverage, not just the regions with significant intensity change but also plainer texture regions, which exhibit less spatial change. The observation of this property in the network's activations demonstrates the richer capacity of the complex network to capture both spatial and phase details compared to the real network counterparts.

are located in the crypts of the iris texture. In object detection and classification, real-valued deep networks are known for their capability to learn object structure via a hierarchical approach where earlier layers detect the basic edges and corners while later layers progressively detect more complex shapes [1]. Unfortunately, this capability is not significant in the iris recognition setting due to the stochastic nature of the iris texture [10]. The fact that the iris texture in iris recognition has no consistent shapes, edges, or structure unlike in classical object detection and classification, causes real-valued networks to struggle to learn any meaningful semantics from the iris texture and is unable to realize the full potential of automatic feature learning in the iris recognition setting.

- For the complex iris network: The complex networks tend to have a more diverse coverage across the entire iris image, not just the regions with significant intensity change. They also focus on plainer texture regions, which visually seem to exhibit less spatial change. The observation of this property in the network's activations demonstrates the richer capacity of the complex network to capture both spatial and phase details compared to its real network counterpart. In contrast to the spatial focus of the real iris networks, the complex iris networks also focus on phase information embedded within the image, making them more suitable for the iris recognition task. This highlights the fundamental value of the proposed complex-valued iris networks compared to the real-valued iris networks which is their ability to capture richer information (spatial and phase versus spatial only). This ability allows complex iris networks to: (1) perform better genuine pair matching; better focus on the texture regions with less intensity value changes to recognize whether they are identical versus real networks which only focus on significant intensity change regions; and (2) perform better impostor pair rejection: through an increased focus on both phase and spatial information to provide more clues to differentiate regions as shown in Fig. 11.

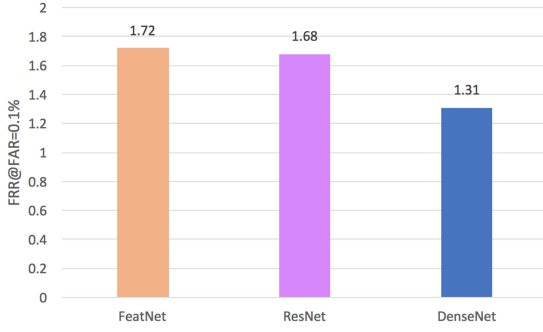


Fig. 12. Performance comparison of different backbone architecture choices for the complex-valued architecture on the ND-CrossSensor-Iris-2013 validation subset.

### 5.3.3 Impact of Backbone Architecture Choices

We now study how the choice of this backbone architecture compared to other architectures in the complex-valued domain. We retain the Gabor Block and swap the following dense blocks with one landmark architecture in general deep learning, *viz.*, ResNet architecture [16], and one landmark architecture in iris deep learning, *viz.*, FeatNet architecture [55]. The FeatNet architecture is the state-of-the-art architecture in applying deep learning to iris recognition. FeatNet has three blocks of *CONV – TANH – POOL*. The feature maps from the *TANH* activation layers are upsampled and stacked together before a fourth *CONV* is applied. The ResNet architecture is the state-of-the-art (and among the most popular) architecture in general deep learning. Compared to architectures such as Inception [47], MobileNet [18] and SENet [19], ResNet is chosen due to its notable generalization capacity together with the simplicity and consistency of its architecture. Details of FeatNet and ResNet architectures can be found in [16], [55]. We apply these two backbone architectures (FeatNet and ResNet) with the complex-valued operations as shown in Section 4.2.

The effectiveness of the proposed architecture based on the dense blocks is validated by the performance on the validation subset as shown in Fig. 12, achieving a FRR = 1.31% compared to 1.68% and 1.72% at FAR = 0.1% of ResNet and FeatNet, respectively.

### 5.4 Comparison With Handcrafted Features

We next compare the proposed method with handcrafted approaches. The major advantage of handcrafted features

compared to deep learning features is fast computation and no requirement for training, which is beneficial for mobile and embedded applications. However, for applications such as large-scale identification systems running on high-end computing infrastructure, the recognition accuracy may be of more interest than the computation cost. The experiments in this section are designed to investigate how much performance can be leveraged by the proposed complex-valued data-driven features.

We compare with the classic IrisCode, its phase-based derivatives (DCT-based [27] and DFT-based [26]), and a state-of-the-art non-phase-based Ordinal features [46]. As reviewed in Section 2, the IrisCode algorithm relied on non-linear encoding of the phase information extracted from the Gabor complex domain using a binarization scheme [6], [11]. We implemented the algorithm based on Daugman's original papers [6], [11], and used recent papers [55], [56] to optimize the parameters for the highest potential performance and comparable experimental settings. We employ a bank of  $M = 40$  Gabor filters with 5 wavelengths and 8 orientations. For a fair comparison, similar to [46], [55], we employ 72 ordinal filters. Other parameters are fine-tuned by a grid search for the best performance on the experimental datasets.

For intra-dataset comparison, experiments are performed on the ND-CrossSensor-2013 dataset. The proposed complex iris networks with different backbones are first trained to learn a feature representation using the train set. Once the deep networks have been trained, they are employed to perform recognition on the test set. Since the handcrafted approaches do not require training, they are directly employed to perform recognition on the test set. The performance of the handcrafted features and complex iris networks with various backbones on the test set is illustrated in Fig. 13a and the intra-dataset column of Table 3. Compared to an EER of 1.67% for the classic IrisCode, the best complex iris network achieves an EER of 0.66%, which translates to a reduction in EER of 60%.

For cross-dataset comparison, the complex iris networks, which have been trained in the intra-dataset experiments, are employed to perform recognition, with and without fine-tuning on the train sets of the CASIA-Iris-Thousand and UBIRIS.v2 datasets. Since the handcrafted approaches do not require training, these are again directly employed to perform recognition on the test sets. On the CASIA-Iris-

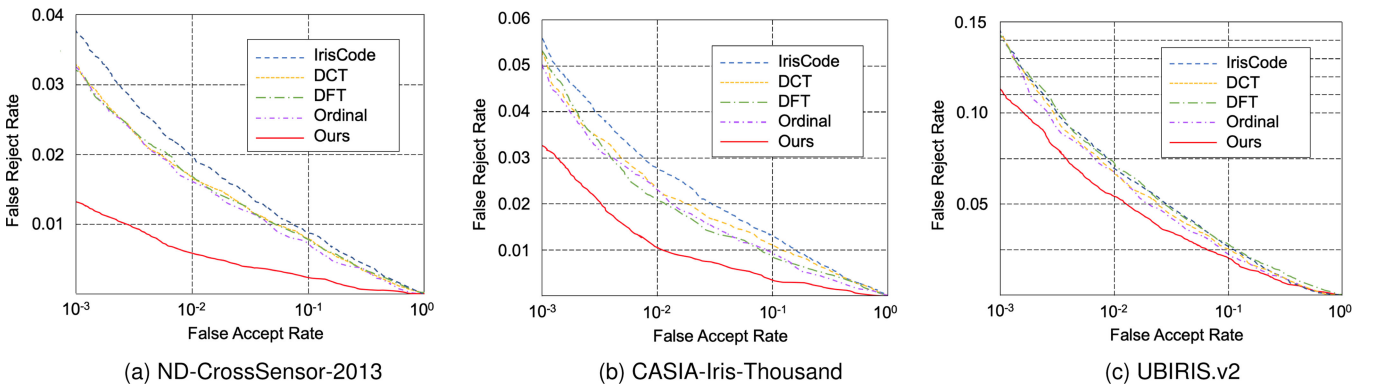


Fig. 13. DET curves for comparison of the proposed complex-valued data-driven features with other classic handcrafted feature representations on the test sets of three datasets. *Best viewed in color.*

TABLE 3

Performance Comparison With State-of-the-Art Handcrafted and Deep Learning Iris Recognition Approaches in Terms of False Rejection Rate at 0.1% False Acceptance Rate and Equal Error Rate on the Test Sets of Three Datasets

	Intra-dataset		Cross-dataset							
	ND-CrossSensor-2013		CASIA-Iris-Thousand				UBIRIS.v2			
			Fine-tuning		No fine-tuning		Fine-tuning		No fine-tuning	
			FRR	EER	FRR	EER	FRR	EER	FRR	EER
IrisCode	3.43%	1.67%	-	-	5.33%	3.37%	-	-	14.26%	8.30%
DCT [27]	3.21%	1.56%	-	-	5.18%	3.29%	-	-	14.25%	8.26%
DFT [26]	3.18%	1.53%	-	-	5.19%	3.30%	-	-	14.22%	8.23%
Ordinal [46]	3.17%	1.52%	-	-	5.01%	3.21%	-	-	14.17%	8.07%
DeepIris [24]	2.60%	1.29%	4.22%	2.15%	4.39%	2.25%	13.21%	7.13%	13.67%	7.56%
FeatNet [55], [56]	1.79%	0.99%	3.96%	1.92%	4.01%	2.03%	12.92%	6.68%	13.24%	7.03%
DRFNet [52]	1.77%	0.92%	3.91%	1.89%	4.05%	2.00%	12.62%	6.74%	13.20%	6.88%
Complex (FeatNet backbone)	1.45%	0.76%	3.66%	1.75%	3.72%	1.81%	12.11%	6.22%	12.32%	6.14%
Complex (ResNet backbone)	1.41%	0.75%	3.67%	1.76%	3.70%	1.79%	12.03%	6.17%	12.33%	6.18%
<b>ComplexIrisNet</b>	<b>1.31%</b>	<b>0.66%</b>	<b>3.25%</b>	<b>1.62%</b>	<b>3.31%</b>	<b>1.64%</b>	<b>11.56%</b>	<b>6.01%</b>	<b>11.98%</b>	<b>6.15%</b>

Thousand dataset, the best complex iris network achieves EERs of 1.62% and 1.64% with and without fine-tuning, respectively. Compared to an EER of 3.37% for the IrisCode, the complex iris network reduces EERs by 52% and 50% with and without fine-tuning, respectively. On the UBIRIS dataset, the best complex iris network achieves EERs of 6.01% and 6.15% with and without fine-tuning, respectively. Compared to an EER of 8.30% for the IrisCode, the complex iris network reduces EERs by 26% and 25% with and without fine-tuning, respectively. Compared to the cross-dataset performance on the CASIA dataset, the performance gained from no-finetuning to finetuning on the UBIRIS dataset is less due to the domain shift from NIR to Visible in the latter.

The data shows that the features learned from the proposed ComplexIrisNets achieve consistently higher performance on all three datasets, reducing EERs by 25% to 60% in comparison with the traditional IrisCode. Similar performance gains over the other three handcrafted approaches are also observed. This not only shows that the feature representation automatically discovered by the ComplexIrisNet is highly discriminative achieving higher competitive performance, but it also illustrates the generalization capability of the discovered feature representation when tested on other datasets, both NIR (CASIA-Iris-Thousand) and Visible (UBIRIS.v2).

We note that the choice of a FAR at 0.1% is to be consistent with the literature. This value has been widely used in deep learning papers for iris recognition such as [24], [55], [56]. Using this value enables us to directly compare our performance with theirs. We acknowledge that, as discussed in [9], iris recognition has a large entropy and the IrisCode can generate very flat DET curves, i.e., the FAR can be reduced by many orders of magnitude, even by factors of 10,000 to 100,000, while only paying a price of roughly doubling the FRR. This point has been confirmed in National Institute of Standards and Technology (NIST) reports [32], [38]. This is one of the advantages of iris recognition that makes IrisCode very successful.

*VeriEye*. We also compare our method with the commercial product VeriEye SDK<sup>4</sup> from NeuroTechnology. In 2020,

Neurotechnology's iris recognition algorithm was judged by NIST as the second most accurate among the IREX 10 participants. We obtained a copy of the VeriEye SDK and experimented on the ND dataset. We randomly selected different 100 irises, each having 10 images. There are a number of irises with less than 10 images in the dataset, for which we simply used all available images. In total, there are 973 images, generating 4,293 genuine image pairs and 468,585 impostor image pairs. We ran the VeriEye SDK, our home-brewed IrisCode and the ComplexIrisNet to compute matching scores. The results are presented in Table 4. The VeriEye SDK achieved a better EER compared to the home-brewed implementation of IrisCode. However, the ComplexIrisNet performs much better than VeriEye, thereby conveying the efficacy of the proposed method.

## 5.5 Comparison With Deep Learning Features

We next compare with other deep-learning-based approaches in the literature. The state of the art of deep iris networks is based on pairwise losses with three state of the art approaches - DeepIris [24], FeatNet [55] and DRFNet [52]. It is noteworthy that [56] uses the same FeatNet as the backbone network to extract its iris feature representation. Since the source codes of these two approaches are not available, we re-implemented and optimized them to a comparable level of performance with those reported in the original papers.

We implemented DeepIris with 9 layers including one pairwise filter layer, one convolutional layer, two pooling layers, two normalization layers, two local layers and one fully connected layer [24]. This network uses the normalized iris images with a size of  $100 \times 100$  pixels; hence, in this experiment, we chose this output size for the segmentation process.

We implemented the FeatNet with 4 convolutional layers, each followed by an activation layer (*Tanh*) and an average

TABLE 4  
Comparison With the Commercial VeriEye SDK

	EER
Our IrisCode	1.66%
VeriEye	0.99%
ComplexIrisNet	0.66%

4. <https://www.neurotechnology.com/verieye.html>



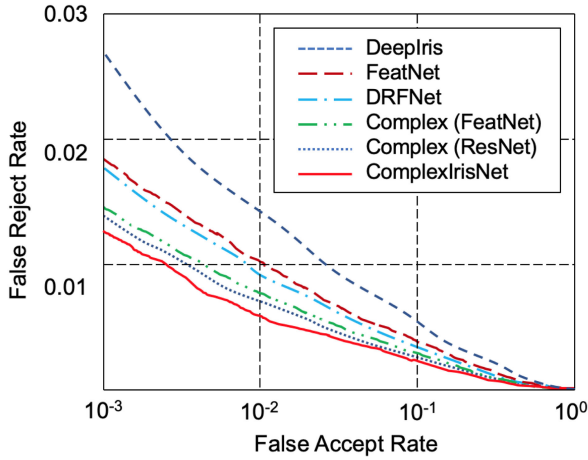


Fig. 14. DET curves for comparison with other deep learning feature representations on the test set of the ND-CrossSensor-2013 dataset. *Best viewed in color.*

pooling layer as detailed in [55]. We implemented the Extended Triple Loss incorporating bit-shifting and masking [55]. This network is trained using the same pairwise loss scheme as the proposed ComplexIrisNet. One major difference is the architecture and the capacity to deal with complex-valued weights and feature maps. To purely compare the feature representation power, we used our own segmentation approach rather than their segmentation approach (MaskNet). Our implementation achieved comparable results to those reported in the paper (FRRs of 1.79% versus 1.78% and same EER of 0.99%); the slight difference in performance may be due to the differences in the segmentation approach and dataset splitting scheme. We also implemented the improved version of FeatNet called DRFNet with dilated convolution and residual connections [52]. DRFNet achieves slightly better performance in comparison with its plain version in FeatNet with FRR of 1.77% and EER of 0.92%.

It is noteworthy that the other two deep iris networks in the literature [13], [30] use cross-entropy losses, which limits the applicability of the iris networks because the joint identities between the training and testing are required. The output of the intermediate layers from these networks can be extracted for use as feature vectors and some classifiers such as the euclidean distance can be used to measure the (dis)similarity. However, their performance is inferior to the pairwise networks and are not included here.

We also include the complex-valued networks with two different backbone architectures (FeatNet and ResNet) for comparison.

For the intra-dataset scenario, experiments are performed on the ND-CrossSensor-2013 dataset. The proposed complex iris networks with various backbones and state of the art deep iris networks (DeepIris, FeatNet, DRFNet) are first trained on the same train set of the ND dataset. Once the networks have been trained, they are employed to perform recognition on the same test set of the ND dataset. Their performance is illustrated in Fig. 14 and the intra-dataset column of Table 3. The data shows that complex iris networks outperform all state of the art deep (real-valued) iris networks by large margins. The best complex iris network reduces the EER by 28% - from 0.92% for the best state of the art DRFNet [52] to 0.66% for the ComplexIrisNet.

TABLE 5  
Complexity Analysis

Approach	#Parameters	Model Size (Byte)	Feature Extraction
DeepIris [24]	55,420.0 K	289.0 M	12.70 ms
FeatNet [55]	129.9 K	1.5 M	8.93 ms
DRFNet [52]	125.3 K	1.3 M	8.12 ms
ComplexIrisNet	421.2 K	20.5 M	21.08 ms

For the cross-dataset scenario, the proposed complex iris networks and all state of the art deep (real-valued) iris networks, which have been pre-trained on the ND train set in the intra-dataset experiments, are employed to perform recognition, with and without fine-tuning, on the CASIA and UBIRIS datasets. The experimental results are shown in Fig. 14 and the bottom half of Table 3. The proposed ComplexIrisNet significantly reduces the EER by 18% from 2.00% to 1.64% on the CASIA no fine-tuning experiments. Similar improvements have been shown in the UBIRIS dataset where the EER reduces from 6.88% to 6.15%. This demonstrates the effectiveness of processing the weights and feature maps in the complex domain.

We also provide a complexity analysis and comparison as shown in Table 5. Three representative methods - DeepIris [24], FeaNet [55] and DRFNet [52] - are compared to our method on the same machine. Our machine configuration is Intel(R) Core(TM) i7-8700K CPU @ 3.70GHz with 32GB DDR4 RAM, and NVIDIA GeForce RTX 2080 card with 8GB memory. It can be seen that the computational time of ComplexIrisNet is longer but acceptable (approximately 50 fps).

## 6 DISCUSSION AND CONCLUSIONS

It is clear the iris stromal texture in iris recognition has no consistent shapes, edges, or structure unlike in classical object detection and classification task. This would make standard real-valued networks struggle to learn any meaningful semantic shapes from the iris texture during automatic feature learning and thereby not realizing the full potential of automatic feature learning in the iris recognition setting.

Complex-valued networks have a solid mathematical and theoretical foundation that make them more suitable for iris recognition when compared to real-valued networks.

- Complex-valued networks explicitly model and retain phase information through the whole network. Phase is very important in iris encoding as it has been shown in the success of many handcrafted features.
- Complex-valued networks are better at capturing features at multiple scales, multiple frequencies and multiple orientations [49], which have the capacity to approximate such non-linear multiwavelet packets as Gabor wavelets in the classic IrisCode.
- Complex-valued networks allow automatic complex-valued feature learning, which combines the strength of modern deep learning (automatic feature learning) and domain specific knowledge of the iris recognition field (complex filters). Compared to automatic real-valued feature learning in real-valued

networks, automatic complex feature learning better suits the stochastic nature of the iris texture, leading to better feature learning.

- Complex-valued networks can be considered as a generalization of the classic IrisCode due to their correspondence with Gabor wavelets. This allows us to understand the optimality of the classic IrisCode, since if it is optimal, the complex-valued networks will converge to its architecture and performance.

Despite these benefits, complex-valued networks have not been explored in the iris recognition literature. The key reason is due to the fact that the iris images themselves are not complex-valued, and so all existing deep iris networks operate on the real-valued intensity of iris images.

We re-purpose the complex-valued response of Gabor filters to enable the development of new complex-valued networks to directly operate on iris images to translate all the above benefits to iris recognition. We also upgrade the dense connection and fully convolutional architecture to the complex-valued domain, which enables the proposed fully complex-valued network architecture to effectively and automatically learn a complex-valued feature representation for iris recognition. The proposed complex-valued iris network interestingly is a generalization of both the classic IrisCode and many state of the art deep iris networks. We show in our ablation study that complex-valued iris networks actually outperform their real-valued counterparts. More importantly, through visualizing the activation maps at the last convolutional layer, we show that while real-valued networks focus on capturing spatial detail, complex-valued networks are more sensitive to phase and are much better at capturing both spatial and phase information. The fundamental value of complex-valued networks compared to real-valued networks for iris recognition is the ability to capture richer detail (phase and spatial versus spatial only) for better genuine pair matching and better impostor pair rejection.

Through comprehensive comparison with handcrafted features and modern deep iris networks, we have shown that the feature representation discovered through the automatic complex feature learning process of the proposed method is more discriminative and informative than the feature representation discovered by the real-valued networks. The accuracy of the proposed complex-valued iris network surpasses the accuracy of the classic IrisCode (as implemented by us, judiciously guessing its parameter values), its phase-based derivatives and other state-of-the-art deep learning approaches by significant margins. The effectiveness and generalization capacity of the discovered feature representation has been demonstrated across three iris datasets under a variety of settings: near-infrared and visible spectral bands, near and far stand-off distances, and in strict and less cooperative scenarios.

Interestingly, in one of his original papers [4], Daugman actually showed that IrisCode could be expressed as a two-layer CNN. The introduction of more layers and complex-valued operations, in theory, will lead to models with increased modeling capacity than classical approaches. The experiments conducted across multiple datasets have also, empirically, confirmed this observation.

This paper has presented another major step in the representation evolution in iris recognition, from classical handcrafted features to current deep learning with standard

real-valued networks to tailored deep learning with complex-valued networks. With many benefits highly suitable for iris recognition, complex-valued networks deserve more attention and investigation from the iris recognition community to fully take advantage of automatic complex-valued feature learning. This work has taken the first step in this direction, and opens other opportunities for explicitly processing phase information in other domains where complex filters are routinely used for texture modeling.

## REFERENCES

- [1] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, Aug. 2013.
- [2] R. Chakraborty, J. Wang, and S. X. Yu, "SurReal: Frechet mean and distance transform for complex-valued deep learning," in *Proc. IEEE Comput. Vis. Pattern Recognit. Workshop*, 2019, pp. 889–897.
- [3] A. Chowdhury, S. Kirchgasser, A. Uhl, and A. Ross, "Can a CNN automatically learn the significance of minutiae points for fingerprint matching?," in *Proc. IEEE Winter Conf. Appl. Comput. Vis.*, 2020, pp. 340–348.
- [4] J. G. Daugman, "Complete discrete 2-D Gabor transforms by neural networks for image analysis and compression," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, no. 7, pp. 1169–1179, Jul. 1988.
- [5] J. Daugman, "How iris recognition works?," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 1, pp. 21–30, Jan. 2004.
- [6] J. Daugman, "New methods in iris recognition," *IEEE Trans. Syst., Man Cybern. B, Cybern.*, vol. 37, no. 5, pp. 1167–1175, Oct. 2007.
- [7] J. Daugman, "Iris encoding and recognition using Gabor wavelets," in *Encyclopedia of Biometrics*, S. Z. Li and A. Jain, Eds., Boston, MA, USA: Springer, 2009, pp. 787–797.
- [8] J. Daugman, "Information theory and the IrisCode," *IEEE Trans. Inf. Forensics Secur.*, vol. 11, no. 2, pp. 400–409, Feb. 2016.
- [9] J. Daugman, "Collision avoidance on national and global scales: Understanding and using big biometric entropy," early access, Feb. 2021, doi: [10.36227/techrxiv.14061671.v1](https://doi.org/10.36227/techrxiv.14061671.v1).
- [10] J. Daugman and C. Downing, "Epigenetic randomness, complexity and singularity of human iris patterns," *Proc. Biol. Sci.*, vol. 268, no. 1477, pp. 1737–1740, 2001.
- [11] J. G. Daugman, "High confidence visual recognition of persons by a test of statistical independence," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 11, pp. 1148–1161, Nov. 1993.
- [12] I. Fogel and D. Sagi, "Gabor filters as texture discriminator," *Biol. Cybern.*, vol. 61, pp. 103–113, 1989.
- [13] A. Gangwar and A. Joshi, "DeepIrisNet: Deep iris representation with applications in iris recognition and cross-sensor iris recognition," in *Proc. IEEE Int. Conf. Image Process.*, 2016, pp. 2301–2305.
- [14] W. Geisler, "Visual perception and the statistical properties of natural scenes," *Annu. Rev. Psychol.*, vol. 59, no. 1, pp. 167–192, 2008.
- [15] N. Guberman, "On complex valued convolutional neural networks," 2016, arXiv:1602.09046.
- [16] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
- [17] A. Hirose Ed., *Complex-Valued Neural Networks: Advances and Applications*. Hoboken, NJ, USA: Wiley, 2013.
- [18] A. Howard et al., "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, arXiv:1704.04861.
- [19] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7132–7141.
- [20] Y. Hu, K. Sirlantzis, and G. Howells, "Optimal generation of iris codes for iris recognition," *IEEE Trans. Inf. Forensics Secur.*, vol. 12, no. 1, pp. 157–171, Jan. 2017.
- [21] G. Huang, Z. Liu, K. Q. Weinberger, and L. van der Maaten, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2261–2269.
- [22] A. W. K. Kong, D. Zhang, and M. S. Kamel, "An analysis of IrisCode," *IEEE Trans. Image Process.*, vol. 19, no. 2, pp. 522–532, Feb. 2010.
- [23] B. Kwolek, "Face detection using convolutional neural networks and Gabor filters," in *Proc. Int. Conf. Artif. Neural Netw.: Biol. Inspirations*, 2005, pp. 551–557.
- [24] N. Liu, M. Zhang, H. Li, Z. Sun, and T. Tan, "DeepIris: Learning pairwise filter bank for heterogeneous iris verification," *Pattern Recognit. Lett.*, vol. 82, pp. 154–161, 2016.

- [25] S. Luan, C. Chen, B. Zhang, J. Han, and J. Liu, "Gabor convolutional networks," *IEEE Trans. Image Process.*, vol. 27, no. 9, pp. 4357–4366, Sep. 2018.
- [26] K. Miyazawa, K. Ito, T. Aoki, K. Kobayashi, and H. Nakajima, "An effective approach for iris recognition using phase-based image matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 10, pp. 1741–1756, Oct. 2008.
- [27] D. M. Monro, S. Rakshit, and D. Zhang, "DCT-Based iris recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 4, pp. 586–595, Apr. 2007.
- [28] K. R. Mopuri, U. Garg, and R. Venkatesh Babu, "CNN fixations: An unraveling approach to visualize the discriminative image regions," *IEEE Trans. Image Process.*, vol. 28, no. 5, pp. 2116–2125, May 2019.
- [29] A. Muron and J. Pospisil, "The human iris structure and its usages," *Acta Univ. Plalcki Phys.*, vol. 39, pp. 87–95, 2000.
- [30] K. Nguyen, C. Fookes, A. Ross, and S. Sridharan, "Iris recognition with Off-the-Shelf CNN Features: A deep learning perspective," *IEEE Access*, vol. 6, pp. 18848–18855, 2018.
- [31] K. Nguyen, C. Fookes, and S. Sridharan, "Constrained design of deep iris networks," *IEEE Trans. Image Process.*, vol. 29, pp. 7166–7175, 2020.
- [32] NIST, "IREX III - Performance of iris identification algorithms," NIST Interagency, Nat. Inst. Sci. Technol., Gaithersburg, MD, USA, Tech. Rep. 7836, 2012.
- [33] C. A. of Sciences Institute of Automation. CASIA iris image database, 2017. [Online]. Available: <http://biometrics.idealtest.org/>
- [34] P. J. Phillips *et al.*, "FRVT 2006 and ICE 2006 large-scale experimental results," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 5, pp. 831–846, May 2010.
- [35] J. K. Pillai, V. M. Patel, R. Chellappa, and N. K. Ratha, "Secure and robust iris recognition using random projections and sparse representations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 9, pp. 1877–1893, Sep. 2011.
- [36] H. Proenca, S. Filipe, R. Santos, J. Oliveira, and L. A. Alexandre, "The UBIRIS.v2: A database of visible wavelength images captured on-the-move and at-a-distance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 8, pp. 1529–1535, Aug. 2010.
- [37] H. Proença and J. C. Neves, "IRINA: Iris recognition (Even) in inaccurately segmented data," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 6747–6756.
- [38] G. Quinn, J. Matey, and P. Grother, "IREX IX part one performance of iris recognition algorithms," *Nat. Inst. Standards Technol.*, early access, Apr. 2018, doi: [10.6028/NIST.IR.8207](https://doi.org/10.6028/NIST.IR.8207).
- [39] O. Rippel, J. Snoek, and R. P. Adams, "Spectral representations for convolutional neural networks," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2015, pp. 2449–2457.
- [40] A. Ross, "Iris recognition: The path forward," *IEEE Comput.*, vol. 43, no. 2, pp. 30–35, Feb. 2010.
- [41] S. S. Sarwar, P. Panda, and K. Roy, "Gabor filter assisted energy efficient fast learning convolutional neural networks," in *Proc. IEEE/ACM Int. Symp. Low-Power Electron. Des.*, 2017, pp. 1–6.
- [42] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 815–823.
- [43] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 618–626.
- [44] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2017.
- [45] A. Stylianou, R. Souvenir, and R. Pless, "Visualizing deep similarity networks," in *Proc. IEEE Winter Conf. Appl. Comput. Vis.*, 2019, pp. 2029–2037.
- [46] Z. Sun and T. Tan, "Ordinal measures for iris recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 12, pp. 2211–2226, Dec. 2009.
- [47] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Inception-v4, inception-ResNet and the impact of residual connections on learning," in *Proc. AAAI Conf. Artif. Intell.*, 2017, pp. 4278–4284.
- [48] C.-W. Tan and A. Kumar, "Accurate iris recognition at a distance using stabilized iris encoding and Zernike moments phase features," *IEEE Trans. Image Process.*, vol. 23, no. 9, pp. 3962–3974, Sep. 2014.
- [49] M. Tygert, J. Bruna, S. Chintala, Y. LeCun, S. Piantino, and A. Szlam, "A mathematical motivation for complex-valued convolutional networks," *Neural Comput.*, vol. 28, pp. 815–825, 2016.
- [50] P. Virtue, S. X. Yu, and M. Lustig, "Better than real: Complex-valued neural nets for MRI fingerprinting," in *Proc. IEEE Int. Conf. Image Process.*, 2017, pp. 3953–3957.
- [51] C. Wang, J. Muhammad, Y. Wang, Z. He, and Z. Sun, "Towards complete and accurate iris segmentation using deep multi-task attention network for non-cooperative iris recognition," *IEEE Trans. Inf. Forensics Secur.*, vol. 15, pp. 2944–2959, 2020.
- [52] K. Wang and A. Kumar, "Toward more accurate iris recognition using dilated residual features," *IEEE Trans. Inf. Forensics Secur.*, vol. 14, no. 12, pp. 3233–3245, Dec. 2019.
- [53] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 818–833.
- [54] Z. Zhang, H. Wang, F. Xu, and Y.-Q. Jin, "Complex-valued convolutional neural network and its application in polarimetric SAR image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 12, pp. 7177–7188, Dec. 2017.
- [55] Z. Zhao and A. Kumar, "Towards more accurate iris recognition using deeply learned spatially corresponding features," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 3829–3838.
- [56] Z. Zhao and A. Kumar, "A deep learning based unified framework to detect, segment and recognize irises using spatially corresponding features," *Pattern Recognit.*, vol. 93, pp. 546–557, 2019.
- [57] S. Zhu, T. Yang, and C. Chen, "Visual explanation for deep metric learning," 2020, *arXiv:1909.12977*.



**Kien Nguyen** (Member, IEEE) is currently a research fellow with the Queensland University of Technology. He has been conducting research in the area of iris recognition and biometrics for the last 10 years, and has published his research in high quality journals and conferences in the area. His research interests include application of computer vision and deep learning techniques to the areas of biometrics, surveillance, and scene understanding. He has been serving as an associate editor of the journal *IEEE Access* since 2016.



**Clinton Fookes** (Senior Member, IEEE) received the BEng (Aero/Av) degree, the MBA degree, and the PhD degree in computer vision. He is currently a professor in vision and signal processing with the Queensland University of Technology. He actively researches across computer vision, machine learning, signal processing, and pattern recognition areas. He serves on the editorial boards of *IEEE Transactions on Image Processing*, *Pattern Recognition*, and *IEEE Transactions on Information Forensics and Security*. He is an Australian Institute of Policy and Science Young Tall Poppy, an Australian Museum Eureka Prize winner, and a Senior Fulbright Scholar.



**Sridha Sridharan** (Life Senior Member, IEEE) received the MSc degree from the University of Manchester, Manchester, U.K., and the PhD degree from the University of New South Wales, Kensington, Australia. He is currently a professor with the Queensland University of Technology (QUT) where he leads the research program in signal processing, artificial intelligence and vision technologies (SAIVT).



**Arun Ross** (Senior Member, IEEE) is currently a professor with Michigan State University and the director of the iProBe Lab. He is the co-author of the books *Handbook of Multibiometrics* and *Introduction to Biometrics*. He is a recipient of the IAPR JK Aggarwal Prize, IAPR Young Biometrics Investigator Award, and the NSF CAREER Award.

► For more information on this or any other computing topic, please visit our Digital Library at [www.computer.org/csdl](http://www.computer.org/csdl).