IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

SPECIAL SECTION ON VISUAL SURVEILLANCE AND BIOMETRICS: PRACTICES, CHALLENGES, AND POSSIBILITIES

Received October 23, 2017, accepted December 9, 2017, date of publication December 18, 2017, date of current version April 23, 2018.

Digital Object Identifier 10.1109/ACCESS.2017.2784352

Iris Recognition With Off-the-Shelf CNN Features: A Deep Learning Perspective

KIEN NGUYEN[1], (Member, IEEE), CLINTON FOOKES[1], (Senior Member, IEEE), ARUN ROSS[2], (Senior Member, IEEE), AND SRIDHA SRIDHARAN[1], (Senior Member, IEEE)
[1]Image and Video Lab, Queensland University of Technology, Brisbane, QLD 4000, Australia
[2]Department of Computer Science and Engineering, Michigan State University, East Lansing, MI 48824, USA

Corresponding author: Kien Nguyen (k.nguyenthanh@qut.edu.au)

ABSTRACT Iris recognition refers to the automated process of recognizing individuals based on their iris patterns. The seemingly stochastic nature of the iris stroma makes it a distinctive cue for biometric recognition. The textural nuances of an individual's iris pattern can be effectively extracted and encoded by projecting them onto Gabor wavelets and transforming the ensuing phasor response into a binary code - a technique pioneered by Daugman. This textural descriptor has been observed to be a robust feature descriptor with very low false match rates and low computational complexity. However, recent advancements in deep learning and computer vision indicate that generic descriptors extracted using convolutional neural networks (CNNs) are able to represent complex image characteristics. Given the superior performance of CNNs on the ImageNet large scale visual recognition challenge and a large number of other computer vision tasks, in this paper, we explore the performance of state-of-the-art pre-trained CNNs on iris recognition. We show that the off-the-shelf CNN features, while originally trained for classifying generic objects, are also extremely good at representing iris images, effectively extracting discriminative visual features and achieving promising recognition results on two iris datasets: ND-CrossSensor-2013 and CASIA-Iris-Thousand. We also discuss the challenges and future research directions in leveraging deep learning methods for the problem of iris recognition.

INDEX TERMS Iris recognition, biometrics, deep learning, convolutional neural network.

I. INTRODUCTION

Iris recognition refers to the automated process of recognizing individuals based on their iris patterns. Iris recognition algorithms have demonstrated very low false match rates and very high matching efficiency in large databases. This is not entirely surprising given the (a) complex textural pattern of the iris stroma that varies significantly across individuals, (b) the perceived permanence of its distinguishing attributes, and (c) its limited genetic penetrance [1]–[5]. A large-scale evaluation conducted by the National Institute of Science and Technology (NIST) has further highlighted the impressive recognition accuracy of iris recognition in operational scenarios [6], [7]. According to a report from 2014 [8], over one billion people worldwide have had their iris images electronically enrolled in various databases across the world. This includes about 1 billion people in the Unique IDentification Authority of India (UIDAI) program, 160 million people from the national ID program of Indonesia, and 10 million people from the US Department of Defense program. Thus, the iris is likely to play a critical role in next generation large-scale identification systems.

A. STATE-OF-THE-ART LITERATURE

The success of iris recognition - besides its attractive physical characteristics - is rooted in the development of efficient feature descriptors, especially the Gabor phase-quadrant feature descriptor introduced in Daugman's pioneering work [3], [5], [9]. This Gabor phase-quadrant feature descriptor (often referred to as the iriscode) has dominated the iris recognition field, exhibiting very low false match rates and high matching efficiency. Researchers have also proposed a wide range of other descriptors for iris based on Discrete Cosine Transforms (DCT) [10], Discrete Fourier Transforms (DFT) [11], ordinal measures [12], class-specific weight maps [13], compressive sensing and sparse coding [14], hierarchical visual codebooks [15], multi-scale Taylor expansion [16], [17], etc. Readers are referred to [18]–[21] for an extensive list of methods that have been appropriated for iris recognition.

B. GAP

Given the widespread use of classical texture descriptors for iris recognition, including the Gabor phase-quadrant feature descriptor, it is instructive to take a step back and answer the following question: how do we know that these hand-crafted feature descriptors proposed in the literature are actually the best representations for the iris? Furthermore, can we achieve better performance (compared to the Gabor-based approach) by designing a novel feature representation scheme that can perhaps attain the upper bound on iris recognition accuracy with low computational complexity?

C. POTENTIAL SOLUTION

One possible solution is to leverage the recent advances in *Deep Learning* to discover a feature representation scheme that is primarily data-driven. By automatically learning the feature representation from the iris data, an optimal representation scheme can potentially be deduced, leading to high recognition results for the iris recognition task. Deep learning techniques often use hierarchical multi-layer networks to elicit feature maps that optimize performance on the training data [22]. These networks allow for the feature representation scheme to be learned and discovered directly from data, and avoid some of the pitfalls in developing hand-crafted features. Deep learning has completely transformed the performance of many computer vision tasks [23], [24]. Therefore, we hypothesize that deep learning techniques, as embodied by Convolutional Neural Networks (CNNs), can be used to design alternate feature descriptors for the problem of iris recognition.

D. WHY ARE DEEP IRIS METHODS NOT WIDELY USED AS YET?

There have been a few attempts to appropriate the principles of deep learning to the task of iris recognition [25], [26]. The limited application of deep learning methods to the problem of iris recognition is due to the fact that deep learning requires a huge amount of training data, which is not available to most iris researchers at this time. In addition, deep learning is also very computationally expensive and requires the power of multiple Graphical Processing Units (GPUs). This is a deterrent to the physical implementation of such deep learning approaches. Most importantly, to date, there has been no insight into why deep learning should work for iris recognition, and no systematic analysis has been conducted to ascertain how best to capitalize on modern deep approaches to design an optimal architecture of deep networks to achieve high accuracy and low computational complexity. Simply stacking multiple layers to design a CNN for iris recognition without intuitive insights would be infeasible (due to the lack of large-scale iris datasets in the public domain), non-optimal (due to *ad hoc* choices for CNN architecture, number of layers, configuration of layers...) and inefficient (due to redundant layers).

We argue that rather than designing and training new CNNs for iris recognition, using CNNs whose architectures have been proven to be successful in large-scale computer vision challenges should yield good performance without the time-consuming architecture design step. A major source of state-of-the-art CNNs is from the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [27] organized annually to evaluate state-of-the-art algorithms for large-scale object detection and image classification. The networks developed as part of this challenge are generally made available in the public domain for extracting deep features from images. Researchers have shown that these off-the-shelf CNN features are extremely effective for various computer vision tasks, such as facial expression classification, action recognition and visual instance retrieval, and are not restricted to just the object detection and image classification tasks for which they were designed [28]. In this paper, we will investigate the performance of those CNNs that won the ILSVRC challenge since 2012 (before 2012, the winners were non-CNN methods that did not perform as well as CNN-based approaches).

The main contributions of this paper are as follows:

- First, we analyze the deep architectures which have been proposed in the literature for iris recognition.
- Second, we appropriate off-the-shelf CNNs to the problem of iris recognition and present our preliminary results using them.
- Third, we discuss the challenges and the future of deep learning for iris recognition.

The remainder of the paper is organized as follows: Section II briefly introduces CNNs: Section II-A discusses related work in general CNNs, while Section II-B discusses CNNs that have been proposed for iris recognition. Section III describes the off-the-shelf CNNs that were used in this work, and our proposed framework for investigating the performance of these CNNs. Section IV presents the experimental results. Section V concludes the paper with a discussion on future work.

II. RELATED WORK

A. CNNs - CONVOLUTIONAL NEURAL NETWORKS

Deep learning methods, especially convolutional neural networks (CNNs), have recently led to breakthroughs in many computer vision tasks such as object detection and recognition, and image segmentation and captioning [22]–[24]. By attempting to mimic the structure and operation of the neurons in the human visual cortex through the use of hierarchical multi-layer networks, deep learning has been shown to be extremely effective in automating the process of learning feature-representation schemes from the training data, thereby eliminating the laborious feature engineering task. CNNs belong to a specific category of deep learning methods designed to process images and videos. By using repetitive blocks of neurons in the form of a convolution layer that is applied hierarchically across images, CNNs have not only been able to automatically learn image feature

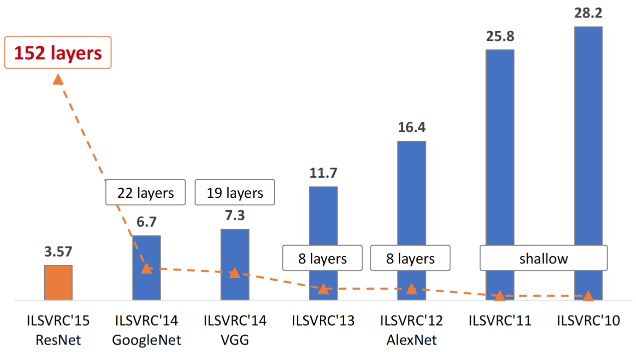


FIGURE 1. The evolution of the winning entries on the ImageNet Large Scale Visual Recognition Challenge from 2010 to 2015. Since 2012, CNNs have outperformed hand-crafted descriptors and shallow networks by a large margin. Image re-printed with permission [36].

representations, but they have also outperformed many conventional hand-crafted feature techniques [29].

In the 1960s, Hubel and Wiesel [30] found that cells in the animal visual cortex were responsible for detecting light in receptive fields and constructing an image. Further, they demonstrated that this visual field could be represented using a topographic map. Later, Fukushima and Miyake [31] proposed the NeoCognitron, which could be regarded as the predecessor of the CNN. The seminal modern CNN was introduced by Yan Lecun *et al.* [32] in the 1990s for Handwritten Digit Recognition with an architecture called LeNet. Many features of modern deep networks are derived from the LeNet, where convolutional connections were introduced and a backpropagation algorithm was used to train the network. CNNs became exceptionally popular in 2012 when Krizhesky *et al.* introduced a CNN called AlexNet, which significantly outperformed previous methods on the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) [33]. The AlexNet is simply a scaled version of the LeNet with a deeper structure, but is trained on a much larger dataset (ImageNet with 14 million images) with a much more powerful computational resource (GPUs).

Since then, many novel architectures and efficient learning techniques have been introduced to make CNNs deeper and more powerful [34]–[37], achieving revolutionary performance in a wide range of computer vision applications. The annual ILSVRC event has become an important venue to recognize the performance of new CNN architectures, especially with the participation of technology giants like Google, Microsoft and Facebook. The depth of the "winning" CNNs has progressively increased from 8 layers in 2012 to 152 layers in 2015, while the recognition error rate has significantly dropped from 16.4% in 2012 to 3.57% in 2015. This phenomenal progress is illustrated in Figure 1.

Pre-trained CNNs have been open-sourced and widely used in other applications and show very promising performance [28].

B. CNNs FOR IRIS RECOGNITION IN THE LITERATURE

A number of deep networks have been proposed for improving the performance of iris recognition. Liu *et al.* proposed a DeepIris network of 9 layers consisting of one pairwise filter layer, one convolutional layer, two pooling layers, two normalization layers, two local layers and one fully-connected layer [38]. This deep network achieved a very promising recognition rate on both the Q-FIRE [39] and CASIA [40] datasets. Gangwar and Joshi [25] employed more advanced layers to create two DeepIrisNets for the iris recognition task [25]. The first network, DeepIrisNet-A, contained 8 convolutional layers (each followed by a batch normalization layer), 4 pooling layers, 3 fully connected layers and two drop-out layers. The second network, DeepIrisNet-B, added two inception layers to increase the modeling capability. These two networks exhibited superior performance on the ND-IRIS-0405 [41] and ND-CrossSensor-Iris-2013 [41] datasets. It is worth mentioning that CNNs have also been used in the iris biometrics community for iris segmentation [42], [43], spoof detection [44], [45] and gender classification [46]. While self-designed CNNs such as DeepIris [38] and DeepIrisNet [25] have shown promising results, their major limit lies in the design of the network since the choice on the number of layers is limited by the number of training samples. The largest public dataset that is currently available is the ND-CrossSensor-2013 dataset which contains only 116,564 iris images. This number is far from the millions of parameters that embody any substantially deep neural network.

To deal with the absence of a large iris dataset, transfer learning can be used. Here, CNNs that have been trained on other large datasets such as ImageNet [27], can be appropriated directly to the iris recognition domain. In fact, CNN models pre-trained on ImageNet, have been successfully transferred to many computer vision tasks [28]. Minaee *et al.* [26] showed that the VGG model, even though pre-trained on ImageNet to classify objects from different categories, works reasonably well for the task of iris recognition. However, since the release of the VGG model in 2014, many other advanced architectures have been proposed in the literature. In this paper, we will harness such CNN architectures - primarily those that have won the ImageNet challenge - for the iris recognition task.

III. METHODS - OFF-THE-SHELF CNN FEATURES FOR IRIS RECOGNITION

Considering the dominance of CNNs in the computer vision field and inspired by recent research [28] which has shown that Off-the-Shelf CNN Features work very well for multiple classification and recognition tasks, we investigate the performance of state-of-the-art CNNs pre-trained on the ImageNet dataset for the iris recognition task. We first review some popular CNN architectures and then present our framework for iris recognition using these CNN Features.

A. CNNs IN USE

We will now analyze each architecture in detail and highlight their notable properties.

1) ALEXNET

ILSVRC 2012 winner: In 2012, Krizhevsky *et al.* [33] achieved a breakthrough in the large-scale ILSVRC challenge, by utilizing a deep CNN that significantly outperformed other hand-crafted features resulting in a top-5 error rate of 16.4%. AlexNet is actually a scaled version of the conventional LeNet, and takes advantage of a large-scale training dataset (ImageNet) and more computational power (GPUs that allow for 10x speed-up in training). Tuning the hyperparameters of AlexNet was observed to result in better performance, subsequently winning the ILSVCR 2013 challenge [47]. The detailed architecture of AlexNet is presented in the Appendix. In this paper, we extract the outputs of all convolutional layers (5) and all fully connected layers (2) to generate the CNN Features for the iris recognition task.

2) VGG

ILSVRC 2014 runner-up: In 2014, Simonyan and Zisserman from Oxford showed that using smaller filters (3×3) in each convolutional layer leads to improved performance. The intuition is that multiple small filters in sequence can emulate the effects of larger ones. The simplicity of using small sized filters throughout the network leads to very good generalization performance. Based on these observations, they introduced a network called VGG which is still widely used today due to its simplicity and good generalization performance [34]. Multiple versions of VGG has been introduced, but the two most popular ones are VGG-16 and VGG-19 that contain 16 and 19 layers, respectively. The detailed architecture of VGG is presented in the Appendix. In this paper, we extract the outputs of all convolutional layers (16) and all fully connected layers (2) to generate the CNN Features for the iris recognition task.

3) GoogLeNet AND INCEPTION

ILSVRC 2014 winner: In 2014, Szegedy *et al.* from Google introduced the Inception v1 architecture that was implemented in the winning ILSVRC 2014 submission called GoogLeNet with a top-5 error rate of 6.7%. The main innovation is the introduction of an inception module, which functions as a small network inside a bigger network [35]. The new insight was the use of 1×1 convolutional blocks to aggregate and reduce the number of features before invoking the expensive parallel blocks. This helps in combining convolutional features in a better way that is not possible by simply stacking more convolutional layers. Later, the authors introduced some improvements in terms of batch normalization, and re-designed the filter arrangement in the inception module to create Inception v2 and v3 [48]. Most recently, they added residual connections to improve the gradient flows in Inception v4 [49]. The detailed architecture of Inception v3 is presented in the Appendix. In this paper, we extract the outputs of all convolutional layers (5) and all inception layers (12) to generate the CNN Features for the iris recognition task.

4) RESNET

ILSVRC 2015 winner: In 2015, He *et al.* from Microsoft introduced the notion of residual connection or skip connection which feeds the output of two successive convolutional layers and bypasses the input to the next layer [36]. This residual connection improves the gradient flow in the network, allowing the network to become very deep with 152 layers. This network won the ILSVRC 2015 challenge with a top-5 error rate of 3.57%. The detailed architecture of ResNet-152 is presented in the Appendix. In this paper, we extract the outputs of all convolutional layers (1) and all bottleneck layers (17) to generate the CNN Features for the iris recognition task.

5) DENSENET

In 2016, Huang *et al.* [37] from Facebook proposed DenseNet, which connects each layer of a CNN to every other layer in a feed-forward fashion. Using densely connected architectures leads to several advantages as pointed out by the authors: "alleviating the vanishing-gradient problem, strengthening feature propagation, encouraging feature reuse, and substantially reducing the number of parameters". The detailed architecture of DenseNet-201 is presented in the Appendix. In this paper, we extract the outputs of a selected number of dense layers (15) to generate the CNN Features for the iris recognition task.

It is worth noting that there are several other powerful CNN architectures in the literature [29], [50]. However, we have only chosen the above architectures for illustrating the performance of pre-trained CNNs on the iris recognition task.

B. IRIS RECOGNITION FRAMEWORK USING CNN FEATURES

The framework we employ to investigate the performance of off-the-shelf CNN Features for iris recognition is summarized in Figure 2.

1) SEGMENTATION

The iris is first localized by extracting two circular contours pertaining to the inner and outer boundaries of the iris region. The integro-differential operator, one of the most commonly used circle detectors, can be mathematically expressed as,

$$max_{r,x_0,y_0}\left|G_\sigma(r) * \frac{\partial}{\partial r}\oint_{r,x_0,y_0}\frac{I(x,y)}{2\pi r}ds\right|, \tag{1}$$

where $I(x, y)$ and G_σ denote the input image and a Gaussian blurring filter, respectively. The symbol $*$ denotes a convolution operation and r represents the radius of the circular arc ds, centered at the location (x_0, y_0). The described operation detects circular edges by iteratively searching for the maximum responses of a contour defined by the parameters

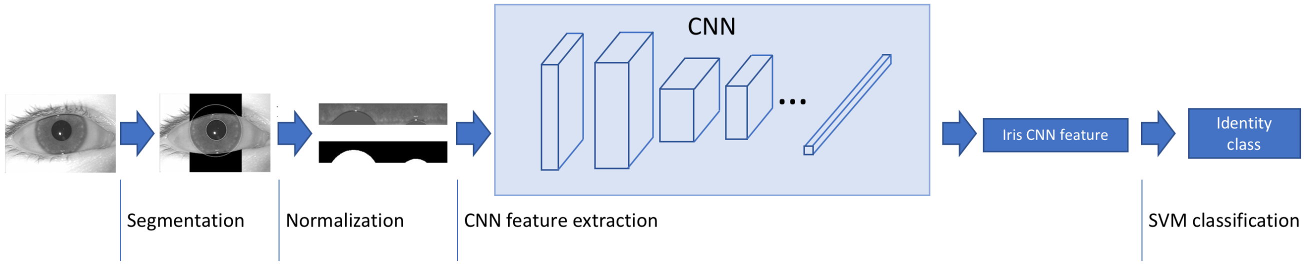


FIGURE 2. The framework for iris recognition using Off-the-shelf CNN Features. The iris image is segmented using two circular contours and then geometrically normalized using a pseudo-polar transformation resulting in a fixed rectangular image. Features are next extracted using off-the-shelf CNNs, and then classified using an SVM.

(x_0, y_0, r). In most cases, the iris region can be obscured by the upper and lower eyelids and eyelashes. In such images, the eyelids can be localized using the above operator with the path of contour integration changed from a circle to an arc. Noise masks distinguish the iris-pixels from the non-iris pixels (e.g., eyelashes, eyelids, etc.) in a given image. Such noise masks, corresponding to each input image, are generated during the segmentation stage and used in the subsequent steps.

2) NORMALIZATION

The area enclosed by the inner and outer boundaries of an iris can vary due to the dilation and contraction of the pupil. The effect of such variations need to be minimized before comparing different iris images. To this end, the segmented iris region is typically mapped to a region of fixed dimension. Daugman proposed the usage of a rubber-sheet model to transform the segmented iris to a fixed rectangular region. This process is carried out by re-mapping the iris region, $I(x, y)$, from the raw Cartesian coordinates (x, y) to the dimensionless polar coordinates (r, θ), and can be mathematically expressed as,

$$I(x(r, \theta), y(r, \theta)) \rightarrow I(r, \theta), \tag{2}$$

where r is in the unit interval [0,1], and θ is an angle in the range of $[0,2\pi]$. $x(r, \theta)$ and $y(r, \theta)$ are defined as the linear combination of both pupillary $(x_p(\theta), y_p(\theta))$ and limbic boundary points $(x_s(\theta), y_s(\theta))$ as,

$$x(r, \theta) = (1 - r)x_p(\theta) + rx_s(\theta), \tag{3}$$
$$y(r, \theta) = (1 - r)y_p(\theta) + ry_s(\theta). \tag{4}$$

An additional benefit of normalization is that the rotations of the eye (e.g., due to the movement of the head) are reduced to simple translations during matching. The corresponding noise mask is also normalized to facilitate easier matching in the later stages.

3) CNN FEATURE EXTRACTION

The normalized iris image is then fed into the CNN feature extraction module. As discussed earlier, five state-of-the-art and off-the-shelf CNNs (AlexNet, VGG, Google Inception, ResNet and DenseNet) are used in this work to extract features from the normalized iris images. Note that there are multiple layers in each CNN. Every layer models different levels of visual content in the image, with later layers encoding finer and more abstract information, and earlier layers retaining coarser information. One of the key reasons why CNNs work very well on computer vision tasks is that these deep networks with tens or hundreds of layers and millions of parameters are extremely good at capturing and encoding complex features of the images, leading to superior performance. To investigate the representation capability of each layer for the iris recognition task, we employ the output of each layer as a feature descriptor and report the corresponding recognition accuracy.

4) SVM CLASSIFICATION

The extracted CNN feature vector is then fed into the classification module. We use a simple multi-class Support Vector Machine (SVM) [51] due to its popularity and efficiency in image classification. The multi-class SVM for N classes is implemented as a one-against-all strategy, which is equivalent to combining N binary SVM classifiers, with every classifier discriminating a single class against all other classes. The test sample is assigned to the class with the largest margin [51].

IV. EXPERIMENTAL RESULTS

A. DATASETS

We conducted our experiments on two large iris datasets: 1) LG2200 dataset: ND-CrossSensor-Iris-2013 is the largest public iris dataset in the literature in terms of the number of images [41]. The ND-CrossSensor-2013 dataset contains 116,564 iris images captured by the LG2200 iris camera from 676 subjects. 2) CASIA-Iris-Thousand: This contains 20,000 iris images from 1,000 subjects, which were collected using the IKEMB-100 camera from IrisKing [40]. Some samples images from the two datasets are depicted in Figure 3.

B. PERFORMANCE METRIC AND BASELINE METHOD

To report the performance, we rely on the Recognition Rate. Recognition Rate is calculated as the proportion of correctly classified samples at a pre-defined False Acceptance Rate (FAR). In this work, we choose to report Recognition Rate at FAR = 0.1%.

The baseline feature descriptor we used for comparison is the Gabor phase-quadrant feature [3]. The popular

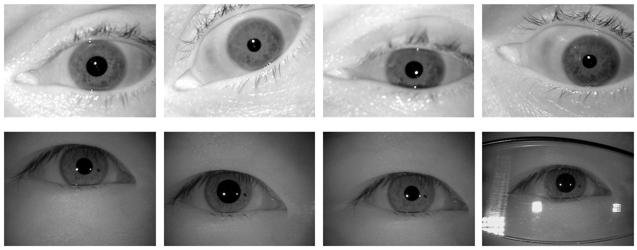


FIGURE 3. Sample images from the LG2200 (first row) and CASIA-Iris-Thousand (second row) datasets.

matching operator coupled with this descriptor is the Hamming distance [3]. This baseline achieved recognition accuracies of 91.1% and 90.7% on the LG2200 and CASIA-Iris-Thousand datasets, respectively.

C. EXPERIMENTAL SETUP

The left and right iris images of every subject are treated as two different classes. Thus, the LG2200 dataset has 1,352 classes and the CASIA-Iris-Thousand has 2,000 classes. We randomly choose 70% of the data corresponding to each class for training and the rest (30%) for testing. It must be noted that the training images were used *only* to train the multi-class SVMs; the pre-trained CNNs were *not* modified at all using the training data. This is one of the main advantage of using pre-trained CNNs.

For iris segmentation and normalization, we used an open-source software, USIT v2.2, from the University of Salzburg [52]. This software takes each iris image as an input, segments it using inner and outer circles, and normalizes the segmented region into a rectangle of size 64×256 pixels.

For CNN feature extraction, we implemented our approach using PyTorch [53]. PyTorch is a recently released deep learning framework by Facebook combining the advantages of both Torch and Python. The two most advanced features of this framework are dynamic graph computation and imperative programming, which make deep network coding more flexible and powerful [53]. With regards to our experiments, PyTorch supplies a wide range of pre-trained off-the-shelf CNNs, making our feature extraction task that much more convenient.

For classification, we used the LIBSVM library [54] with a Python wrapper implemented in the scikit-learn library [55] which allows for ease of integration with the feature extraction step.

D. PERFORMANCE ANALYSIS

As stated earlier, different layers encode different levels of visual content. To investigate the performance due to each layer, we estimate the recognition accuracy after using the output from each layer as a feature vector to represent the iris. The recognition accuracies are illustrated in Figure 4 for the two datasets: LG2200 and CASIA-Iris-Thousand.

Interestingly, the recognition accuracy peaks in some middle layers for all CNNs. On the LG2200 dataset: layer 10 for VGG, layer 10 for Inception, layer 11 for ResNet and layer 6 for DenseNet. On the CASIA-Iris-Thousand dataset: layer 9 for VGG, layer 10 for Inception, layer 12 for ResNet and layer 5 for DenseNet. The difference in the "peak layers" can be explained by the properties of each CNN. Since Inception uses intricate inception layers (actually each layer is a network inside a larger network), it quickly converges to the peak than others. In contrast, ResNet with its skip connections is very good at allowing the gradient to flow through the network, making the network perform well at a deeper depth, leading to a later peak in the iris recognition accuracy. DenseNet, with its rich dense connections, allows neurons to interact easily, leading to the best recognition accuracy among all CNNs for the iris recognition task.

As can be seen, peak results do not occur toward the later layers of the CNNs. This can be explained by the fact that the normalized iris image is not as complex as the images in the ImageNet dataset where large structural variations are present in a wide range of objects. Hence, it is not necessary to have a large number of layers to encode the normalized iris. Thus, peak accuracies are achieved in the middle layers.

Among all five CNNs, DenseNet achieves the highest peak recognition accuracy of 98.7% at layer 6 on the LG2200 dataset and 98.8% at layer 5 on the

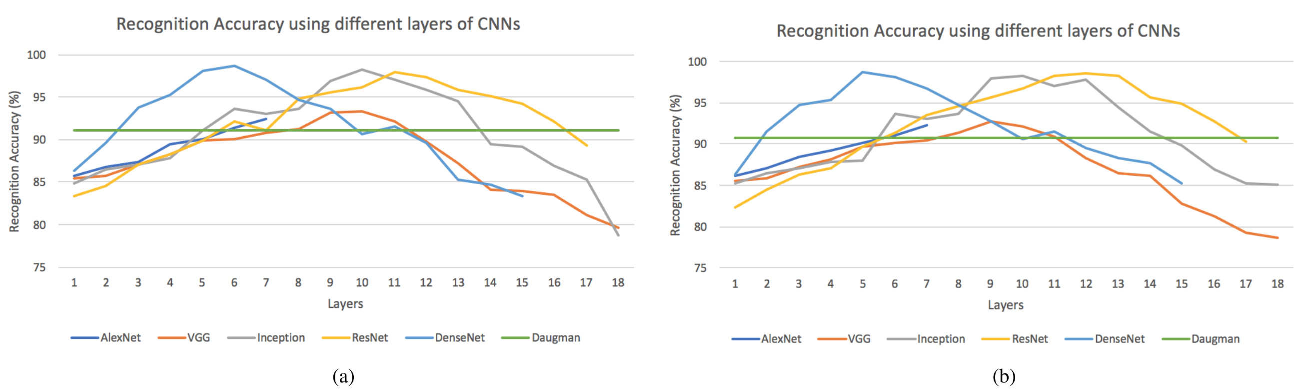


FIGURE 4. Recognition accuracy of different layers in the CNNs on the two datasets: (a) LG2200 and (b) CASIA-Iris-Thousand.

CASIA-Iris-Thousand dataset. ResNet and Inception achieve similar peak recognition accuracies of 98.0% and 98.2% at layers 11 and 10, respectively, on the LG2200 dataset; and 98.5% and 98.3% at layers 12 and 10, respectively, on the CASIA-Iris-Thousand dataset. VGG, with its simple architecture, only achieves 92.7% and 93.1% recognition accuracy, both at layer 9, on the LG2200 and CASIA-Iris-Thousand datasets, respectively. The continually increasing recognition accuracy of AlexNet indicates that the number of layers considered in its architecture may not fully capture the discriminative visual features in iris images.

V. DISCUSSIONS AND CONCLUSION

In this paper, we have approached the task of iris recognition from a deep learning point of view. Our experiments have shown that off-the-shelf pre-trained CNN features, even though originally trained for the problem of object recognition, can be appropriated to the iris recognition task. By harnessing state-of-the-art CNNs from the ILSVRC challenge and applying them on the iris recognition task, we achieve state-of-the-art recognition accuracy in two large iris datasets, viz., ND-CrossSensor-2013 and CASIA-Iris-Thousand. These preliminary results show that off-the-shelf CNN features can be successfully transferred to the iris recognition problem, thereby effectively extracting discriminative visual features in iris images and eliminating the laborious feature-engineering task. The benefit of CNNs in automated feature-engineering is critical in learning new iris encoding schemes that can benefit large scale applications.

A. OPEN PROBLEMS

This work, together with previous work on DeepIris [38] and DeepIrisNet [25], indicates that CNNs are effective in encoding discriminative features for iris recognition. Notwithstanding this observation, there are several challenges and open questions when applying deep learning to the problem of iris recognition.

- **Computational complexity**: The computational complexity of CNNs during the training phase is very high due to the millions of parameters used in the network. In fact, powerful GPUs are needed to accomplish training. This compares unfavorably with hand-crafted iris features, especially Daugman's Gabor features, where thousands of iriscodes can be extracted and compared on a general purpose CPU within a second. Model reduction techniques such as pruning and compressing may, therefore, be needed to eliminate redundant neurons and layers, and to reduce the size of the network.
- **Domain adaptation and fine-tuning**: Another approach to use off-the-shelf CNNs is by fine-tuning, which would entail freezing the early layers and only re-training a few selected later layers to adapt the representation capability of the CNNs to iris images. Fine-tuning is expected to learn and encode iris-specific features, as opposed to generic image features. In addition, domain adaptation can be used to transform the representation from the ImageNet domain to the iris image domain.
- **Few-shot learning**: If the network for iris recognition has to be designed and trained from scratch, the problem of limited number of training images can be partially solved with a technique called few-shot learning, which would allow the network to perform well after seeing very few samples from each class.
- **Architecture evolution**: Recent advances in Evolution Theory and Deep Reinforcement Learning allow the network to change itself and generate better instances for the problem at hand. Such an approach can be used to evolve off-the-shelf CNNs in order to generate powerful networks that are more suited for iris recognition.
- **Other architectures**: In the field of deep learning, there are other architectures such as unsupervised Deep Belief Network (DBN), Stacked Auto-Encoder (SAE) and Recurrent Neural Network (RNN). These architectures have their own advantages and can be used to extract features for iris images. They can be used either separately or in combination with classical CNNs to improve the representation capacity of iris templates.

REFERENCES

[1] A. Muron and J. Pospisil, "The human iris structure and its usages," *Acta Univ. Plalcki Physica*, vol. 39, pp. 87–95, Mar. 2000.
[2] A. K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 1, pp. 4–20, Jan. 2004.
[3] J. Daugman, "How iris recognition works," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 1, pp. 21–30, Jan. 2004.
[4] J. Daugman and C. Downing, "Searching for doppelgängers: Assessing the universality of the IrisCode impostors distribution," *IET Biometrics*, vol. 5, no. 2, pp. 65–75, 2016.
[5] J. Daugman, "Information theory and the iriscode," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 2, pp. 400–409, Feb. 2016.
[6] NIST, "IREX III—Performance of iris identification algorithms," Nat. Inst. Sci. Technol., Gaithersburg, MD, USA, Tech. Rep. NIST Interagency 7836, 2012.
[7] NIST, "IREX IV—Evaluation of iris identification algorithms," Nat. Inst. Sci. Technol., Gaithersburg, MD, USA, Tech. Rep. NIST Interagency 7949, 2013.
[8] J. Daugman, "Major international deployments of the iris recognition algorithms: A billion persons," Univ. Cambridge, Cambridge, U.K., Tech. Rep., Dec. 2014.
[9] J. Daugman, "New methods in iris recognition," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 37, no. 5, pp. 1167–1175, Oct. 2007.
[10] 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.
[11] 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, Sep. 2008.
[12] 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.
[13] W. Dong, Z. Sun, and T. Tan, "Iris matching based on personalized weight map," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 9, pp. 1744–1757, Sep. 2011.
[14] 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.
[15] Z. Sun, H. Zhang, T. Tan, and J. Wang, "Iris image classification based on hierarchical visual codebook," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 6, pp. 1120–1133, Jun. 2014.
[16] A. Bastys, J. Kranauskas, and R. Masiulis, "Iris recognition by local extremum points of multiscale Taylor expansion," *Pattern Recognit.*, vol. 42, no. 9, pp. 1869–1877, 2009.

[17] A. Bastys, J. Kranauskas, and V. Krüger, ''Iris recognition by fusing different representations of multi-scale Taylor expansion,'' *Comput. Vis. Image Understand.*, vol. 115, no. 6, pp. 804–816, 2011.

[18] K. W. Bowyer, K. Hollingsworth, and P. J. Flynn, ''Image understanding for iris biometrics: A survey,'' *Comput. Vis. Image Understand.*, vol. 110, no. 2, pp. 281–307, 2008.

[19] K. W. Bowyer, K. Hollingsworth, and P. J. Flynn, ''A survey of iris biometrics research 2008–2010,'' in *Handbook Iris Recognition*. London, U.K.: Springer-Verlag, 2013, ch. 2.

[20] I. Nigam, M. Vatsa, and R. Singh, ''Ocular biometrics: A survey of modalities and fusion approaches,'' *Inf. Fusion*, vol. 26, pp. 1–35, Nov. 2015.

[21] K. Nguyen, C. Fookes, R. Jillela, S. Sridharan, and A. Ross, ''Long range iris recognition: A survey,'' *Pattern Recognit.*, vol. 72, pp. 123–143, Dec. 2017.

[22] Y. LeCun, Y. Bengio, and G. Hinton, ''Deep learning,'' *Nature*, vol. 521, pp. 436–444, May 2015.

[23] J. Schmidhuber, ''Deep learning in neural networks: An overview,'' *Neural Netw.*, vol. 61, pp. 85–117, Jan. 2015.

[24] 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.

[25] 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. (ICIP)*, Sep. 2016, pp. 2301–2305.

[26] S. Minaee, A. Abdolrashidiy, and Y. Wang, ''An experimental study of deep convolutional features for iris recognition,'' in *Proc. IEEE Signal Process. Med. Biol. Symp. (SPMB)*, Dec. 2016, pp. 1–6.

[27] O. Russakovsky *et al.*, ''ImageNet large scale visual recognition challenge,'' *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015.

[28] A. S. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, ''CNN features off-the-shelf: An astounding baseline for recognition,'' in *Proc. IEEE CVPR*, May 2014, pp. 512–519.

[29] A. Canziani, A. Paszke, and E. Culurciello, ''An analysis of deep neural network models for practical applications,'' *CoRR*, pp. 1–7, May 2016.

[30] D. H. Hubel and T. N. Wiesel, ''Receptive fields and functional architecture of monkey striate cortex,'' *J. Physiol.*, vol. 195, no. 1, pp. 215–243, 1968.

[31] K. Fukushima and S. Miyake, *Neocognitron: A Self-Organizing Neural Network Model for a Mechanism of Visual Pattern Recognition*. Berlin, Germany: Springer, 1982, pp. 267–285.

[32] Y. LeCun *et al.*, ''Handwritten digit recognition with a back-propagation network,'' in *Proc. Adv. Neural Inf. Process. Syst.*, 1990, pp. 396–404.

[33] A. Krizhevsky, I. Sutskever, and G. E. Hinton, ''ImageNet classification with deep convolutional neural networks,'' in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[34] K. Simonyan and A. Zisserman, ''Very deep convolutional networks for large-scale image recognition,'' *CoRR*, pp. 1–14, Sep. 2014.

[35] C. Szegedy *et al.*, ''Going deeper with convolutions,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.

[36] K. He, X. Zhang, S. Ren, and J. Sun, ''Deep residual learning for image recognition,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[37] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, ''Densely connected convolutional networks,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2261–2269.

[38] 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, no. 2, pp. 154–161, 2015.

[39] P. A. Johnson, P. Lopez-Meyer, N. Sazonova, F. Hua, and S. Schuckers, ''Quality in face and iris research ensemble (Q-FIRE),'' in *Proc. 4th IEEE Int. Conf. Biometrics, Theory, Appl. Syst. (BTAS)*, Sep. 2010, pp. 1–6.

[40] Chinese Academy of Sciences Institute of Automation. (Aug. 2017). *CASIA Iris Image Database*. [Online]. Available: http://biometrics.idealtest.org/

[41] P. J. Phillip *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.

[42] N. Liu, H. Li, M. Zhang, J. Liu, Z. Sun, and T. Tan, ''Accurate iris segmentation in non-cooperative environments using fully convolutional networks,'' in *Proc. Int. Conf. Biometrics (ICB)*, Jun. 2016, pp. 1–8.

[43] E. Jalilian and A. Uhl, *Iris Segmentation Using Fully Convolutional Encoder–Decoder Networks*. Cham, Switzerland: Springer, 2017, pp. 133–155.

[44] L. He, H. Li, F. Liu, N. Liu, Z. Sun, and Z. He, ''Multi-patch convolution neural network for iris liveness detection,'' in *Proc. IEEE 8th Int. Conf. Biometrics Theory, Appl. Syst. (BTAS)*, Sep. 2016, pp. 1–7.

[45] D. Menotti *et al.*, ''Deep representations for iris, face, and fingerprint spoofing detection,'' *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 4, pp. 864–879, Apr. 2015.

[46] J. Tapia and C. Aravena, *Gender Classification From NIR Iris Images Using Deep Learning*. Cham, Switzerland: Springer, 2017, pp. 219–239.

[47] M. D. Zeiler and R. Fergus, ''Visualizing and understanding convolutional networks,'' in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Aug. 2014, pp. 818–833.

[48] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, ''Rethinking the inception architecture for computer vision,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2818–2826.

[49] C. Szegedy, V. Vanhoucke, S. Ioffe, 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.

[50] J. Gu *et al.*, ''Recent advances in convolutional neural networks,'' *CoRR*, pp. 187–332, Dec. 2017.

[51] B. Scholkopf and A. J. Smola, *Learning With Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA, USA: MIT Press, 2001.

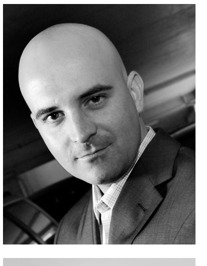
[52] C. Rathgeb, A. Uhl, P. Wild, and H. Hofbauer, ''Design decisions for an iris recognition SDK,'' in *Handbook of Iris Recognition* (Advances in Computer Vision and Pattern Recognition), 2nd ed., K. Bowyer and M. J. Burge, Eds. London, U.K.: Springer-Verlag, 2016.

[53] (Sep. 2017). *PyTorch Tensors and Dynamic Neural Networks in Python With Strong GPU Acceleration*. [Online]. Available: http://pytorch.org/

[54] C. C. Chang and C. J. Lin, ''LIBSVM: A library for support vector machines,'' *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–27, 2011.

[55] F. Pedregosa *et al.*, ''Scikit-learn: Machine learning in Python,'' *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Oct. 2011.

KIEN NGUYEN received the Ph.D. degree from the Queensland University of Technology (QUT), Australia. He has been researching iris recognition for seven years. He is currently a Post-Doctoral Research Fellow with QUT. His research interests are applications of computer vision and deep learning techniques for biometrics, surveillance, and scene understanding.



CLINTON FOOKES (SM'14) received the B.Eng. (Aerospace/Avionics), MBA, and Ph.D. degrees in computer vision from the Queensland University of Technology (QUT), Australia. He is currently a Professor in vision and signal processing with the QUT. He is a Senior Member of an AIPS Young Tall Poppy, an Australian Museum Eureka Prize winner, and a Senior Fulbright Scholar.

ARUN ROSS is currently a Professor with Michigan State University and the Director of the iPRoBe Lab. He has co-authored the books *Handbook of Multibiometrics* and *Introduction to Biometrics*. He was a recipient of the IAPR JK Aggarwal Prize, the IAPR Young Biometrics Investigator Award, and the NSF CAREER Award.

SRIDHA SRIDHARAN received the M.Sc. degree from the University of Manchester, Manchester, U.K., and the Ph.D. degree from the University of New South Wales, Australia. He is currently a Professor with the Queensland University of Technology where he leads the research program in Speech, Audio, Image and Video Technologies.