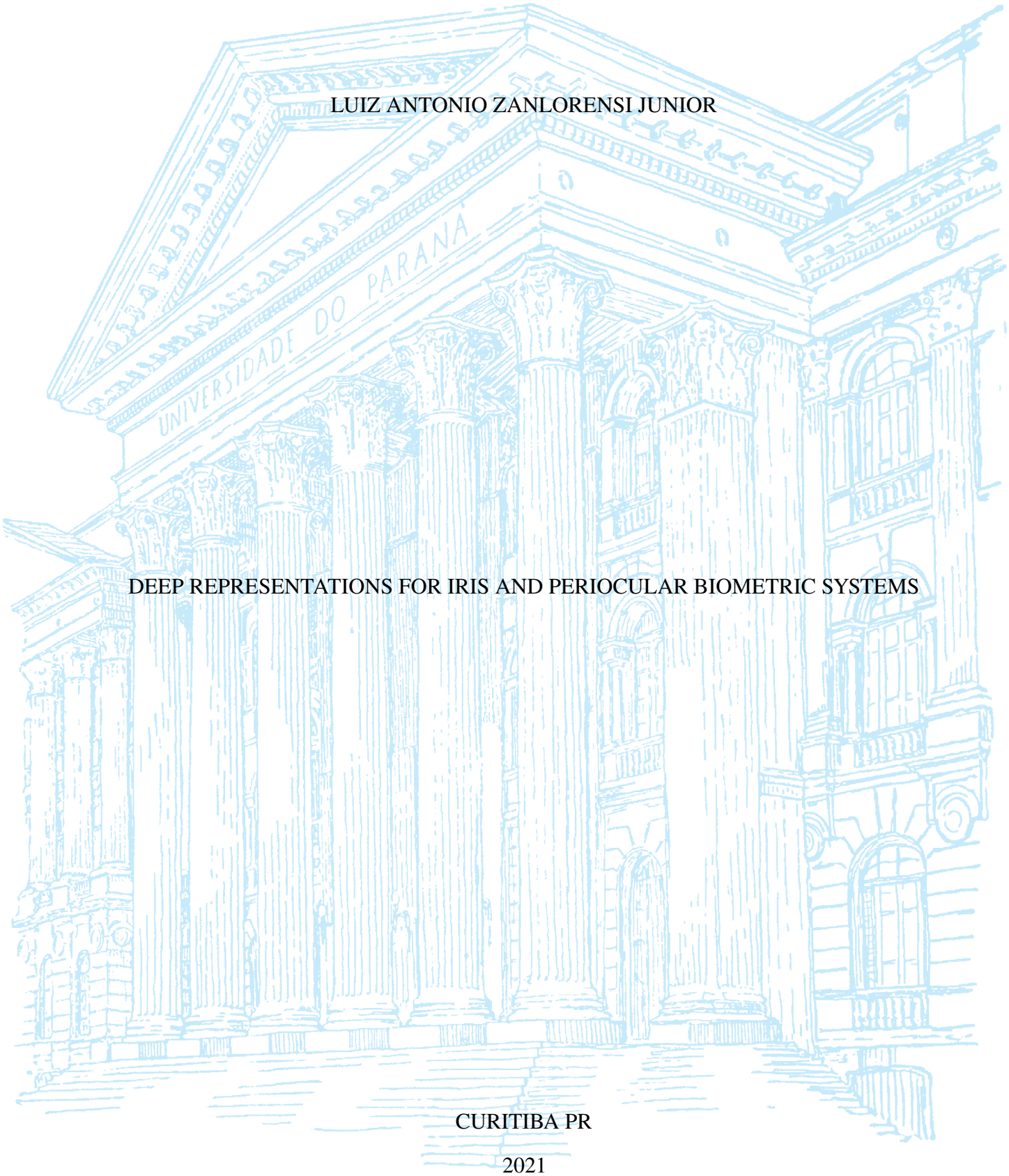
UNIVERSIDADE FEDERAL DO PARANÁ

LUIZ ANTONIO ZANLORENSI JUNIOR



DEEP REPRESENTATIONS FOR IRIS AND PERIOCULAR BIOMETRIC SYSTEMS

CURITIBA PR

2021

LUIZ ANTONIO ZANLORENSI JUNIOR

DEEP REPRESENTATIONS FOR IRIS AND PERIOCULAR BIOMETRIC SYSTEMS

Thesis presented as a partial requirement for the degree of Doctor in Computer Science in the Graduate Program in Informatics, Exact Sciences Sector, of the Federal University of Paraná, Brazil.

Field: Computer Science.

Advisor: David Menotti.

Co-advisor: Alceu S. Britto Jr. and Hugo Proença.

CURITIBA PR

2021

Catalogação na Fonte: Sistema de Bibliotecas, UFPR
Biblioteca de Ciência e Tecnologia

Z31d

Zanlorensi Junior, Luiz Antonio
Deep representations for iris and periocular biometric systems [recurso eletrônico] / Luiz Antonio Zanlorensi Junior. – Curitiba, 2021.

Tese - Universidade Federal do Paraná, Setor de Ciências Exatas, Programa de Pós-Graduação em Informática, 2021.

Orientador: David Menotti Gomes – Coorientador: Alceu S. Britto Junior - Coorientador: Hugo Proença

1. Identificação biométrica. 2. Reconhecimento de padrões. 3. Íris (Olhos). 4. Análise espectral. I. Universidade Federal do Paraná. II. Gomes, David Menotti. III. Britto Junior, Alceu S. IV. Proença, Hugo. V. Título.

CDD: 006.248

Bibliotecário: Elias Barbosa da Silva CRB-9/1894

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em INFORMÁTICA da Universidade Federal do Paraná foram convocados para realizar a arguição da tese de Doutorado de **LUIZ ANTONIO ZANLORENSI JUNIOR** intitulada: **DEEP REPRESENTATIONS FOR IRIS AND PERIOCULAR BIOMETRIC SYSTEMS**, sob orientação do Prof. Dr. DAVID MENOTTI GOMES, que após terem inquirido o aluno e realizada a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

A outorga do título de doutor está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

CURITIBA, 22 de Abril de 2021.

Assinatura Eletrônica
03/05/2021 09:52:30.0
DAVID MENOTTI GOMES
Presidente da Banca Examinadora

Assinatura Eletrônica
03/05/2021 10:23:19.0
EDUARDO TODT
Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

Assinatura Eletrônica
03/05/2021 09:33:46.0
LUIZ EDUARDO SOARES DE OLIVEIRA
Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

Assinatura Eletrônica
05/05/2021 10:48:13.0
RODRIGO MINETTO
Avaliador Externo (UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ)

Assinatura Eletrônica
05/05/2021 13:53:58.0
EDUARDO JOSÉ DA SILVA LUZ
Avaliador Externo (UNIVERSIDADE FEDERAL DE OURO PRETO)

To the memory of my father, Luiz Antonio Zanlorensi, and my grandfather, Antoninho Zanlorensi.

ACKNOWLEDGEMENTS

First, I would like to thank my entire family for the support and encouragement they have always given me, from the beginning of my graduation to the complete masters and doctorate period. To my mother, Josiane, and my sister Heloisa, for always being there for me and for all the strength they have always given me.

To my life partner, Alana, for being always present, continuously supporting me in all the necessary decisions. Your support and encouragement were essential during this journey.

I also thank the Federal University of Paraná, the National Council for Scientific and Technological Development (CNPq), and the Higher Education Personnel Improvement Coordination (CAPES) for the financial and infrastructure support that were extremely important and indispensable for the realization of this work.

In addition, I would like to thank my advisors who have become great friends along the way: David Menotti, Alceu Britto Jr., and Hugo Proença, for all their attention and help during the development of this research. Thank you not only for the example of extreme professionalism but also for all the extra-professional support. Here is also my thanks to all the professors who helped me during the development of this research and to the qualification and defense examining committee members for all the notes that were essential for the improvement of this thesis.

Thanks to all the friends and colleagues in the Vision, Robotics and Imaging (VRI) lab with whom I have spent most of my time over the past 5 years. For all the moments of collaboration, exchange of experiences and moments of relaxation during coffees.

Finally, I want to thank all the friends of SOCIA-LAB (University of Beira Interior), for all their support and help during my stay in Covilhã. These are friendships that I will carry, forever, throughout my life.

This thesis was partially prepared at IT - Instituto de Telecomunicações, Pattern and Image Analysis Group, Covilhã . The support of the University of Beira Interior, and IT: Instituto de Telecomunicações is acknowledged, in the scope of the UIDB//EEA/50008/2020 Project, funded by the Fundação para a Ciência e a Tecnologia (FCT).

"Never look down on anybody unless you're helping him up". (Jesse Jackson)

RESUMO

Traços oculares são amplamente utilizados em sistema biométricos devido a alta distinção e unicidade da íris, e a viabilidade de projetar métodos robustos de reconhecimento periocular em ambientes sem restrições. Sistemas biométricos que empregam imagens de íris obtidas no infravermelho próximo (NIR) e capturadas em ambientes controlados podem ser considerados uma tecnologia madura, provando serem eficazes em diferentes cenários. Um dos maiores desafios atuais da biometria ocular é a utilização de imagens obtidas no espectro visível em ambientes não controlados. O principal problema com essas imagens é que elas geralmente possuem vários ruídos causados por diversos fatores como desfoque, desfoque de movimento, baixo contraste, reflexo especular, ângulo do olhar, olho fora de ângulo e oclusão. Esses ruídos geralmente aumentam as variações intra e interclasses, degradando a acurácia dos sistemas biométricos oculares tanto para a íris quanto para as regiões perioculares. Com o recente avanço das técnicas de aprendizado profundo, várias abordagens aplicando Redes Neurais Convolucionais (CNN) para reconhecimento ocular vem sendo desenvolvidas. A principal vantagem das aplicações baseadas em deep learning é que, ao contrário da engenharia de características, existe um processo de aprendizagem de características. Dessa forma, estas aplicações podem produzir modelos de extração de características invariantes a algumas variações intra e interclasses, dependendo das amostras de imagens presentes no conjunto de treinamento. Considerando a necessidade de constante evolução dos métodos biométricos, nesta tese, exploramos e investigamos representações profundas para o reconhecimento da íris e da região periocular em diferentes cenários. Nossa hipótese principal é que é possível alcançar resultados a nível do estado da arte empregando técnicas de aprendizagem profunda em diferentes etapas de sistemas biométricos oculares baseados nos traços da região periocular e da íris. Para testar esta hipótese, investigamos, propomos e avaliamos diversas abordagens para sistemas biométricos de reconhecimento da íris e da região periocular, produzindo as seguintes contribuições: desenvolvendo uma abordagem para o reconhecimento da íris em ambientes sem restrições removendo as etapas de pré-processamento, projetando um único modelo capaz de aprender diretamente representações de imagens de íris e de regiões perioculares obtidas em espectros diferentes, propondo um método de normalização de atributos presentes em imagens perioculares para reduzir a variabilidade intraclasse causada por atributos não inerentes dos sujeitos, e utilizando soft-biometria no estágio de treinamento de um modelo multitarefa para melhorar a discriminabilidade da representação profunda da região periocular. Outra importante contribuição é a coleta de uma nova base de dados de imagens perioculares capturadas por dispositivos móveis em ambientes sem restrições. Até onde sabemos, essa base de dados é a que possui a maior quantidade de indivíduos presente na literatura e está públicamente disponível para a comunidade científica. Experimentos extensivos com as abordagens propostas utilizando bases de dados publicas demonstram que as técnicas de aprendizado profundo aplicadas ao reconhecimento ocular empregando os traços da íris e da região periocular podem alcançar resultados interessantes, mesmo em ambientes irrestritos e não controlados.

Palavras-chave: Biometria ocular, Reconhecimento de íris, Reconhecimento periocular, ambientes sem restrição, Reconhecimento em espectros cruzados.

ABSTRACT

Ocular traits in biometric systems are widely used because of the iris' high distinction and uniqueness, and the feasibility of designing robust periocular recognition methods in unconstrained environments. Biometric systems employing Near-infrared (NIR) iris images captured in controlled environments can be considered a mature technology, proving to be effective in different scenarios. One of the current greatest challenges in ocular biometrics is the use of images obtained at the visible spectrum (VIS) under uncontrolled environments. The main problem with these images is that they usually have several noises caused by factors such as blur, motion blur, low contrast, specular reflection, eye gaze, off-angle, and occlusion. These noises generally increase intra and inter-class variations, degrading the ocular biometric systems' accuracy for both iris and periocular regions. With the recent advancement of deep learning techniques, several approaches applying Convolutional Neural Networks (CNN) to ocular recognition have been designed. The main advantage of applications based on deep learning is that, unlike the handcrafted features, there is a process of representation learning. This process can produce feature extractor models invariant for some intra and inter-class variations, depending on the image samples present in the training set. Considering the need for the constant evolution of biometric methods, in this thesis, we explored and investigated deep representations for iris and periocular recognition in different scenarios. Our main hypothesis is that it is possible to achieve state-of-the-art results by employing deep learning techniques at different stages of ocular biometric systems based on periocular and iris traits. To support this hypothesis, we investigate, propose, and evaluate several approaches for iris and periocular recognition producing the following contributions: an approach for iris recognition in unconstrained environments without preprocessing steps, a single model to directly learn representations from cross-spectral images of iris and periocular regions, an attribute normalization method to reduce the intra-class variability present in periocular images caused by subjects' noninherent attributes, and the use of soft biometrics in the training stage of a multi-task model to improve the periocular deep representation's discriminability. Another important contribution is the release of a new periocular database captured by mobile devices in unconstrained environments. To the best of our knowledge, the collected database is the largest one in terms of the number of subjects publicly available to the research community. Extensive experiments with our proposed approaches using publicly ocular databases support that deep learning techniques applied to ocular recognition for both the iris and periocular traits can achieve impressive results even in unconstrained and uncontrolled environments.

Keywords: Ocular biometrics, Iris recognition, Periocular recognition, Unconstrained Environments, Cross-spectral recognition.

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

ANN	Artificial Neural Network
AUC	Area Under the Curve
CMC	Cumulative Match Characteristic
CNN	Convolutional Neural Network
EER	Equal Error Rate
FAR	False Acceptance Rate
FCN	Fully Convolutional Network
FMR	False Match Rate
FN	False Negative
FNMR	False Non-Match Rate
FP	False Positive
FPR	False Positive Rate
FRR	False Rejection Rate
FTA	Failure-to-acquire
FTE	Failure-to-enroll
GABOR	Gabor Spectral Decomposition
GAN	Generative Adversarial Network
GFAR	Generalized False Accept Rate
GFFR	Generalized False Reject Rate
HOG	Histogram of Oriented Gradients
LBP	Local Binary Patterns
LDA	Linear Discriminant Analysis
NIR	Near-infrared
PCA	Principal Component Analysis
PFB	Pairwise Filter Bank
PNN	Probabilistic Neural Network
RBFNN	Radial Basis Function Neural Network
RBM	Restricted Boltzmann Machines
ROC	Receiver Operating Characteristic
RR	Recognition Rate
SAFE	Symmetry Patterns
SGD	Stochastic Gradient Descent
SIFT	Scale-Invariant Feature Transform
SOM	Self-Organizing Map
SVM	Support Vector Machine

TN	True Negative
TP	True Positive
TPR	True Positive Rate
VIS	Visible

CONTENTS

1 INTRODUCTION

Several corporations and governments fund biometrics research due to various applications such as combating terrorism and the social networks, showing that this is a strategically important research area [30, 31]. A biometric system exploits pattern recognition techniques to extract distinctive information/signatures of a person [3]. Such signatures are stored and used to compare and determine the identity of a person sample within a population. As biometric systems require robustness against acquisition and/or preprocessing fails, as well as high accuracy, the challenges and the methodologies for identifying individuals are constantly evolving.

Methods that identify a person based on their physical or behavioral features are particularly important since such characteristics cannot be lost or forget, as may occur with passwords or identity cards [32]. In this context, the use of ocular information as a biometric trait is interesting regarding a noninvasive technology and also because the biomedical literature indicates that irises are one of the most distinct biometric sources [33]. Moreover, the periocular region can provide discriminative patterns even in noisy images when the iris recognition is difficult [27, 34, 35, 25, 24].

The most common ocular biometrics task is recognition, divided into verification (1:1 comparison) and identification (1:N comparisons). Furthermore, recognition can be performed in two distinct protocols called closed-world (subject-dependent) and open-world (subject-independent). In the closed world protocol, samples of an individual are present in the training and test set. On the other hand, there may be samples in the test set belonging to individuals that are not present in training set in the open-world protocol. The identification process generally is performed on the closed-world protocol (except the open-set scenario, which has imposters that are only in the test set, i.e., individuals who should not match any subject in the gallery set), while verification (authentication) can be performed in both, being the open-world most common protocol adopted in this setup. In additional to identification and verification, there are other tasks in ocular biometrics such as spoofing and liveness detection [36, 37], recognition of mislabeled left and right iris images [38], gender classification [39], iris and periocular region detection [40, 41, 42], iris and sclera segmentation [43, 44], and sensor model identification [45].

Nowadays, with the advancement of deep learning-based techniques, several methodologies applying this kind of frameworks have been developed for iris and periocular recognition [46, 23, 47, 48, 49, 50, 24, 25, 34, 51, 35]. The advancement of the ocular biometric systems can be observed by the recent contests that have been conducted to evaluate the evolution of the state-of-the-art methods for different applications, such as iris recognition in heterogeneous lighting conditions (NICE.I and NICE.II) [1, 52], iris recognition using mobile images (MICHE.I and MICHE.II) [15, 27], iris and periocular recognition in cross-spectral scenarios (Cross-Eyed 1 and 2) [18, 19], and periocular recognition using mobile images captured in different lighting conditions (VISOB 1 and 2) [14, 28]. Furthermore, several approaches using soft-biometrics as a second (auxiliary) biometric information have been explored to improve biometric tasks [53, 54, 55, 56, 17]. Generally, soft-biometric information does not have a high level of discrimination. However, it can be used to index databases or to enhance the recognition accuracy of primary biometric traits as the face, periocular region, and iris [54].

Considering the constant evolution of biometric methods as stated by the aforementioned competitions and ocular recognition methods, in this thesis, we investigated the following topics regarding iris and periocular recognition employing deep learning techniques:

- The impact of preprocessing on deep representations extracted from iris images.

- Iris and periocular deep representations for cross-spectral ocular biometrics.
- Periocular attribute normalization employing generative adversarial networks.
- Benchmark of several CNN architectures in a large periocular database in the unconstrained environment.
- Periocular multitask learning with soft-biometrics.

Classical biometric methods employing the iris traits usually apply preprocessing techniques such as iris detection, segmentation, and normalization [57]. Recent works on ocular biometrics stated that CNN models could automatically define the region of interest and extract discriminative representations of this region [34, 51]. Thus, in our first study, we analyze the impact of preprocessing steps on iris recognition. This investigation was carried out by an ablation study about the impact of the following preprocessing steps on deep representations of NIR and VIS iris images: iris detection, segmentation for noise removal and normalization using the rubber sheet model [57, 4, 30].

As we stated that it is possible to develop a robust iris biometric system by directly employing the iris bounding box region [23] as input of CNN models, we evaluated the use of this approach combined with the periocular region in a cross-spectral scenario. We performed extensive experiments on two publicly available cross-spectral ocular databases for both the closed and open-world protocols for this study. We also evaluated different weights to fuse the iris and the periocular verification scores, and the performance of the approach in terms of EER varying the depth of the layer from where representations were taken.

Regarding, only the periocular region, we observed that factors as eye-gaze and eyeglasses present in the images, generally increase the intra-class variability, degrading the performance of the biometric system [24]. To handle with these problems, we proposed an attribute normalization method employing a state-of-the-art Generative Adversarial Network (GAN) model for automatic image editing. This method is a prepossessing step to correct different attributes of a pair of images. As proof of concept, we considered the "eyeglasses" and "eye-gaze" factors, comparing the levels of performance of different recognition methods based on deep representations and hand-crafted features with/without using the proposed normalization strategy.

Finally, to investigate the scalability of periocular biometric systems developed with CNN models and soft-biometrics information to improve the recognition accuracy of these systems, we developed a new database (currently the largest one in terms of the number of subjects) containing images obtained by mobile devices in unconstrained scenarios.

1.1 MOTIVATION

Periocular recognition has been demonstrated to be an alternative when the iris trait is not available due to occlusions or low image resolution. However, the iris trait has a high uniqueness than the periocular region. Thus, the study of both traits for biometrics is essential to design and develop robust biometric systems. Furthermore, a recent study [58] stated that the use of masks (currently due to the Covid-19 pandemic) decreases significantly the verification performance of face biometric systems. In this sense, ocular recognition can be employed as an alternative since masks usually do not occlude the periocular region.

Machine learning techniques based on deep learning have achieved great popularity in the last years due to the literature results. Recently works reported results outperforming the state of the art in several problems, such as object detection [59, 60, 61, 62, 63], speech recognition [64, 65, 66, 67, 68], natural language processing [69, 70, 71, 72], medical research [73, 74], optical

character recognition [75, 76, 77, 78, 79], handwritten digit recognition [80, 81, 82, 83], and face recognition [84, 85, 86, 87]. In the field of ocular biometrics, the use of deep learning representation has been advocated both for the periocular [46, 48, 51, 88, 89] and iris regions [90, 91, 92, 93, 34, 29, 35, 47, 94], with interesting and promising results being reported.

As stated in previous works [95, 90], an often and open problem in ocular recognition is matching heterogeneous images captured at different resolutions, distances, and devices (cross-sensor and cross-spectral). It is difficult to design a robust handcrafted feature extractor to address the intra-class variations present in these scenarios regarding these problems. In this sense, several recent works demonstrate that deep representations report better results compared to handcrafted features in iris and periocular region recognition [90, 46, 51, 29].

Another recent advancement is the use of deep learning techniques for automatic facial attribute editing. Approaches based on Generative Adversarial net (GAN) [96] and Variational Autoencoder (VAE) [97] architectures reported promising results performing these tasks [98, 99, 100, 101, 26, 102, 103, 104, 105]. The models for face attributes editing can be divide based on their ability to manipulate a single [102, 103] or multiple attributes [98, 99, 100, 101, 26], such as eyeglasses, hair color, age, mustache, gender, beard, among others. Also, there are strategies for image attribute editing by transferring face attributes [100, 105, 104]. The concept of this task is to modify a face image based on attributes contained in another image, preserving the subject's identity. Regarding the intra-class variability present in the periocular trait caused by attributes in the image such as eyeglasses, eye-gaze, makeup, and contact lenses, we hypothesize that this family of frameworks for automatic image editing can be explored and employed to reduce this intra-class variability.

Finally, Soft biometrics, such as gender and age classification, using ocular traits are tasks that have gained attention in research in recent years [53, 54, 55, 56, 17]. It can be used as second biometric information to improve the accuracy of biometric systems [54]. Some works employing ocular traits (iris and periocular region) using VIS images for gender and age estimation/classification based on deep learning techniques achieved promising results [106, 55, 56, 107, 17].

1.2 PROBLEM DEFINITION

Ocular recognition using iris images captured at controlled NIR wavelength environments is a mature technology, proving to be effective in different scenarios [32, 108, 109, 52, 34, 94]. Currently, one of the greatest challenges in ocular biometrics is the use of images obtained in the *visible spectrum* (VIS) under uncontrolled environments [52, 27]. The main problem with these images is that they may have some noises caused by factors such as blur, motion blur, low contrast, specular reflection, eye gaze, off-angle, and occlusion. These noises generally increase intra and inter-class variations, degrading the ocular biometric systems' accuracy for both iris and periocular regions. Samples of some of these problems are shown in Fig. 1.1.

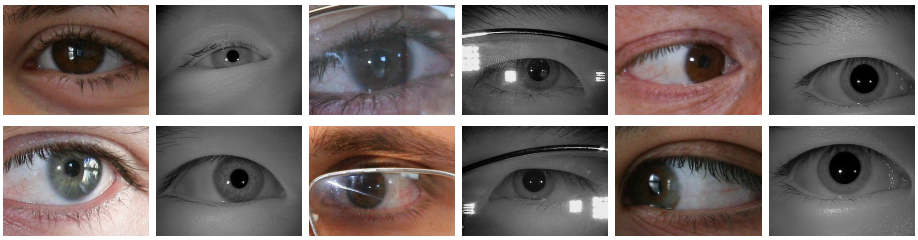


Figure 1.1: Some problems found in VIS (UBIRIS.V2 [1] database) and NIR (CASIA-THOUSAND [2] database) iris images. (a) VIS and (b) NIR specular reflection, (c) VIS and (d) NIR noise caused by glasses, (e) VIS eye-gaze and (f) NIR pupil dilatation.

Regarding the problems caused by noises present in the images, classical ocular biometric systems employ some preprocessing methods to correct or reduce these factors' impact. As previously described, an iris biometric system can be decomposed in some steps. These steps usually consist of image preprocessing, feature extraction (representation), and classification (e.g., matching). On preprocessing, commonly, three stages are performed. The first process consists of iris region detection and/or iris and pupil delineation. Then, a segmentation approach is applied for noise removal. With delineated and segmented image, the last preprocess is realized to normalize the effect of scale and pupil dilation/constriction. Finally, the feature extraction/representation and classification (matching or identification) are performed using the preprocessed images. Considering that preprocessing such as iris detection and segmentation for noise removal are still a complex and an open problem, some errors may occur in these stages. As shown in Fig. 1.2, these errors are propagated to the next steps, decreasing the system performance.

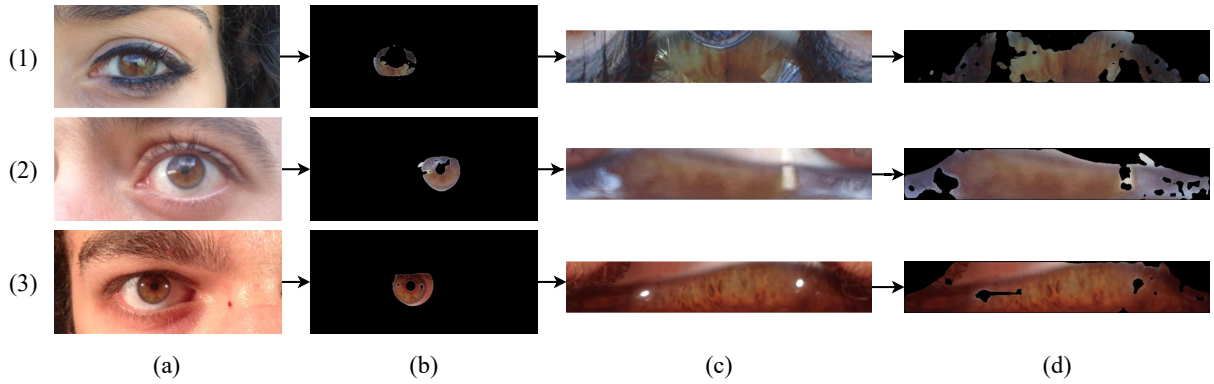


Figure 1.2: Errors in iris preprocessing stages. (a) original image, (b) delineated and segmented iris, (c) normalized iris, and (d) segmented and normalized iris image. (1) and (2) error in pupil boundary detection caused by reflection, (3) error in iris boundary detection.

A recent challenge in ocular recognition is the application of biometric systems in a cross-spectral scenario/setting. The term cross-spectral refers to matching features extracted from images captured at different wavelengths, usually VIS images against NIR. Based on the configuration of the experiment, the feature extraction training step can be performed using images obtained at only one wavelength (VIS or NIR) or both (VIS and NIR). The problem in this scenario is that the features present in NIR images are not always the same as that extracted in VIS images. Several recent approaches have been developed [18, 19, 20, 29, 24], showing that ocular biometric systems based on both iris and periocular traits still need improvements for the cross-spectral scenario.

In periocular recognition, another problem that still needs attention is the subject's non-inherent attributes, such as eyeglasses, eye gaze, and makeup present in the images captured under unconstrained environments. These problems generate high intra-class variability, degrading the uniqueness of the features extracted from the biometric trait.

Finally, the evaluation of the proposed ocular biometric methods' scalability remains a problem since the available databases do not have samples from a large number of subjects. The term scalability refers to a biometric system's ability to maintain efficiency (accuracy) even when applied to databases with a large number of images and subjects. The largest NIR iris database available in the literature in terms of the number of subjects is CASIA-IrisV4-Thousand [2], which has 20,000 images taken in a controlled environment from 1,000 subjects. In an uncontrolled environment and with VIS ocular images, the largest database is VISOB [14], which is composed of 158,136 images from 550 subjects. Although several proposed methodologies achieve high decidability index in these databases [93, 34, 94, 14, 110, 111, 112], indicating that these approaches have impressive and high separation of the intra- and inter-class comparison distribution, can we state that these methodologies are scalable? In this sense, it is necessary to research new methods and new databases with a larger number of images/subjects to evaluate the scalability of existing approaches in the literature.

1.3 CHALLENGES

Considering the aforementioned problems, the main challenge is how to improve the performance in terms of accuracy of ocular biometric systems employing deep learning techniques. In this sense, to design robust biometric ocular systems using irises and the periocular region separately and also to fuse these traits, we can present the following specific challenges:

- The preprocessing steps in iris recognition methods as segmentation for noise removal and normalization remain a complex problem, usually leading the biometric system to mismatched, specifically in unconstrained scenarios. Thus, a recent challenge is how to develop a biometric approach employing a deep learning technique using the iris trait that does not require these preprocessing steps.
- Design a methodology approach that can learn specific representations from different ocular biometrics sources such as NIR and VIS images obtained in constrained and unconstrained environments from cross-spectral scenarios.
- Regarding the intra-class variation problem present in the periocular recognition under unconstrained scenarios, a challenge is how we can employ deep learning frameworks to handle and reduce this problem.
- Since the available ocular databases do not have samples from a larger number of subjects, and also taking into account that several recent ocular approaches have been developed using mobile images, it is important to create/collect a new database considering theses features.
- Finally, a recent challenge is how to use soft biometrics as gender and age to improve the accuracy of ocular biometric systems.

1.4 HYPOTHESES

The main hypothesis of this work is that **it is possible to achieve state-of-the-art results by employing deep learning techniques at different stages of ocular biometric systems based**

on periocular and iris traits. For such aim, we have to develop ocular biometric systems based on deep learning frameworks and compare the proposed approaches with state-of-the-art methods. Summarizing the hypothesis, we aim to answer the following research questions:

- Can we eliminate the preprocessing stages, i.e., directly use an iris squared/rectangular bounding box to extract iris deep representations?
- It is possible to use a single model to directly learn representations from iris and periocular region captured under different wavelengths (cross-spectral representation)?
- Can we employ GAN architectures for automatic image editing to normalize different attributes present in periocular images reducing the intra-class variability?
- Finally, can we use soft biometrics information to improve the accuracy of ocular biometric systems?

To address all these research questions, we have to study and investigate recent ocular biometrics approaches and also recent works employing deep learning techniques for several tasks such as segmentation, recognition, and automatic image editing. All the investigations and experiments are described in the next chapters.

1.5 OBJECTIVES

This thesis's main objective is to investigate and develop ocular biometric systems employing recent deep learning techniques from iris and periocular traits. For this purpose, it was necessary to design and evaluate ocular biometric methods using the iris and periocular region in different scenarios. The developed methods and their specific objectives can be organized into the following topics:

1. Iris recognition in unconstrained environments without preprocessing as segmentation and normalization: this approach aims to create an iris biometric system removing the preprocessing steps since it usually returns errors degrading the feature matching process. To design an iris biometric system addressing these features, we have to perform an ablation study of the impact of preprocessing stages as segmentation for noise removal and normalization using iris images captured under unconstrained environments.

2. Iris and periocular recognition on the cross-spectral scenario: the objective is to investigate and develop an ocular biometric system that can directly learn representations from ocular images captured at VIS and NIR wavelengths.

3. Periocular attribute normalization to reduce the intra-class variability: the objective is to develop an attribute normalization process employing recent GAN models to automatic remove or correct eyeglasses and eye-gaze factors present in the periocular images. Performing this normalization process, the intra-class variation will be reduced, improving periocular recognition systems' accuracy.

4. Benchmark of state-of-the-art CNN architectures employing them for periocular recognition in unconstrained environments: the objective is to perform a benchmark for the identification and verification tasks extracting deep representations for periocular images captured by mobile devices in the unconstrained environment;

5. Soft biometrics to improve the accuracy of periocular recognition methods: the objective is to employ soft biometric information such as age and gender in the deep representation extractor's training stage to improve the learning process.

Another important objective is creating a new periocular database containing periocular images collected by the participants using their own smartphone. This database will be used to investigated soft biometrics and the scalability of the existing ocular biometrics approaches.

To accomplish the main objective, some secondary or specific objectives are required, as follows:

- To study and compare different CNN models to extract deep representations from the iris and the periocular region.
- To evaluate shallow and deep ocular representations from different CNN architectures.
- To study approaches to train CNN models, such as transfer learning and fine-tuning.
- To investigate recent approaches for automatic image editing keeping the discriminative features in the image.
- To create a mobile application enabling the subject himself to capture his periocular images.
- To study which soft biometrics can be used to improve the performance of the ocular biometric systems in terms of accuracy.

1.6 CONTRIBUTIONS

Addressing the aforementioned objectives, some methodologies and approaches were investigated and developed, generating the following contributions:

- A survey of ocular databases and the most challenging biometric problems employing the iris and periocular traits. We research and explore most of the ocular databases found in literature and their applications, as well as competitions on ocular biometric recognition and the methodologies that reported the best results to overview the recent and challenging problems. The produced survey can provide a general overview of the challenges in ocular recognition over the years, the databases used in the literature, and some future directions in this research field.
- An approach for iris recognition not requiring preprocessing steps as segmentation and normalization, achieving state-of-the-art results.
- Extensive experiments and evaluation of a method employing CNN models to directly learn representations from iris and periocular and fusing these traits to achieve state-of-the-art results on the cross-spectral ocular recognition scenario.
- A new and original attribute normalization process employing recent GAN architectures to reduce the intra-class variability in periocular images. The proposed normalization method was validated for periocular recognition approaches based on hand-crafted features and deep representations, improving both techniques' accuracy.

- A new periocular database (UFPR-Periocular) containing 33,660 images from 1,122 subjects. This database is currently the largest one in the literature in terms of the number of subjects. We also manually annotated each image's eye corners and store information about the subjects as gender and age. The UFPR-Periocular, manual annotations, and information are available to the research community and can be employed to study new ocular biometric methods for realistic unconstrained scenarios.

1.7 LIST OF PUBLICATIONS

This document generated the following original produced and published works:

- **Ocular Recognition Databases and Competitions: A Survey**; L. A. Zanlorensi, R. Laroca, E. Luz, A. S. Britto Jr., L. S. Oliveira, D. Menotti. *Artificial Intelligence Review*; 2021 [13].
- **The Impact of Preprocessing on Deep Representations for Iris Recognition on Unconstrained Environments**; L. A. Zanlorensi, E. Luz, R. Laroca, A. S. Britto Jr., L. S. Oliveira, D. Menotti;*31st Conference on Graphics, Patterns and Images (SIBGRAPI)*; 2018 [23].
- **Deep Representations for Cross-spectral Ocular Biometrics**; L. A. Zanlorensi, D. R. Lucio, A. S. Britto Jr., H. Proença, D. Menotti; *IET Biometrics*, 9(2):68–77; 2020 [24].
- **Unconstrained Periocular Recognition: Using Generative Deep Learning Frameworks for Attribute Normalization**; L. A. Zanlorensi, H. Proença, D. Menotti; *IEEE International Conference on Image Processing (ICIP);* 2020 [25].
- **UFPR-Periocular: A Periocular Dataset Collected by Mobile Devices in Unconstrained Scenarios**; L. A. Zanlorensi, R. Laroca, D. R. Lucio, L. R. Santos, A. S. Britto Jr., D. Menotti; [Under Review with required major changes in Journal Qualis A1]; 2020 [17].

1.8 DOCUMENT ORGANIZATION

This work is further organized in 6 chapters. Chapter 2 contains the theoretical foundation about iris and periocular biometrics, Deep learning, and Convolutional Neural Networks. In Chapter 3, the literature review is described. First, we discuss some surveys on ocular recognition. The next subsection details ocular databases and their applications. To illustrate the state-of-the-art challenges, we describe and discuss the major recent competitions on ocular recognition and the approaches that have performed the best results. The last subsection of this chapter presents works applying deep learning methodologies to several ocular biometric tasks. The proposed methodologies are detailed in Chapter 4. The results are described and discussed in Chapter 5. Finally, the conclusion is given in Chapter 6.

2 THEORETICAL FOUNDATION

This Chapter introduces biometric systems based on ocular traits and deep learning fundamentals. We describe a mature biometric method for iris recognition and explain when periocular recognition can be employed. Finally, we detail different CNN architectures and models that have been explored to design ocular biometric systems.

2.1 BIOMETRICS

Several human traits can be employed for diverse biometric applications. The design of a biometric system using a particular biometric trait depends on a variety of issues besides its performance [3]. The suitability of a physical or a behavioral trait to be used in a biometric application can be determined by the following factors [113, 3]:

1. **Universality**: Every individual should possess the trait.
2. **Uniqueness**: The trait should be different across individuals from the population.
3. **Permanence**: The biometric trait should be invariant over a period of time.
4. **Measurability**: It should be possible to acquire the biometric trait using suitable devices in a non-inconvenient way. Furthermore, it should be possible to extract representative features from the acquired data.
5. **Performance**: The recognition accuracy and the resources required to achieve that accuracy should meet the constraints imposed by the application.
6. **Acceptability**: Individuals that will utilize the application should be willing to provide their biometric trait to the system.
7. **Circumvention**: This refers to the ease with which a biometric trait of an individual can be imitated/simulated.

No single biometric is expected to meet all the requirements imposed by all applications effectively. The relevance of a specific trait to a biometric system depends on the nature and requirements of that application, as well as on the properties of the biometric characteristics [3]. Unlike a password-based system where a perfect match is required between two values, a biometric system seldom encounters two samples of the same individual having the same feature set. This is due to imperfect sensing conditions caused by noises, sensor malfunctions, lighting problems, occlusions, and reflections. When two sets of feature biometric match perfectly, there is a high probability that one set comes from an attack on the system [3].

Biometrics systems can operate at least in two main modes: verification(1:1 comparison) and identification(1:N comparison) as shown in Figure 2.1.

The Verification task refers to the problem of verifying whether an individual is whom he claims to be. If two samples match sufficiently, the identity is verified; otherwise, it is rejected. In this way, verification can result in four possibilities [32]:

- **True accept**: the system accepts an identity claim, and the claim is true.

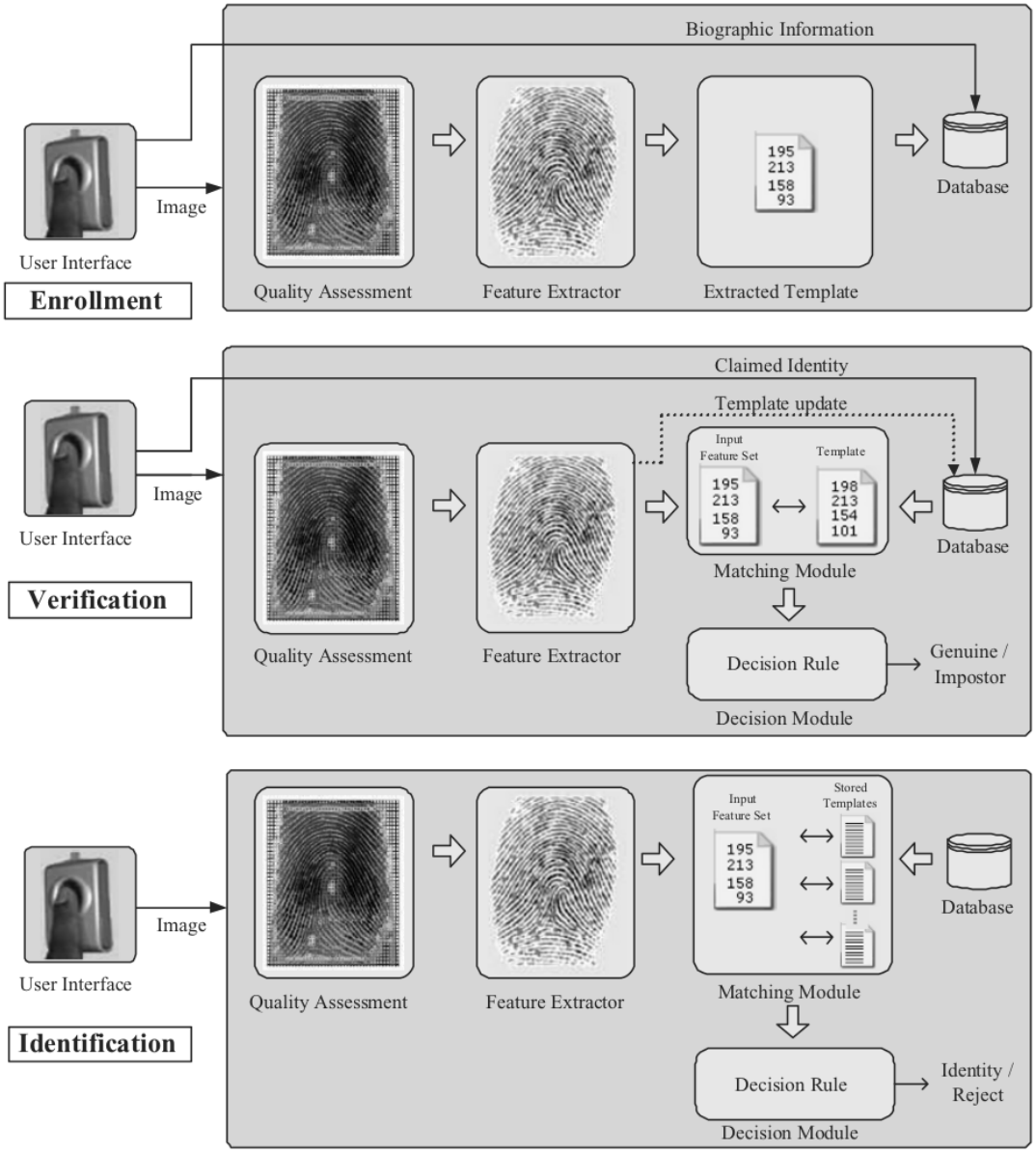


Figure 2.1: Biometric system stages. In enrollment, the features are extracted and stored in the gallery (database). The matching (verification or identification) is performed with the new input data features. Extracted from [3].

- **False accept**: the system accepts an identity claim, but the claim if false.
- **True reject**: the system rejects an identity claim, and the claim is false.
- **False reject**: the system rejects an identity claim, but the claim is true.

The two errors that can occur are false acceptance, measured by False Acceptance Rate (FAR), and false rejection, measured by False Rejection Rate (FRR). Regarding the supra-cited possibilities, the performance of biometric systems operating in a verification task is usually measured by the EER, which is computed from the ROC curve. This curve plots the TPR by the FPR, or alternatively, the FRR by the FAR. Then, the EER is the value where TPR and FPR are equals. Another metric that can be extracted from the ROC curve is the AUC, which informs

the quality of the predictions (verification matching) based on different thresholds. The EER and AUC metrics are show in Figure 2.2.

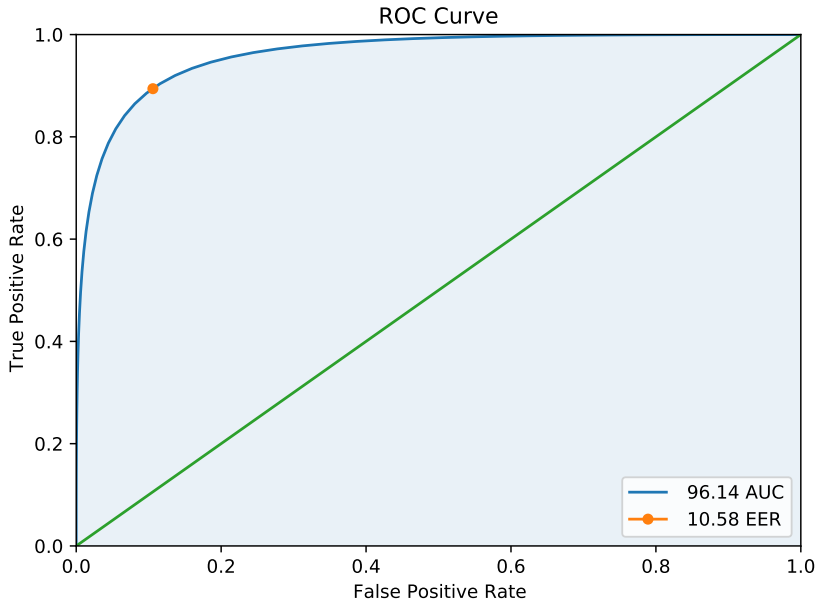


Figure 2.2: Receiver Operating Characteristic (ROC) curve. The EER is the value where TPR equals the FPR. The AUC measures the entire area underneath the ROC curve.

The decidability index d' [57] can be employed to compute the dissimilarity between samples. The metric or index d' measures how well separated are the two types of distributions (*genuines* and *impostors*), in the sense that recognition errors correspond to the regions where both distributions overlap:

$$d' = \frac{|\mu_E - \mu_I|}{\sqrt{\frac{1}{2}(\sigma_E^2 + \sigma_I^2)}}, \tag{2.1}$$

where the means and standard deviations of the genuine and impostor distributions are given by μ_I, μ_E, σ_I, and σ_E, respectively. Whereas the index d' can be related to the feature vector discrimination ability of an approach, the EER metric measures a biometric system's real performance.

For the identification task, the problem is to establish an individual sample's identity within a known database. Generally, the known database is called gallery, and the sample that will be classified is called probe. The probe sample is matched against all samples in the gallery, and the closest match is considered the individual's identity. Similar to verification, the identification results in four possibilities ([32]):

- **True Positive (TP)**: the system says that an unknown sample matches a particular person in the gallery and the match is correct.
- **False Positive (FP)**: the system says that an unknown sample matches a particular person in the gallery and the match is not correct.
- **True Negative (TN)**: the system says that the sample does not match any of the entries in the gallery, and the sample in fact does not.

- **False Negative (FN)**: the system says that the sample does not match any of the entries in the gallery, but the sample in fact does belong to someone in the galery.

The performance in an identification system is usually evaluated in a Cumulative Match Characteristic (CMC) curve. This curve plots the percent correctly recognized against the cumulative rank considered as a correct match. For example, for a cumulative rank of 3, if the correct individual is among the first 3 closest combinations, it is classified as correct. Usually, the rank value 1 of a CMC curve is highlighted to evaluate the identification performance.

Verification is usually used for positive recognition, where the goal is to prevent multiple people from using the same identity. Whereas identification is a critical component in negative recognition where the goal is to prevent a single person from using multiple identities [3].

2.1.1 Ocular Recognition

Ocular biometric systems can be designed employing information from the iris, periocular region, or both traits. The iris trait comprises the region between the sclera and pupil. The term periocular is associated with the region around the eye, composed of eyebrows, eyelashes, and eyelids [114, 115, 116], as illustrated in Fig. 2.3.

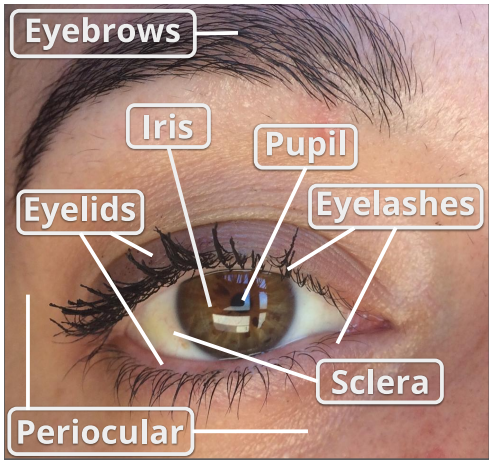


Figure 2.3: Ocular components.

As described by Daugman [4], to capture the rich details of iris patterns, an iris image should have a minimum of 70 pixels in iris radius. Being typical of a resolved iris radius of 80 – 130 pixels. Generally, the pupil center is nasal and inferior to the iris center and its radius can range from 0.1 to 0.8 of the iris radius [4].

The iris recognition method proposed by Daugman [57, 4, 30, 117], can be considered a mature and consolidated approach, being employed in several works as the baseline. However, its performance is usually better using NIR images obtained in a controlled environment. This approach consists of four stages: iris and pupil localization (segmentation), iris normalization, feature extraction, and matching. Each one of these steps is detailed below.

The first process (segmentation) is performed by integrodifferential operators [4], which consist of circular edge detectors used to locate the limbal and pupil boundaries of the iris [118]. These operators define the pupillary circle parameters separately from those of the iris due to the non-concentricity of the pupil and iris. Considering the abrupt intensity transition between the iris and sclera, the search for the iris boundary sets the smoothing function for a coarse scale of analysis [118]. This search is exhaustive across the image. With the iris delimited, the same process is performed for the pupil search but only looking at the detected iris region.

In the next stage, the normalization process is performed using the iris boundaries parameters. The homogeneous rubber sheet model (Fig. 2.4) assigns to each point on the iris a pair of real coordinates $((r, \theta))$ where r is on the unit interval $[0, 1]$ and θ is angle$[0, 2\pi]$ [4]. This remapping can be represented as

$$I(x(r, \theta), y(r, \theta)) \rightarrow I(r, \theta) \tag{2.2}$$

Also, $x(r, \theta)$ and $y(r, \theta)$ are defined as

$$x(r, \theta) = (1 - r)x_p(\theta) + rx_s(\theta) \tag{2.3}$$

$$y(r, \theta) = (1 - r)y_p(\theta) + ry_s(\theta) \tag{2.4}$$

where $(x_p(\theta), y_p(\theta))$ and $(x_s(\theta), y_s(\theta))$ represent the points between inner and outer boundary of the iris, respectively.

Figure 2.4: Rubber sheet model normalization proposed by Daugman [4].

The features are extracted from the normalized images applying 2-D Gabor wavelets. Considering that amplitude information is not very discriminating, and it depends upon extraneous factors (contrast, illumination), only phase information is used for recognizing irises [4]. With this information, a binary iris code is created. An iris mask is also generated, indicating which area will be used (iris) and discarded (noise/occlusion) in the matching step.

At last, the dissimilarity(matching) between two irises is measured computing the fractional Hamming Distance, whose two-phase code bit vectors are denoted $codeA, codeB$ and mask bit vectors are denoted $maskA, maskB$:

$$HD = \frac{\|(codeA \bigotimes codeB) \bigcap maskA \bigcap maskB\|}{\|maskA \bigcap maskB\|} \tag{2.5}$$

where $\bigotimes$ represents the XOR operator and $\bigcap$ represents the AND operator, and irises from the same class should produce $HD = 0$.

Considering that Daugman's approach is one of the first proposals for iris recognition and one of the most widespread, several current methods based on handcrafted and deep representations use segmented and normalized iris images.

Iris recognition under controlled environments at NIR demonstrates impressive results, and as reported in several works [32, 108, 109, 34, 94] can be considered a mature technology. However, iris trait in uncontrolled environments and images captured at VIS wavelength still is one of the greatest challenges in ocular biometrics [52, 14]. This kind of image usually presents noise caused by illumination, occlusion, reflection, and motion blur. Therefore, to improve the biometric system's performance in these scenarios, recent approaches have used information extracted only from the periocular region [16, 51, 46] or fusing them with iris features [119, 120, 121, 122]. Usually, the periocular region is used when there are poor quality in the iris region, commonly in VIS images or part of the face is occluded (in face

images) [114, 46]. In the literature regarding the periocular region, some works kept the iris and sclera regions [46, 52, 27] and others that removed them [18, 19, 51].

2.2 CONVOLUTIONAL NEURAL NETWORK (CNN)

As described by Lecun et al. [123], conventional machine-learning techniques were limited to process natural data in their raw form. Thus, pattern recognition or machine learning system requires domain expertise to design a feature extractor that transformed the raw data into representations or feature vectors used by the learning system to detect or classify patterns.

Deep learning methods learn multiple level representations, obtained by composing simple but non-linear modules that each transform the representation at one level into a representation at a higher, slightly more abstract level [123]. Deep learning somehow seeks to imitate the human brain. There are several reasons to believe that the human visual system contains multilayer generative models [124, 125] in which top-down connections can be used to generate low-level sample features from high-level representations and that bottom-up connections can be used to infer high-level representations that would have generated an observed set of low-level features.

Its important to emphasize that the deep learning representation generates generative feature models, whereas a conventional classifier, e.g., Support Vector Machine (SVM) [126, 127] is a discriminative model/classifier. Thus, representations obtained by deep learning can still be classified (inferred) by another generative model, such as a multilayer neural network or even an SVM.

One particular model of the deep network generalized much better than networks with full connectivity between adjacent layers. This model was the CNN [128, 129]. It achieved many practical successes when neural networks were out of favor, and it has recently been widely adopted by the computer-vision community [123].

In general, the CNNs is composed of multiple layers, each of which performs a filtering process through a convolution, activation, pooling, and normalization. There are four key ideas behind CNN models that take advantage of natural signals' properties: local connections, shared weights, pooling, and the use of many layers [123].

The architecture of a typical CNN model presents a series of stages, as can be seen in Fig. 2.5. The first few stages are composed of two types of layers: convolutional layers and pooling layers. Units in a convolutional layer are organized in feature maps. Each unit is connected to local patches in the previous layer's feature maps through a set of weights called a filter bank. The result of this locally weighted sum is then passed through a non-linearity function. All units in a feature map share the same filter bank. Different feature maps in a layer use different filter banks.

The convolution layer (filter bank) extracts features through the input sample's convolution operation with a kernel and detects local conjunctions of features from the previous layer. The activation layer plays an important role in the network information flow, improving the robustness of the features, rectifying the convoluted sample's output, and discarding less important information. Pooling is an operation that aims at bringing translational invariance to the features, employing operations, as a maximum or average, of certain regions of the sample, merging semantically similar features into one. Finally, the normalization layer promotes competition among filters, forcing the use of filters with the best response, according to a criterion [130, 123].

As detailed in [5], the convolutional layer consists of a set of learnable kernels or filters, which extract local features from the input and calculate feature maps. Each feature map is generated by sliding a filter over the input and computing the dot product. Then, a non-linear

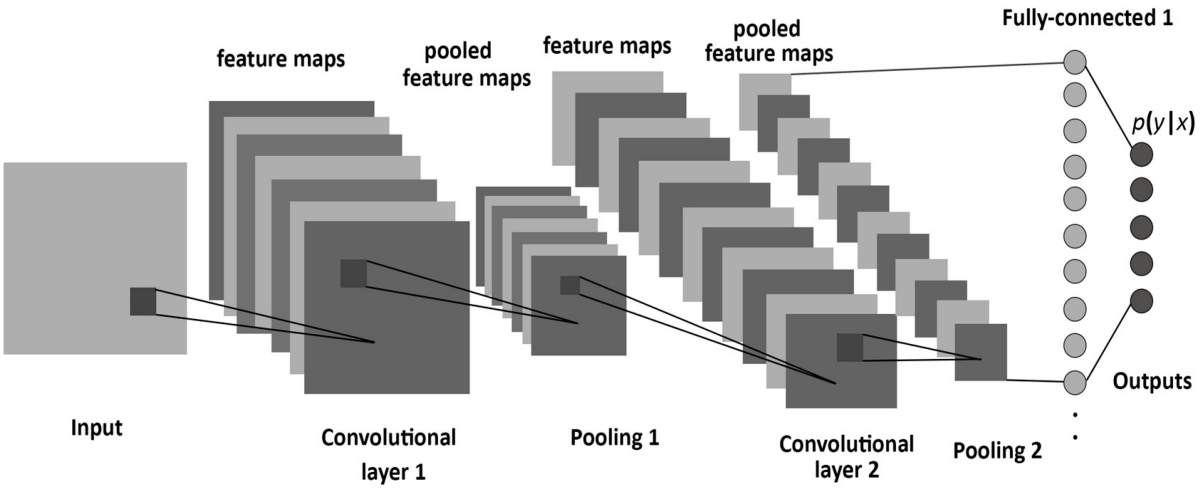


Figure 2.5: Generic structure of a CNN, consisting of convolutional, pooling, and fully-connected layers. Extracted from [5].

activation function is applied to introduce non-linearity into the model. All units share the same weights (filters) among each feature map. The advantage of sharing weights is the reduced number of parameters and the ability to detect the same feature, regardless of its location in the inputs.

There are several nonlinear activation functions, such as Logistic or Sigmoid (Eq. 2.6), TanH (Eq. 2.7), ArcTan (Eq. 2.8), ReLU (Eq. 2.9), among others. However, ReLU and some modifications of it are more used because they make the training faster than others [131, 132].

$$f(x) = \sigma x = \frac{1}{1 + e^{-x}} \tag{2.6}$$

$$f(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \tag{2.7}$$

$$f(x) = tan^{-1}(x) \tag{2.8}$$

$$f(x) = \begin{cases} 0, & \text{if } x < 0 \\ x, & \text{if } x \geq 0 \end{cases} \tag{2.9}$$

The size of the output feature map is based on the filter size and stride [5]. Thus, convolving the input image with a size of $(I \times I)$ over a filter with a size of $(F \times F)$ and a stride of (S), the output size of $(W \times W)$ is given by:

$$W = \left\lfloor \frac{H - F}{S} \right\rfloor + 1 \tag{2.10}$$

Usually, after one or a few convolutional layers, there is a pooling or down-sampling layer, which reduces the previous feature maps' resolution. Layers of this type split the inputs into disjoint regions with a size of $(R \times R)$ to produce one output with max or average pixel

values from each region [5]. If a given input with a size of $(W \times W)$ is fed to the pooling layer, then the output size will be obtained by:

$$P = \left\lfloor \frac{W}{R} \right\rfloor \tag{2.11}$$

At last, the top layers of CNNs models are one or more fully-connected layers, also called in some cases dense layers, similar to a feed-forward neural network, which aims to extract the global features of the inputs. The last layer has a softmax function (classifier), which reports the posterior probability of each class [5].

According to Lecun et al. [123] ,there have been numerous applications of CNN models going back to the early 1990s, starting with time-delay neural networks for speech recognition [133] and document reading [129]. More recently, works report results outperforming state of the art in several problems, such as speech recognition [64, 65, 66], natural language processing [69, 70], face recognition [84, 85], among others. With greater effort in generic object recognition, especially after the creation of the Imagenet database [131], which has more than 14 million images, several CNN architectures were created for multi-class classification as VGG-16 and VGG-19 [134], ResNet-50 [6], Inception and Inception-ResNet [7], Xception [135], DenseNet [136], NASNet [137], MobileNetV2 [138], among others. Furthermore, a recently interesting topic regarding deep learning is Neural Architecture Research [139, 140, 141, 142], which aims to design convolutional network architectures automatically. In the next subsections, we describe the most applied CNN architectures and models in ocular recognition systems.

2.2.1 Multi-class Classification

Multi-class classification is the task of classifying instances into three or more classes, where each sample must have a single unique class/label. Several techniques [143, 144, 145] have been proposed combining multiple binary classifiers to solve multi-class classification problems. Deep learning-based approaches usually address this problem through CNN models with softmax cross-entropy loss. In summary, these models' architecture has several convolutional, pooling, activation, and fully-connected layers, as shown in Fig. 2.6.

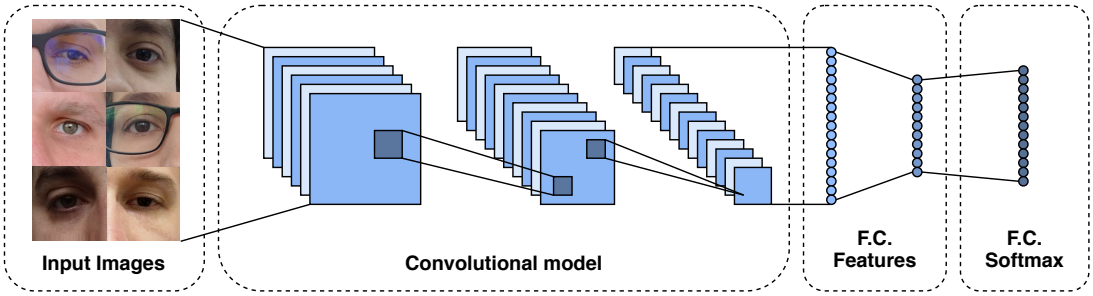


Figure 2.6: Multi-class classification CNN architecture.

In the training stage, a batch of images and their labels feed these models. The model extracts the image features through convolutional, pooling, and fully connected (dense) layers. The last layer is composed of a fully connected layer using the softmax cross-entropy as a loss function. Note that the model exactly as shown in Fig. 2.6 can be only applied in the closed-world protocol. However, its also possible to use a similar model for open-world protocol by using the last layer (softmax) only in the training stage and removing it to use the remaining architecture as a feature extractor. Below we describe the main characteristics of each model.

The VGG model, proposed by Simonyan and Zisserman [134], consists of a CNN using small convolution filters (3×3) with a fixed stride of 1 pixel. The spatial polling is computed by 5 max-pooling layers over a 2×2 pixel window. Two models were proposed varying the number of convolutional layers: VGG16 and VGG19. Both models have two fully connected layers at the top with 4096 channels each – these architectures achieved the first and second places in the localization and classification tracks on the ImageNet Challenge 2014. The authors also stated that it is possible to improve prior-art configurations by increasing the models' depth. Parkhi et al. [84] applied these models (called VGG16-Face) on the face recognition problem, showing that a deep CNN with a simpler network architecture can achieve results comparable to the state of the art. Furthermore, recent approaches for ocular (iris/periocular) biometrics employing VGG models have demonstrated the ability to produce discriminant features [23, 48, 46, 29, 146, 24, 147].

The Residual Network (ResNet) was introduced by He et al. [6] and applied to biometrics for face recognition [85], iris recognition [23, 148, 24, 29, 146] and periocular recognition [24, 49, 147, 149]. The authors addressed the degradation (vanishing gradient) problem caused by deeper network architectures proposing a deep residual learning framework. They added shortcut connections between residual blocks to insert residual information, as shown in Fig. 2.7. These residual blocks are composed of a weighted layer followed by batch normalization, an activation function, another weighted layer, and batch normalization. Let $F(x)$ be a residual block, and x the input of this block (identity map), the residual information consists of adding x to $F(x)$, i.e., $F(x) + x$, and using it as input to the next residual block.

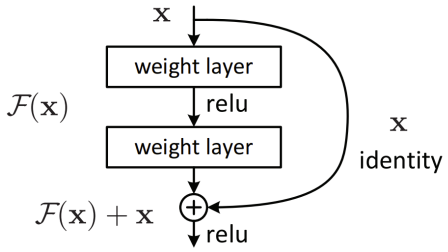


Figure 2.7: Residual building block. Extracted from [6]

Different architectures were proposed and evaluated, varying the models' depth: ResNet50, ResNet101, and ResNet152. These models achieved promising results on the ImageNet database [131]. In [150], He et al. proposed the ResNetV2 by changing the residual block by adding a pre-activation into it. Empirical experiments showed that the proposed method improved the network generalization ability, reporting better results than ResNetV1 on ImageNet. Besides the depth, one of the main differences of the ResNet models compared to the VGG models is the insertion of a global average pooling layer instead of a fully connected layer to the top of the network. The VGG models have fully connected layers at the top of the architecture to classify the feature maps generated by the convolutional layers. A common problem in this strategy is the overfitting of the fully connected layers, reducing the generalization ability of the entire model [151]. Regarding this problem, Hinton et al. [152] proposed the dropout technique, which works as a regularizer that randomly removes some activations from the fully connected layers. The ResNet models employ the global average polling at the top of the models. The global average polling strategy, proposed by Lin et al. [151] consists of generating one feature map for each corresponding category of the classification task in the last convolutional layer. Then this feature map is directly fed into the softmax layer. The main advantage of the global average polling over the fully connected layer is that the generated feature maps can be interpreted as

categories confident maps, and no parameter is needed to be optimized. Also, the global average polling is more robust to the spatial translations since it sums out the spatial information. [151]

The InceptionResNet model [7] combines the residual connections [6] and the inception architecture [7]. The first inception model [153], known as GoogLeNet, introduced the Inception module aiming to increase the network depth while keeping a relatively low computational cost. The main idea of inception is to approximate a sparse CNN with a normal dense construction. The inception module consists of several convolutional layers, where their output filter banks are concatenated and used as the input to the next module. The model version difference is based on the organization inside its inception module. Combining the residual connections with the InceptionV3 and InceptionV4 models, the author developed InceptionResNetV1 and InceptionResNetV2, respectively. The InceptionResNet building blocks are shown in Fig. 2.8. Experiments performed on the ImageNet database showed that the InceptionResNet models trained faster and reached slightly better results than the inception architecture [7].

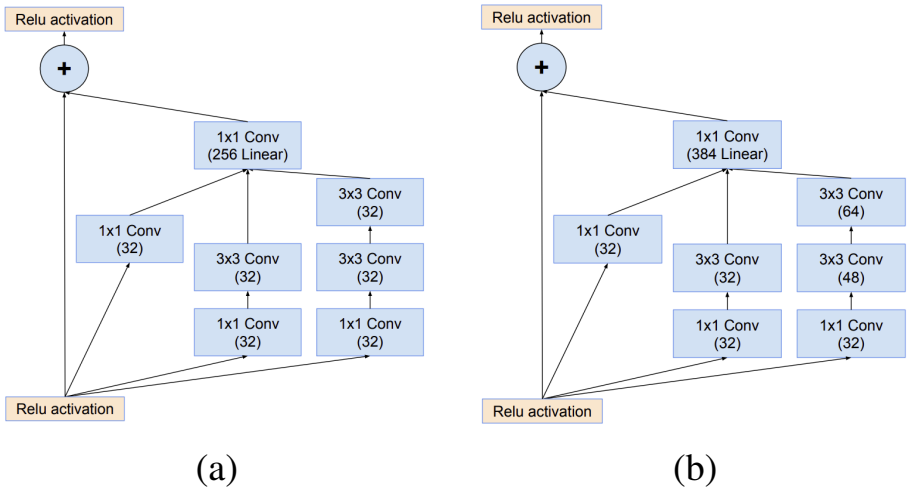


Figure 2.8: The schema for 35×35 grid (Inception-ResNet-A) module of the InceptionResNetV1 (a) and InceptionResNetV2 (b). Adapted from [7].

The first version of the MobileNet model (MobileNetV1) [154] was developed focusing on mobile and embedded vision applications, in which the CNN model should have a small size and high computational efficiency. This model is based on depthwise separable filters, which are composed of depthwise and pointwise convolutions. As described in [154], depthwise convolutions apply a single filter for each input channel, and pointwise convolutions use a 1×1 convolution to compute a linear combination of the depthwise output. Both layers use batch normalization and ReLU activation. MobileNetV1 achieved promising results in performance and accuracy on several tasks such as fine-grained recognition, large-scale geolocation, face attributes classification, object detection, and face recognition [154]. MobileNetV2 [138] combines the first version architecture with an inverted ResNet [6] structure, which has shortcut connections between the bottleneck layers. Experiments performed in different tasks such as image classification, object detection, and image segmentation showed that the MobileNetV2 could achieve high accuracy with low computation costs compared to state-of-the-art methods [138].

The Dense Convolutional Network (DenseNet) model [136] consists of a CNN architecture where each layer is connected to every other layer in a feed-forward way. Thus, let L be the number of layers from a network, a DenseNet layer has $\frac{L(L+1)}{2}$ direct connections with subsequent layers – instead of L as a traditional CNN model. As in the ResNet models [6, 150], these connections can handle the vanishing-gradient problem and ensure maximum information

flow between layers. The feed-forward is preserved, passing the output from all layers as an additional input to the subsequent ones in a channel-wise concatenation. The DenseNet models achieved state-of-the-art accuracies in image classification on the CIFAR10/100 and ImageNet databases [131, 136]. The authors proposed different models varying the depth of the network.

Inception modules inspired the creation of the Xception model, which can be defined as an intermediate step between regular convolution and the depthwise separable convolution operation [135]. The proposed architecture replaces the standard inception modules with depthwise separable convolutions and also has residual connections. The Xception architecture has the same number of parameters as InceptionV3 but outperforms it on the ImageNet database [131].

2.2.2 Multi-task Learning

Multi-task learning uses the domain information of related tasks as an inductive bias to improve generalization [155]. A Multi-task network can learn several tasks using a shared CNN model, where each task can help the generalization for other tasks. Caruana [155] introduced the Multi-task learning concept and evaluated it in different domains, demonstrating that this method can achieve better results than single-task learning models for related tasks. In deep neural networks, multi-task learning can be performed by using hard or soft parameter sharing [156]. The most common one is the hard parameter sharing, where all the hidden (convolutional) layers weights are shared, i.e., the model learns a single representation for all tasks. Then, different tasks use these shared features by adding some layers for each specific task. On the other hand, in soft parameter sharing, one model is employed for each task. Then, the parameters of these models are regularized to encourage similarities among them. Fig. 2.6 shows a Multi-task network sharing all the convolutional layers and some dense layers. The model has exclusive dense layers for each task, followed by the prediction layers, using the softmax cross-entropy as function loss.

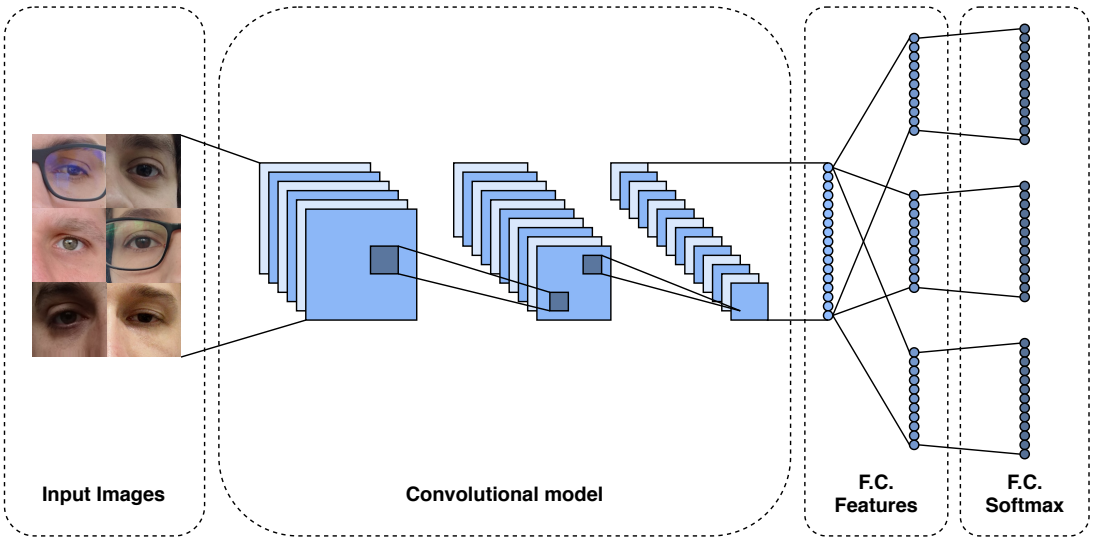


Figure 2.9: Multi-task CNN architecture. In this model, each task has its own output, and all tasks share the convolutional layers. The loss of all tasks is used to update the weights of the convolutional layers.

2.2.3 Pairwise Filters Network

This kind of model directly learns the similarity between a pair of images through pairwise filters. The Pairwise Filters Network is a Multi-class classification model that contains one or two outputs informing whether the input pairs are from the same or different classes. The difference

is that the network input is a pair of images instead of a single image. The network architecture is usually composed f convolutional, pooling, activation, and fully connected layers, as shown in Fig. 2.10.

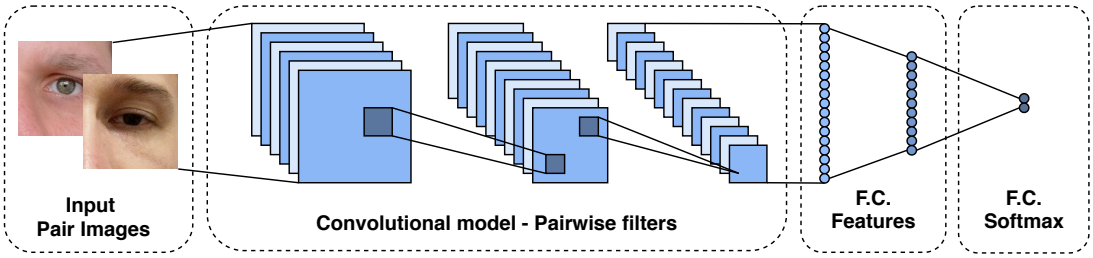


Figure 2.10: Pairwise filters CNN architecture. This model contains filters that directly learn the similarity between a pair of images. The output informs whether the images are of the same person or not.

Liu et al. [90] proposed one of the first works applying deep learning for ocular biometrics (iris verification) employing a pairwise filters network. As this model requires a pair of images as input, the authors generated the input pairs by concatenating the images at the depth level. Let two RGB images with shapes of $224 \times 224 \times 3$, concatenating both images by their channels; the resulting input image will have a shape of $224 \times 224 \times 6$. For the verification problem, which has only two classes, this model's output can also have only one neuron using a binary cross-entropy loss function.

2.2.4 Siamese Network

Introduced by Bromley et al. [157] for signature verification, Siamese networks consist of twin branches sharing their parameters (trainable parameters). Such models learn similarities/distances between a pair of inputs, being used mainly for verification tasks. As illustrated in Fig. 2.11, each branch of the Siamese structure is composed of a CNN model followed by some dense layers. These models can also have shared and non-shared dense layers at the top.

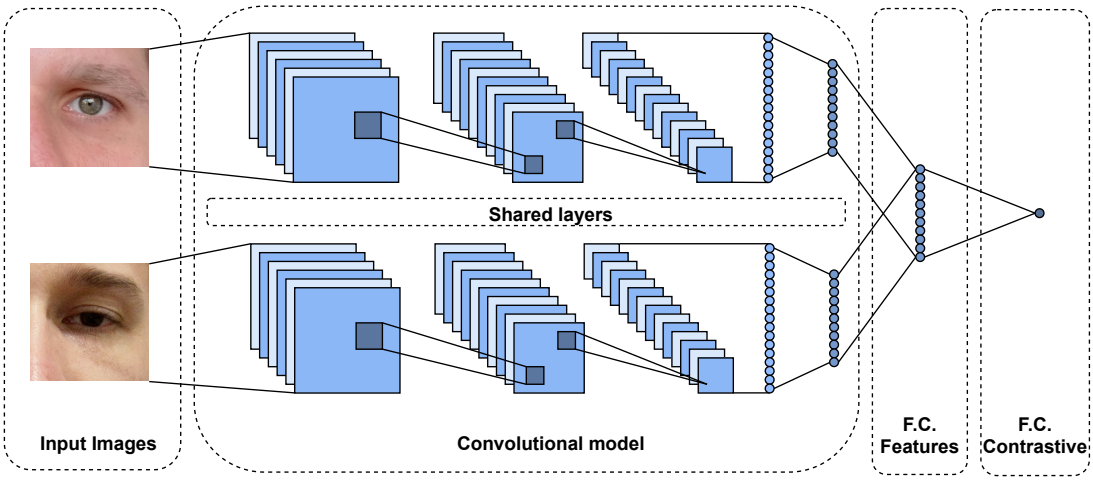


Figure 2.11: Siamese CNN architecture. This model is composed of two twin branches of convolutional layers sharing their trainable parameters. The output computes a distance between the input image pairs.

A typical Siamese architecture employs the contrastive loss as an activation function in the last layer. The contrastive loss was proposed and applied to face verification [158, 159] and has been employed for periocular recognition [88, 147] and iris recognition [29]. As described

in [159], let D_W be the Euclidean distance between two input vectors, the contrastive loss can be written as follows:

$$C(W) = \sum_{i=1}^{P} L(W, (Y, \vec{X}_1, \vec{X}_2)^i), \tag{2.12}$$

where

$$L(W, (Y, \vec{X}_1, \vec{X}_2)^i) = (1 - Y)L_S(D_W^i) + YL_D(D_W^i)\,, \tag{2.13}$$

and P is the number of training pairs, $(Y, \vec{X}_1, \vec{X}_2)^i$ corresponds to the i-th label (Y) of the sample pair $\vec{X}_1, \vec{X}_2$, and L_S and L_D are partial losses for a pair of similar and dissimilar points, respectively. The objective of this function is to minimize L for L_S and L_D by computing low and high values of D_W for similar and dissimilar pairs, respectively.

2.2.5 Final Remarks

This Chapter described and detailed biometrics fundamentals and ocular recognition employing iris and periocular traits. We also detailed several CNN architectures and models that have been employed to design robust ocular biometric approaches. Note that we employed all the described CNN architectures and models as a benchmark to our new collected dataset (UFPR-Periocular), detailed in Section 4.4.3 and evaluated in Section 5.4.

3 LITERATURE REVIEW

This chapter presents a literature review describing surveys, databases, competitions, and deep learning-based methods on ocular biometrics. The remainder of this chapter is organized as follows: Section 3.1 presents works that survey state-of-the-art methods on ocular recognition. To summarize and describe applications using ocular images, Section 3.2 presents several ocular databases. The evolution of the problems and challenges, as well as some solutions, are presented in Section 3.3, which describes competitions in ocular recognition. Finally, Section 3.4 presents recent deep learning-based approaches applied to iris and periocular biometrics.

3.1 SURVEYS ON OCULAR RECOGNITION

One of the first surveys on iris recognition was presented by Wildes [33], who examined iris recognition biometric systems and issues in the design and operation of such systems. This work explored the typical steps present in ocular biometric systems employing iris images, such as image acquisition, iris detection, and pattern matching. Also, some iris recognition methodologies that are still used as baseline are described [57, 160].

Bowyer et al. [32] described both the historical and the state-of-the-art development in iris biometrics, focusing on segmentation and recognition methodologies. This survey detailed several works organized by the steps present in standard iris recognition systems, such as image acquisition, segmentation, analysis and iris texture representation, and feature matching.

In [161], the authors surveyed researches for iris image acquisition, preprocessing techniques, segmentation approaches, feature extraction methods, matching, and indexing methods. The work also addressed problems such as off-angle iris recognition, spoofing, template aging (change in the iris with time), and recognition in uncontrolled, cross-spectral, and cross-sensor environments. Software and databases for ocular recognition were also described, organized into the iris, periocular, iris/periocular, and eye movement. For future research, the authors proposed the following directions: improvement of sensing technology, exploration of advanced machine learning algorithms for better representation and classification algorithms, heterogeneous recognition, ocular recognition at a distance, multimodal ocular biometrics, benchmark standards, and open-source software.

Focusing only on recognition, DeMarsico et al. [95] surveyed iris recognition through machine learning techniques. The work presented researches using different machine learning methods such as Artificial Neural Network (ANN), Self-Organizing Map (SOM), Radial Basis Function Neural Network (RBFNN), Fuzzy Neural Networks, Probabilistic Neural Network (PNN), Gabor Wavelet Neural Networks, Restricted Boltzmann Machines (RBM), and SVM. The authors highlighted the different types and architectures of artificial neural networks and their specific advantages, emphasizing the potential of deep learning for feature representation.

Nguyen et al. [162] discussed the design and implementation of iris recognition systems at a distance addressing long-range iris recognition. The authors also presented a solution for an iris recognition system at a distance with hardware and algorithms and a discussion about the fusion of ocular information and iris to improve the system's performance.

Regarding ocular biometrics in the visible spectrum, Rattani and Derakhshani [163] described state-of-the-art methods for periocular, iris, and conjunctival vasculature recognition. The authors also proposed a hardware-based acquisition set-up for ocular data and reported results for intra-ocular fusion.

3.2 OCULAR DATABASES

Currently, there are various databases of ocular images, constructed in different scenarios and for different purposes. These databases can be classified by VIS and NIR images and separated into controlled (cooperatives) and uncontrolled (non-cooperatives) environments, according to the process of image acquisition. Controlled databases contain images captured in environments with controlled conditions, such as lighting, distance, and focus. On the other hand, uncontrolled databases are composed of images obtained in uncontrolled environments and usually present problems such as defocus, occlusion, reflection, off-angle, to cite a few. A database containing images captured at different wavelengths is referred to as cross-spectral, while a database with images acquired by different sensors is referred to as cross-sensor. The summary of all databases cited in this work as well as links to find more information about how they are available can be found at [`www.inf.ufpr.br/vri/publications/ocularDatabases.html`].

The ocular databases described bellow are presented and organized into three subsections. First, we describe databases that contain only NIR images, as well as synthetic iris databases. Then, we present databases composed of images captured at both VIS and cross-spectral scenarios (i.e., VIS and NIR images from the same subjects). Finally, we describe multimodal databases, which contain data from different biometric traits, including iris and/or periocular.

3.2.1 Near-Infrared Ocular Images Databases

Ocular images captured at NIR wavelength are generally used to study the features present in the iris [2, 108, 109]. As even darker pigmentation irises reveal rich and complex features [4], most of the visible light is absorbed by the melanin pigment while longer wavelengths of light are reflected [32]. Other studies can also be performed with this kind of databases, such as methodologies to create synthetic irises [164, 165], vulnerabilities in iris recognition and liveness detection [166, 167, 168, 169], impact of contact lenses in iris recognition [170, 9, 12, 8], template aging [171, 172], influence of alcohol consumption [173], and study of gender recognition through the iris [174]. The databases used for these and other studies are described in Table 3.1 and detailed in this session. Some samples of ocular images from NIR databases are shown in Figure 3.1.

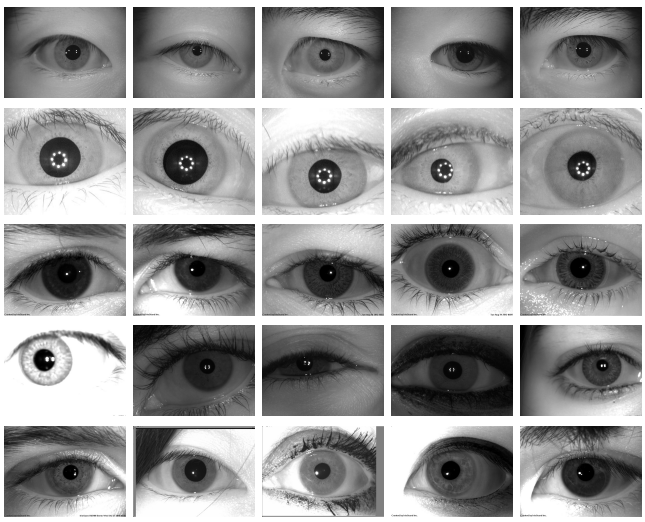


Figure 3.1: From top to bottom: NIR ocular image samples from the CASIA-IrisV3-Lamp [2], CASIA-IrisV3-Interval [2], NDCLD15 [8], IIITD CLI [9, 10], and ND Cosmetic Contact Lenses [11, 12] databases. Extracted from [13].

Table 3.1: NIR ocular databases. Modalities: Iris [IR] and Periocular [PR]. Extracted from [13].

Database	Year	Controlled Environment	Cross-sensor	Subjects	Images	Modality
CASIA-IrisV1 [2]	2002	Yes	No	*108 eyes	756	[IR]
CASIA-IrisV2 [2]	2004	Yes	Yes	*120 classes	2,400	[IR]
ND-IRIS-0405 [109]	2005	Yes	No	356	64,980	[IR]
ICE 2005 [108]	2005	Yes	No	132	2,953	[IR]
ICE 2006 [109]	2006	No	No	240	59,558	[IR]
WVU Synthetic Iris Texture Based [164]	2006	N/A	N/A	*1,000 classes	7,000	[IR]
WVU Synthetic Iris Model Based [165]	2007	N/A	N/A	5,000	160,000	[IR]
Fake Iris Database [166]	2008	N/A	No	50	800	[IR]
CASIA-IrisV3-Interval [2]	2010	Yes	No	249	2,639	[IR]
CASIA-IrisV3-Lamp [2]	2010	Yes	No	411	16,212	[IR]
CASIA-IrisV3-Twins [2]	2010	Yes	No	200	3,183	[IR]
CASIA-IrisV4-Thousand [2]	2010	Yes	No	1,000	20,000	[IR]
CASIA-IrisV4-Syn [2]	2010	N/A	N/A	*1,000 classes	10,000	[IR]
IIT Delhi Iris [175]	2010	Yes	No	224	1,120	[IR]
ND Iris Contact Lenses 2010 [170]	2010	Yes	No	124	21,700	[IR]
ND Iris Template Aging [171]	2012	Yes	No	322	22,156	[IR]
ND TimeLapseIris [172]	2012	Yes	No	23	6,797	[IR]
IIITD IUAI [173]	2012	Yes	No	55	440	[IR]
IIITD CLI [9]	2013	Yes	Yes	101	6,570	[IR]
ND Cosmetic Contact Lenses [11, 12]	2013	Yes	Yes	N/A	5,100	[IR]
ND Cross-Sensor-Iris-2013 [176]	2013	Yes	Yes	676	146,550	[IR]
Database of Iris Printouts [167]	2013	Yes	No	*243 eyes	1,976	[IR]
IIITD Iris Spoofing [168]	2014	Yes	Yes	101	4,848	[IR]
NDCLD15 [8]	2015	Yes	Yes	N/A	7,300	[IR]
IIITD Combined Spoofing [169]	2016	N/A	Yes	1,872	20,693	[IR]
ND-GFI [174]	2016	Yes	No	1,500	3,000	[IR]
BERC mobile-iris database [177]	2016	No	No	100	500	[IR]
Cataract Surgery on Iris [178]	2016	Yes	No	84	504	[IR]
ORNL [179]	2016	Yes	No	50	1,100	[IR]
MUID [180]	2016	Yes	No	111	24,360	[IR]
CASIA-Iris-Mobile-V1.0 [89]	2018	Yes	Yes	630	11,000	[IR]/[PR]
OpenEDS [181]	2019	Yes	No	152	356,649	[IR]

One of the first iris databases found in the literature was created and made available by CASIA (Chinese Academy of Science). The first version, called CASIA-IrisV1, was made available in 2002. The CASIA-IrisV1 database has 756 images of 108 eyes with a size of 320×280 pixels. The NIR images were captured in two sections with a homemade iris camera [2]. In a second version (CASIA-IrisV2), made available in 2004, the authors included two subsets captured by an OKI IRISPASS-h and CASIA-IrisCamV2 sensors. Each subset has 1,200 images belonging to 60 classes with a resolution of 640×480 pixels [2]. The third version of the database (CASIA-IrisV3), made available in 2010, has a total of 22,034 images from more than 700 individuals, arranged among its three subsets: CASIA-Iris-Interval, CASIA-Iris-Lamp and CASIA-Iris-Twins. Finally, CASIA-IrisV4, an extension of CASIA-IrisV3 and also made available in 2010, is composed of six subsets: three from the previous version and three new ones: CASIA-Iris-Distance, CASIA-Iris-Thousand and CASIA-Iris-Syn. All six subsets together contain 54,601 ocular images belonging to more than 1,800 real subjects and 1,000 synthetic ones. Each subset will be detailed below, according to the specifications described in [2].

The CASIA-Iris-Interval database has images captured under a near-infrared LED illumination. In this way, these images are used to study the texture information contained in the iris traits. The database is composed of 2,639 images, obtained in two sections, from 249 subjects and 395 classes with a resolution of 320×280 pixels.

The images from the CASIA-Iris-Lamp database were acquire by a non-fixed sensor (OKI IRISPASS-h) and thus the individual collected the iris image with the sensor in their own hands. During the acquisition, a lamp was switched on and off to produce more intra-class variations due to contraction and expansion of the pupil, creating a non-linear deformation.

Therefore, this database can be used to study problems such as iris normalization and robust iris feature representation. A total of 16,212 images, from 411 subjects, with a resolution of 640×480 pixels were collected in a single section.

During an annual twin festival in Beijing, iris images from 100 pairs of twins were collected to form the CASIA-Iris-Twins database, enabling the study of similarity between iris patterns of twins. This database contains 3,183 images (400 classes from 200 subjects) captured in a single section with the OKI IRISPASS-h camera at a resolution of 640×480 pixels.

The CASIA-Iris-Thousand database is composed of 20,000 ocular images from 1,000 subjects, with a resolution of 640×480 pixels, collected in a single section by an IKEMB-100 IrisKing camera [182]. Due to a large number of subjects, this database can be used to study the uniqueness of iris features. The main source of intra-class variations that occur in this database is due to specular reflections and eyeglasses.

The last subset of CASIA-IrisV4, called CASIA-IRIS-Syn, is composed of iris images generated with iris textures automatically synthesized from the CASIA-IrisV1 subset. The generation process applied the segmentation approach proposed by Tan et al. [183]. Factors such as blurring, deformation, and rotation were introduced to create some intra-class variations. In total, this database has 10,000 images belonging to 1,000 classes.

The images from the ND-IRIS-0405 [109] database were captured with the LG2200 imaging system using NIR illumination. The database contains 64,980 images from 356 subjects and there are several images with subjects wearing contact lenses. Even the images being captured under a controlled environment, some conditions such as blur, occlusion of part of the iris region, and problems like off-angle may occur. The ND-IRIS-0405 is a superset of the databases used in the ICE 2005 [108] and ICE 2006 [109] competitions.

The ICE 2005 database was created for the Iris Challenge Evaluation 2005 competition [108]. This database contains a total of 2,953 iris images from 132 subjects. The images were captured under NIR illumination using a complete LG EOU 2200 acquisition system with a resolution of 640×480 pixels. Images that did not pass through the automatic quality control of the acquisition system were also added to the database. Experiments were performed independently for the left and right eyes. The results of the competition can be seen in [108].

The ICE 2006 database has images collected using the LG EOU 2200 acquisition system with a resolution of 640×480 pixels. For each subject, two 'shots' of 3 images of each eye were performed per session, totaling 12 images. The imaging sessions were held in three academic semesters between 2004 and 2005. The database has a total of 59,558 iris images from 240 subjects [109].

The WVU Synthetic Iris Texture Based database, created at West Virginia University, has 1,000 classes with 7 grayscale images each. It consists exclusively of synthetic data, with the irises being generated in two phases. First, a Markov Random Field model was used to generate the overall iris appearance texture. Then, a variety of features were generated (e.g., radial and concentric furrows, crypts and collarette) and incorporated into the iris texture. This database was created to evaluate iris recognition algorithms since, at the time of publication, there were few available iris databases and they had a small number of individuals [164].

The WVU Synthetic Iris Model Based database also consists of synthetically generated iris images. This database contains 10,000 classes from 5,000 individuals, with degenerated images by a combination of several effects such as specular reflection, noise, blur, rotation, and low contrast. The image gallery was created in five steps using a model and anatomy-based approach [165], which contains 40 randomized and controlled parameters. The evaluation of their synthetic iris generation methodology was performed using a traditional Gabor filter-based

iris recognition system. This database provides a large amount of data that can be used to evaluate ocular biometric systems.

The Fake Iris Database was created using images from 50 subjects belonging to the BioSec baseline database [184] and has 800 fake iris images [166]. The process for creating new images is divided into three steps. The original images were first reprocessed to improve quality using techniques such as noise filtering, histogram equalization, opening/closing, and top hat. Then, the images were printed on paper using two commercial printers: an HP Deskjet 970cxi and an HP LaserJet 4200L, with six distinct types of papers: white paper, recycled paper, photographic paper, high-resolution paper, butter paper, and cardboard [166]. Finally, the printed images were recaptured by an LG IrisAccess EOU3000 camera.

The IIT Delhi Iris database consists of 1,120 images, with a resolution of 320×240 pixels, from 224 subjects captured with the JIRIS JPC1000 digital CMOS camera. This database was created to provide a large-scale database of iris images of Indian users. In [175], Kumar and Passi employed these images to compare the performance of different approaches for iris identification (e.g., Discrete Cosine Transform, Fast Fourier Transform, Haar wavelet, and Log-Gabor filter) and to investigate the impact in recognition performance using a score-level combination.

The images from the ND Iris Contact Lenses 2010 database were captured using the LG 2200 iris imaging system. Visual inspections were performed to reject low-quality images or those with poor results in segmentation and matching. To compose the database, the authors captured 9,697 images from 124 subjects that were not wearing contact lenses and 12,003 images from 87 subjects that were wearing contact lenses. More specifically, the images were acquired from 92 subjects not wearing lenses, 52 subjects wearing the same lens type in all acquisitions, 32 subjects who wore lenses only in some acquisitions and 3 subjects that changed the lens type between acquisitions [170]. According to Baker et al. [170], the purpose of this database is to verify the degradation of iris recognition performance due to non-cosmetic prescription contact lenses.

The ND Iris Template Aging database, described and used by Fenker and Bowyer [171], was created to analyze the template aging in iris biometrics. The images were collected from 2008 to 2011 using an LG 4000 sensor, which captures images at NIR. This database has 22,156 images, being 2,312 from 2008, 5,859 from 2009, 6,215 from 2010 and 7,770 from 2011, corresponding to 644 irises from 322 subjects. The ND-Iris-Template-Aging-2008-2010 subset belongs to this database.

All images from the ND TimeLapseIris database [172] were taken with the LG 2200 iris imaging system, without hardware or software modifications throughout 4 years. Imaging sessions were held at each academic semester over 4 years, with 6 images of each eye being captured per individual in each session. From 2004 to 2008, a total of 6,797 images were obtained from 23 subjects who were not wearing eyeglasses, 5 subjects who were wearing contact lenses, and 18 subjects who were not wearing eyeglasses or contact lenses in any session. This database was created to investigate template aging in iris biometrics.

To investigate the effect of alcohol consumption on iris recognition, Arora et al. [173] created the Iris Under Alcohol Influence (IIITD IUAI) database, which contains 440 images from 55 subjects, with 220 images being acquired before alcohol consumption and 220 after it. The subjects consumed approximately 200 ml of alcohol (with 42% concentration level) in approximately 15 minutes, and the second half of the images were taken between 15 and 20 minutes after consumption. Due to alcohol consumption, there is a deformation in iris patterns caused by the dilation of the pupil, affecting iris recognition performance [173]. The images were captured using the Vista IRIS scanner at NIR wavelength.

The IIITD Contact Lens Iris (IIITD CLI) database is composed of 6,570 iris images belonging to 101 subjects. The images were captured by two different sensors: Cogent CIS 202 dual iris sensor and VistaFA2E single iris sensor with each subject (i) not wearing contact lenses, (ii) wearing color cosmetic lenses, and (iii) wearing transparent lenses. Four lens colors were used: blue, gray, hazel and green. At least 5 images of each iris were collected in each lens category for each sensor [9].

The images from the ND Cosmetic Contact Lenses database [11] were captured by two iris cameras, an LG4000 and an IrisGuard AD100, in a controlled environment under NIR illumination with a resolution of 640×480 pixels. These images are divided into four classes, (i) no contact lenses, (ii) soft, (iii) non-textured and (iv) textured contact lenses. Also, this database is organized into two subsets: Subset1 (LG4000) and Subset2 (AD100). Subset1 has 3,000 images in the training set and 1,200 images in the validation set. Subset2 contains 600 and 300 images for training and validation, respectively [12, 10, 42]. Both subsets have 10 equal folds of training images for testing purposes.

The ND Cross-Sensor-Iris-2013 database [176] is composed of 146,550 NIR images belonging to 676 unique subjects, being 29,986 images captured using an LG4000 and 116,564 taken by an LG2200 iris sensor with 640×480 pixels of resolution. The images were captured in 27 sessions over three years, from 2008 to 2010, and in at least two sessions there are images of the same subject. The purpose of this database is to investigate the effect of cross-sensor images on iris recognition. Initially, this database was released for a competition to be held at the BTAS 2013 Conference, but the competition did not have enough submission.

The Database of Iris Printouts was created for liveness detection in iris images and contains 729 printout images of 243 eyes, and 1,274 images of imitations from genuine eyes. The database was constructed as follows. First, the iris images were obtained with an IrisGuard AD100 camera. Then, they were printed using the HP LaserJet 1320 and Lexmark c534dn printers. To check the print quality, the printed images were captured by the Panasonic ET-100 camera using an iris recognition software, and the images that were successfully recognized were recaptured by an AD100 camera with a resolution of 640×480 pixels to create the imitation subset. Initially, images from 426 distinct eyes belonging to 237 subjects were collected. After the process of recognizing the printed images, 243 eyes images (which compose the database) were successfully verified [167].

The IIITD Iris Spoofing (IIS) database was created to study spoofing methods. To this end, printed images from the IIITD CLI [9] database were used. Spoofing was simulated in two ways. In the first, the printed images were captured by a specific iris scanner (Cogent CIS 202 dual eye), while in the second, the printed images were scanned using an HP flatbed optical scanner. The database contains 4,848 images from 101 individuals [168].

The Notre Dame Contact Lenses 2015 (NDCLD15) database contains 7,300 iris images. The images were obtained under consistent lighting conditions by an LG4000 and an IrisGuard AD100 sensor. All images have 640×480 pixels of resolution and are divided into three classes based on the lens type: no lens, soft, and textured. This database was created to investigate methods to classify iris images based on types of contact lenses [8].

The IIITD Combined Spoofing database was proposed to simulate a real-world scenario of attacks against iris recognition systems. This database consists of joining the following databases: IIITD CLI [9], IIITD IIS [168], SDB [164], IIT Delhi Iris [175] and, to represent genuine classes, iris images from 547 subjects were collected. The CSD database has a total of 1,872 subjects, with 9,325 normal image samples and 11,368 samples of impostor images [169].

The Gender from Iris (ND-GFI) database was created to study the recognition of the subject's gender through the iris, specifically using the binary iris code (which is normally used

in iris recognition systems) [174]. The images were obtained at NIR wavelength by an LG4000 sensor and labeled by gender. The ND-GFI database contains a single image of each eye (left and right) from 750 men and 750 women, totaling 3,000 images. About a quarter of the images were captured with the subjects wearing clear contact lenses. This database has another set of images that can be used for validation, called UND_V, containing 1,944 images, being 3 images of each eye from 175 men and 149 women. In this subset, there are also images using clear contact lenses and some cosmetics [174].

According to [185], an iris image has good quality if the iris diameter is larger than 200 pixels, and if the diameter is between 150 and 200 pixels, the image is classified as adequate quality. In this context, the images from the BERC mobile-iris database have irises with a diameter between 170 and 200 pixels, obtained at NIR wavelength with 1280×960 pixels of resolution. Using a mobile iris recognition system, the images were taken in sequences of 90 shots [177] moving the device at three distances: 15 to 25 cm, 25 to 15 cm, and 40 to 15 cm. In total, the database has 500 images from 100 subjects, which were the best ones selected by the authors of each sequence.

Raghavendra et al. [178] created the Cataract Surgery on Iris database to analyze the impact of cataract surgery on the verification performance of iris recognition systems. The database contains 504 images belonging to 84 subjects who were affected by cataracts. The subjects' ages vary from 50 to 80 years, being 34 males and 49 females. Three eye samples of each subject were collected before (24 hours) and after (36 - 42 hours) the surgery to remove the cataractous lens. The images were captured using a commercial dual-iris NIR device with a resolution of 640×480 pixels.

The Oak Ridge National Laboratory (ORNL) Off-angle database was created to study how the gaze angle affects the performance of iris biometrics [186, 179, 187]. This database encompasses 1,100 NIR iris images from 50 subjects varying the angle acquisition from $-50°$ to $+50°$ with a step-size of $10°$. The gender distribution consists of 56% male and 44% female subjects, and iris color of 64% with dark colors and 36% with light-colors. The images were collected by a Toshiba Teli CleverDragon series camera and have a resolution of 4096×3072 pixels.

The Meliksah University Iris Database (MUID) was collected to investigate the off-angle iris recognition. The authors developed an iris image capture system composed of two cameras to simultaneously capture frontal and off-angle samples. Thus, it is possible to isolate the effect of the gaze angle from pupil dilation and accommodation [180]. In total, the database has 24,360 NIR images from 111 subjects, 64 males and 57 females, with an average age of 26 years. The images were captured by two infrared-sensitive IDS-UI-3240ML-NIR cameras varying from $-50°$ to $+50°$ angles with a step-size of $10°$ and have a resolution of 1280×1024 pixels. More details about the iris image acquisition platform are described in [180].

The CASIA-Iris-Mobile-V1.0 database is composed of 11,000 NIR images belonging to 630 subjects, divided into three subsets: CASIA-Iris-M1-S1 [188], CASIA-Iris-M1-S2 [189] and a new one called CASIA-IRIS-M1-S3. The images were captured simultaneously from the left and right eyes and stored in 8 bits gray-level JPG files. The CASIA-Iris-M1-S1 subset has 1,400 images from 70 subjects with a resolution of 1920×1080 pixels, acquired using a NIR imaging module attached to a mobile phone. The CASIA-Iris-M1-S2 subset has images captured using a similar device. In total, this subset contains 6,000 images from 200 subjects with a resolution of 1968×1024 pixels, collected at three distances: 20, 25 and 30 cm. At last, the CASIA-Iris-M1-S3 subset is composed of 3,600 images belonging to 360 subjects with a resolution of 1920×1920 pixels, which were taken with a NIR iris-scanning technology equipped on a mobile phone.

The Open Eye Dataset (OpenEDS) was created to investigate the semantic segmentation of eyes components, and background [181]. This database is composed of 356,649 eye images,

being 12,759 images with pixel-level annotations, 252,690 unlabeled ones, and 91,200 images from video sequences belonging from 152 subjects. The images were captured with a head-mounted display with two synchronized cameras under controlled NIR illumination with a resolution of 640×400 pixels.

3.2.2 Visible and Cross-spectral Ocular Images Databases

Iris recognition using images taken at controlled NIR wavelength environments is a mature technology, proving to be effective in different scenarios [32, 108, 109, 52, 34, 94]. Databases captured under controlled environments have few or no noise factors in the images. However, these conditions are not easy to achieve and require a high degree of collaboration from subjects. In a more challenging/realistic scenario, investigations on biometric recognition employing iris images obtained in uncontrolled environments and at VIS wavelength have begun to be conducted [190, 1]. There is also research on biometric recognition using cross-spectral databases, i.e., databases with ocular images from the same individual obtained at both NIR and VIS wavelengths [191, 192, 20, 193, 29]. Currently, many types of research have been performed on biometric recognition using iris and periocular region with images obtained from mobile devices, obtained in an uncontrolled environment and by different types of sensors [15, 194, 14]. In this subsection, we describe databases with these characteristics. Table 3.2 summarize these databases. Some samples of ocular images from VIS and Cross-spectral databases are shown in Figure 3.2.

Figure 3.2: From top to bottom: VIS and Cross-spectral ocular image samples from the VISOB [14], MICHE-I [15], UBIPr [16], UFPR-Periocular [17], CROSS-EYED [18, 19], PolyU Cross-Spectral [20] databases. Extracted from [13].

The UPOL (University of Palackeho and Olomouc) database has high-quality iris images obtained at VIS wavelength using the optometric framework (TOPCON TRC501A) and the Sony DXC-950P 3CCD camera. In total, 384 images of the left and right eyes were obtained from 64 subjects at a distance of approximately 0.15 cm with a resolution of 768×576 pixels, stored in 24 bits (RGB) [195].

Table 3.2: Visible and Cross-spectral ocular databases. Wavelengths: Near-Infrared (NIR), Visible (VIS) and Night Vision (NV). Modalities: Iris [IR] and Periocular [PR]. Extracted from [13].

Database	Year	Controlled Environment	Wavelength	Cross-sensor	Subjects	Images	Modality
UPOL [195]	2004	Yes	VIS	No	64	384	[IR]
UBIRIS.v1 [190]	2005	No	VIS	No	241	1,877	[IR]
UTIRIS [191]	2007	Yes	VIS / NIR	Yes	79	1,540	[IR]
UBIRIS.v2 [1]	2010	No	VIS	No	261	11,102	[IR]
UBIPr [16]	2012	No	VIS	No	261	10,950	[PR]
BDCP [196]	2012	No	VIS / NIR	Yes	99	4,314	[IR]/[PR]
MobBIOfake [197]	2013	No	VIS	No	N/A	1,600	[IR]
IIITD Multi-spectral Periocular [192]	2014	Yes	VIS / NIR / NV	Yes	62	1,240	[PR]
PolyU Cross-Spectral [20]	2015	N/A	VIS / NIR	Yes	209	12,540	[IR]
MICHE-I [15]	2015	No	VIS	Yes (Mobile)	92	3,732	[IR]
VSSIRIS [194]	2015	No	VIS	Yes (Mobile)	28	560	[IR]
CSIP [198]	2015	No	VIS	Yes (Mobile)	50	2,004	[IR]/[PR]
VISOB [14]	2016	No	VIS	Yes (Mobile)	550	158,136	[PR]
CROSS-EYED [18, 19]	2016	No	VIS / NIR	Yes	120	3,840	[IR]/[PR]
Post-mortem Human Iris [199]	2016	Yes	VIS / NIR	Yes	6	104	[IR]
QUT Multispectral Periocular [193]	2017	N/A	VIS / NIR / NV	Yes	53	212	[PR]
VISOB 2.0 [28]	2020	No	VIS	Yes	150	75,428	[PR]
I-SOCIAL-DB [200]	2020	No	VIS	No	400	3,286	[IR]/[PR]
UFPR-Periocular [17]	2020	No	VIS	No	1,122	33,660	[PR]
UFPR-Eyeglasses [25]	2020	No	VIS	No	83	2,270	[PR]

The UBIRIS.v1 database [190] was created to provide images with different types of noise, simulating image capture with minimal collaboration from the users. This database has 1,877 images belonging to 241 subjects, obtained in two sections by a Nikon E5700 camera. For the first section (enrollment), some noise factors such as reflection, lighting, and contrast were minimized. However, in the second section, natural lighting factors were introduced by changing the location to simulate an image capture with minimal or without active collaboration from the subjects. The database is available in three formats: color with a resolution of 800×600 pixels, color with 200×150 pixels, and 200×150 pixels in grayscale [190].

The UTIRIS is one of the first databases containing iris images captured at two different wavelengths (cross-spectral) [191]. The database is composed of 1,540 images of the left and right eyes from 79 subjects, resulting in 158 classes. The VIS images were obtained by a Canon EOS 10D camera with 2048×1360 pixels of resolution. To capture the NIR images, the ISW Lightwise LW camera was used, obtaining iris images with a resolution of 1000×776 pixels. As the melanin pigment provides a rich source of features at the VIS spectrum, which is not available at NIR, this database can be used to investigate the impact of the fusion of iris image features extracted at both wavelengths.

The UBIRIS.v2 database was built representing the most realistic noise factors. For this reason, the images that constitute the database were obtained at VIS without restrictions such as distance, angles, light, and movement. The main purpose of this database is to provide a tool for the research on the use of VIS images for iris recognition in an environment with adverse conditions. This database contains images captured by a Canon EOS 5D camera, with a resolution of 400×300 pixels, in RGB from 261 subjects containing 522 irises and 11,102 images taken in two sessions [1].

The UBIPr (University of Beira Interior Periocular) database [16] was created to investigate periocular recognition using images taken under uncontrolled environments and setups. The images from this database were captured by a Canon EOS 5D camera with a 400mm focal length. Five different distances and resolutions were configured: 501×401 pixels (8m), 561×541 pixels (7m), 651×501 pixels (6m), 801×651 pixels (5m), and 1001×801 pixels (4m). In total, the database has 10,950 images from 261 subjects (the images from 104 subjects were obtained in 2 sessions). Several variability factors were introduced in the images, for example,

different distances between the subject and the camera, as well as different illumination, poses and occlusions levels.

The BDCP (Biometrics Development Challenge Problem) database [196] contains images from two different sensors: an LG4000 sensor that captures images in gray levels, and a Honeywell Combined Face and Iris Recognition System (CFAIRS) camera [196], which captures VIS images. The resolutions of the images are 640×480 pixels for the LG4000 sensor and 750×600 pixels for the CFAIRS camera. To compose the database, 2,577 images from 82 subjects were acquired by the CFAIRS sensor and 1,737 images belonging to 99 subjects were taken by an LG4000 sensor. Images of the same subject were obtained for both sensors [201]. The main objective of this database is the cross-sensor evaluation, matching NIR against VIS images [163]. It should be noted that this database was used only in [201] and no availability information is reported.

Sequeira et al. [197] built the MobBIOfake database to investigate iris liveliness detection using images taken from mobile devices under an uncontrolled environment. It consists of 1,600 fake iris images obtained from a subset of the MobBIO database [21]. The fake images were generated by printing the original images using a professional printer in a high-quality photo paper and recapturing the image with the same device and environmental conditions used in the construction of MobBIO.

The images that compose the IIITD Multi-spectral Periocular database were obtained under a controlled environment at NIR, VIS, and night-vision spectra. The NIR images were captured by a Cogent iris Scanner sensor at a distance of 6 inches from the subject, while the night vision subset was created using the Sony Handycam camera in night vision mode at a distance of 1.3 meters. The VIS images were captured with the Nikon SLR camera, also at a distance of 1.3 meters. The database contains 1,240 images belonging to 62 subjects, being 310 images, 5 from each subject, at VIS and night vision spectra, and 620 images, 10 from each subject, at NIR spectrum [192].

Nalla and Kumar [20] developed the PolyU Cross-Spectral database to study iris recognition in the cross-spectral scenario. The images were obtained simultaneously under VIS and NIR illumination, totaling 12,540 images from 209 subjects with 640×480 pixels of resolution, being 15 images from each eye in each spectrum.

To evaluate the state of the art on iris recognition using images acquired by mobile devices, the Mobile Iris Challenge Evaluation (MICHE) competition (Part I) was created [15]. The MICHE-I (or MICHEDB) database consists of 3,732 VIS images obtained by mobile devices from 92 subjects. To simulate a real application, the iris images were obtained by the users themselves, indoors and outdoors, with and without glasses. Images of only one eye of each individual were captured. The mobile devices used and their respective resolutions are iPhone5 (1536×2048), Galaxy S4 (2322×4128) and Galaxy Tablet II (640×480). Due to the acquisition mode and the purpose of the database, several noises are found in images such as specular reflections, focus, motion blur, lighting variations, occlusion due to eyelids, among others. The authors also proposed a subset, called MICHE FAKE, containing 80 printed iris images. Such images were created as follows. First, they were captured using the iPhone5 the Samsung Galaxy S4 mobile devices. Then, using a LaserJet printer, the images were printed and captured again by a Samsung Galaxy S4 smartphone. There is still another subset, called MICHE Video, containing videos of irises from 10 subjects obtained indoor and outdoor. A Samsung Galaxy S4 and a Samsung Galaxy Tab 2 mobile devices were used to capture these videos. In total, this subset has 120 videos of approximately 15 seconds each.

The VSSIRIS database, proposed by Raja et al. [194], has a total of 560 images captured in a single session under an uncontrolled environment from 28 subjects. The purpose of this

database is to investigate the mixed lighting effect (natural daylight and artificial indoor) for iris recognition at the VIS spectrum with images obtained by mobile devices [194]. More specifically, the images were acquired by the rear camera of two smartphones: an iPhone 5S, with a resolution of 3264×2448 pixels, and a Nokia Lumia 1020, with a resolution of 7712×5360 pixels.

Santos et al. [198] created the CSIP (Cross-Sensor Iris and Periocular) database simulating mobile application scenarios. This database has images captured by four different device models: Xperia Arc S (Sony Ericsson), iPhone 4 (Apple), w200 (THL) and U8510 (Huawei). The resolutions of the images taken with these devices are as follows: Xperia Arc S (Rear 3264×2448), iPhone 4 (Front 640×480, Rear 2592×1936), W200 (Front 2592×1936, Rear 3264×2448) and U8510 (Front 640×480, Rear 2048×1536). Combining the models with front and rear cameras, as well as flash, 10 different setups were created with the images obtained. In order to simulate noise variation, the image capture sessions were carried out in different sites with the following lighting conditions: artificial, natural and mixed. Several noise factors are presented in these images, such as different scales, off-angle, defocus, gaze, occlusion, reflection, rotation and distortions [198]. The database has 2,004 images from 50 subjects and the binary iris segmentation masks were obtained using the method described by Tan et al. [183] (winners of the NICE I contest).

The VISOB database was created for the ICIP 2016 Competition on mobile ocular biometric recognition, whose main objective was to evaluate methods for mobile ocular recognition using images taken at the visible spectrum [14]. The front cameras of 3 mobile devices were used to obtain the images: iPhone 5S at 720p resolution, Samsung Note 4 at 1080p resolution and Oppo N1 at 1080p resolution. The images were captured in 2 sessions for each one of the 2 visits, which occurred between 2 and 4 weeks, counting in the total 158,136 images from 550 subjects. At each visit, it was required that each volunteer (subject) capture their face using each one of the three mobile devices at a distance between 8 and 12 inches from the face. For each image capture session, 3 light conditions settings were applied: regular office light, dim light, and natural daylight. The collected images were preprocessed using the Viola-Jones eye detector and the region of the image containing the eyes was cropped to a size of 240×160 pixels.

Sequeira et al. [18, 19] created the Cross-Spectral Iris/Periocular (CROSS-EYED) database to investigate iris and periocular region recognition in cross-spectral scenarios. CROSS-EYED is composed of VIS and NIR spectrum images obtained simultaneously with 2K×2K pixel resolution cameras. The database is organized into three subsets: ocular, periocular (without iris and sclera regions) and iris. There are 3,840 images from 120 subjects (240 classes), being 8 samples from each of the classes for every spectrum. The periocular/ocular images have dimensions of 900×800 pixels, while the iris images have dimensions of 400×300 pixels. All images were obtained at a distance of 1.5 meters, under uncontrolled indoor environment, with a wide variation of ethnicity and eye colors, and lightning reflexes.

The Post-mortem Human Iris database was collected to investigate the post-mortem human iris recognition. Due to the difficulty and restriction in collecting such images, this database has only 104 images from 6 subjects. The images were acquired in three sessions with an interval of approximately 11 hours using the IriShield M2120U NIR and Olympus TG-3 VIS cameras.

The QUT Multispectral Periocular database was developed and used by Algashaam et al. [193] to study multi-spectral periocular recognition. In total, 212 images belonging to 53 subjects were captured at VIS, NIR and night vision spectrum with 800×600 pixels of resolution. The VIS and NIR images were taken using a Sony DCR-DVD653E camera, while the night vision images were acquired with an IP2M-842B camera.

Regarding some ocular biometrics problems caused by substantial degradation due to variations on illumination, distance, noise, and blur when using single-frame mobile captures, Nguyen et al. [28] created the VISOB 2.0 database. This database comprises multi-frame captures and has stacks of eye images acquired using the burst mode of two mobile devices: Samsung Note 4 and Oppo N1. It is the second version of the VISOB database and was used in the 2020 IEEE WCCI competition [28]. The images were collected in two visits. At each visit, the subjects collected their own images under three lighting conditions in two sessions. The available subset of the VISOB 2.0 database (competition training set) has 75,428 images of left and right eyes belonging to 150 subjects. The VISOB 2.0 can also be employed to investigate the probing fairness of ocular biometrics across gender [53].

The Iris Social Database (I-SOCIAL-DB) has 3,286 VIS images from 400 subjects, being 43.75% male and 56.25% female. It is composed of images of public persons such as artists and athletes. This database was created by collecting 1,643 high-resolution portrait images using Google Image Search. Then, the ocular regions were cropped as rectangles of 350×300 pixels. The binary masks for the iris region (created by a human expert) are also available. This database can be employed to evaluate iris segmentation and recognition under unconstrained scenarios.

The UFPR-Periocular database has VIS images acquired in unconstrained environments by mobile devices. These images were captured by the subjects themselves using their own smartphone models through a mobile application (app) developed by the authors [17]. In total, this database contains 33,660 samples from 1,122 subjects acquired during 3 sessions by 196 different mobile devices. The image resolutions vary from 360×160 to 1862×1008 pixels. The main intra- and inter-class variability are caused by occlusion, blur, motion blur, specular reflection, eyeglasses, off-angle, eye-gaze, makeup, facial expression, and variations in lighting, distance, and angles. The authors manually annotated the eye corners and used them to normalize the periocular images regarding scale and rotation. This database can also be employed to investigate gender recognition, age estimation, and the effect of intra-class variability in biometric systems. The UFPR-Periocular database, which includes the manual annotations of the eye corners, as well as information on the subjects' age and gender, is publicly available for the research community.

Zanlorensi et al. [25] created the UFPR-Eyeglasses database to investigate intra-class variability and also the effect of the occlusion by eyeglasses in periocular recognition under uncontrolled environments. This database has 2,270 images captured by mobile devices from 83 subjects with a resolution of 256×256 pixels. The subjects captured the images using the same mobile app used to collect the UFPR-Periocular database. This database can be considered a subset of the UFPR-Periocular database containing some additional images. The authors manually annotated the iris's bounding box in each image and used it to perform scale and rotation normalization. The intra-class variations in this database are mainly caused by illumination, occlusions, distances, reflection, eyeglasses, and image quality. The UFPR-Eyeglasses database, which includes the authors' manual annotations, is publicly available to the research community.

3.2.3 Multimodal Databases

In addition to the databases proposed specifically to assist the development and evaluation of new methodologies for iris/periocular recognition, some multimodal databases can also be used for this purpose. Table 3.3 show these databases. As described in this subsection, most of these databases consist of iris images obtained at NIR wavelength. Figure 3.3 shows samples of ocular images from some multimodal databases.

The BioSec baseline database, proposed by Fierrez et al. [184], has biometric data of fingerprint, face, iris and voice. Data were acquired from 200 subjects in two acquisition sessions,

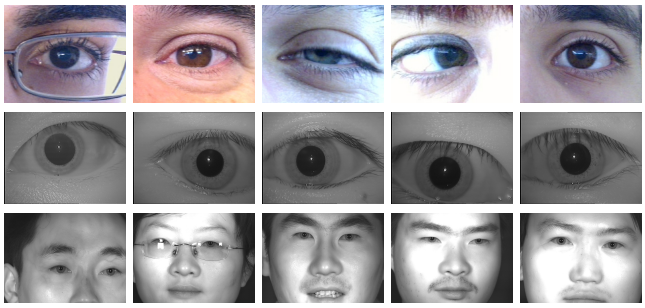


Figure 3.3: From top to bottom: ocular image samples from the MobBIO [21], SDUMLA-HMT [22] and CASIA-IrisV4-Distance [2] multimodal databases. Extracted from [13].

Table 3.3: Multimodal databases. Modalities: Face [FC], Fingerprint [FP], Finger vein [FV], Gait [GT], Hand [HD], Handwriting [HW], Iris [IR], KeyStroking [KS], Periocular [PR], Signature [SG], Speech [SP], and Voice [VC]. Extracted from [13].

Database	Year	Controlled Environment	Wavelength	Cross-sensor	Subjects	Images	Modality
BioSec [184]	2006	No	NIR	No	200	3,200	[IR]/[FC]/[FP]/[VC]
BiosecurID [202]	2007	Yes	NIR	No	400	12,800	[IR]/[FC]/[SP]/[SG]/[FP]/[HD]/[HW]/[KS]
BMDB [203]	2008	Yes	NIR	No	667	5,336	[IR]/[FC]/[SP]/[SG]/[FP]/[HD]
MBGC [204]	2009	No	NIR	No	*268 eyes	*986 videos	[IR]/[FC]
Q-FIRE [205]	2010	No	NIR	No	195	N/A	[IR]/[FC]
FOCS [206]	2010	No	NIR	No	136	9,581	[IR]/[PR]/[FC]
CASIA-IrisV4-Distance [2]	2010	Yes	NIR	No	142	2,567	[IR]/[PR]/[FC]
SDUMLA-HMT [22]	2011	Yes	NIR	No	106	1,060	[IR]/[FC]/[FV]/[GT]/[FP]
MobBIO [21]	2013	No	VIS	No	105	1,680	[IR]/[FC]/[VC]
gb2sμ MOD [207]	2015	Yes	NIR	No	60	*600 videos	[IR]/[FC]/[HD]

with environmental conditions (e.g., lighting and background noise) not being controlled to simulate a real situation. There are 3,200 NIR iris images, being 4 images of each eye for each session, captured by an LG IrisAccess EOU3000 camera [184].

The BiosecurID multimodal database consists of 8 unimodal biometric traits: iris, face, speech, signature, fingerprints, hand, handwriting, and keystroking [202]. The authors collected data from 400 subjects in four acquisition sessions through 4 months at six Spanish institutions. The iris images were captured at NIR by an LG Iris Access EOU 3000 camera with a resolution of 640×480 pixels. Four images of each eye were obtained in each of the 4 sessions, totaling 32 images per individual and a final set of 12,800 iris images.

The BMDB (multienvironment multiscale BioSecure Multimodal Database) [203] has biometric data from more than 600 subjects, obtained from 11 European institutions participating in the BioSecure Network of Excellence [203]. This database contains biometric data of iris, face, speech, signature, fingerprint and hand, and is organized into three subsets: DS1, which has data collected from the Internet under unsupervised conditions; DS2, with data obtained in an office environment under supervision; and DS3, in which mobile hardware was used to take data indoor and outdoor. The iris images belong to the DS2 subset and were obtained in 2 sessions at NIR wavelength in an indoor environment with supervision. For the acquisition, the use of contact lenses was accepted, but glasses needed to be removed. Four images (2 of each eye) were obtained in each session for each of the 667 subjects, totaling 5,336 images. These images have a resolution of 640×480 pixels and were acquired by an LG Iris Access EOU3000 sensor.

The goal of the Multiple Biometrics Grand Challenge (MBGC) [204] was the evaluation of iris and face recognition methods using data obtained from still images and videos under unconstrained conditions [31]. The MBGC is divided into three problems: the portal challenge problem, the still face challenge problem, and the video challenge problem [204]. This competition has two versions. The first one was held to introduce the problems and protocol, whereas version

2 was released to evaluate the approaches in large databases [31]. The iris images were obtained from videos captured at NIR by an Iridian LG EOU 2200 camera [208]. The videos present variations such as pose, illumination, and camera angle. The MBGC database has 986 iris videos from 268 eyes collected in 2008 [208].

The Q-FIRE database (Quality in Face and Iris Research Ensemble) has iris and face images from 195 subjects, obtained through videos at different distances [205]. This database has 28 and 27 videos of face and iris, respectively, captured in 2 sections, with varying camera distance between 5, 7, 11, 15 and 25 feet. The videos have approximately 6 seconds each and were captured at approximately 25 frames per second. A Dalsa 4M30 infrared camera equipped with a Tamron AF 70-300mm 1:4.5-5.6 LD DI lens were used to capture iris videos. For distances of 15 and 25 feet, a Sigma APO 300-800mm F5.6 EX DG HSM lens was used. The most attractive distance of capture for iris is 5 (300×280 pixels), 7 (220×200 pixels), and 11 (120×100 pixels) feet since they respectively represent high, medium and low resolution, based on the number of pixels in the iris diameter. The images also have variations of illumination, defocus, blur, eye angles, motion blur, and occlusions [205].

The NIR images from the ocular region (iris and periocular) of the FOCS database [206] were extracted from the MBGC database [204] videos, which were collected from moving subjects [209]. These videos were captured in an uncontrolled environment presenting some variations such as noise, gaze, occlusion and lighting. The database has 9,581 images (4,792 left, 4,789 right) with a resolution of 750×600 pixels from 136 subjects [201].

The CASIA-IrisV4-Distance database consists of iris images acquired by a long-range multi-modal biometric image acquisition and recognition system developed by the database authors [2]. Their system can recognize users from up to 3 meters (10 feet) using a system with an active search for iris, face or palmprint patterns. The images were taken using a camera with high resolution so that a single image includes regions of interest for both eyes and face traits. Information from the face trait such as skin pattern can also be used for multi-modal fusion. The database has 2,567 images from 142 individuals and 284 classes with a resolution of 2352×1728 pixels.

The SDUMLA-HMT multimodal database contains biometric traits of iris, face, finger vein, gait, and fingerprint [22]. All data belong to 106 subjects and were collected at Shandong University in China. The iris images were collected at NIR and under a controlled environment at a distance of 6 cm to 32 cm between the camera and the subject. In total, the authors collected 1,060 iris images with 768×576 pixels of resolution, being 10 images (5 of each eye) from each subject [22].

Sequeira et al. [21] created the MobBIO database due to the growing interest in mobile biometric applications, as well as the growing interest and application of multimodal biometrics. This database has data from iris, face, and voice belonging to 105 subjects. The data were obtained using an Asus TPad TF 300T mobile device, and the images were captured using the rear camera of this device in 8 MP of resolution. The iris images were obtained at VIS and in two different illumination conditions varying eye orientations and occlusion levels. For each subject, 16 images (8 of each eye, cropped from an image of both eyes) were captured. The cropped images have a resolution of 300×200 pixels. Manual annotations of the iris and pupil contours are provided along with the database, but iris illumination noises are not identified.

The gb2sμMOD database is composed of 8,160 iris, face and hand videos belonging to 60 subjects and captured in three sessions with environment condition variation [207]. Sessions 1 and 2 were obtained in a controlled environment, while session 3 was acquired in an uncontrolled environment. The iris videos were recorded only in sessions 1 and 2 with a NIR camera (850 nm) held by the subject himself as close to the face as possible capturing both eyes. The diameter of

the iris in such videos is approximately 60 pixels. Ten iris videos were collected in two (5 in each session) for each one of the 60 subjects. Along with the videos, information such as name, ID card number, age, gender, and handedness are also available.

All databases described in this subsection contain iris and/or periocular subsets, however, some databases that do not have such subsets can also be employed for iris/periocular recognition. For example, the FRGC [210] database, which is a database of face images, has already been used for iris [120] and periocular [211, 115, 201] recognition in the literature.

3.3 OCULAR RECOGNITION COMPETITIONS

In this section, we describe the major recent competitions and the algorithms that achieved the best results in iris and/or periocular region information. Through these competitions, it is possible to demonstrate the advancement in terms of methodologies for ocular biometrics and also the current challenges in this research area.

The competitions usually provide a database in which the competitors must perform their experiments and submit their algorithms. Once submitted, the algorithms are evaluated with another subset of the database, according to the metrics established by the competition protocol. In this way, it is possible to fairly assess the performance of different methodologies for specific objectives.

In ocular biometrics including iris and periocular recognition, there are several competitions aimed at evaluating different situations, such as recognition in images captured at NIR and/or VIS wavelengths, images captured in an uncontrolled environment, images obtained with mobile devices, among others. For each competition, we describe the approaches that achieved the best results using fused information from iris and periocular region, and also the best performing methodologies using only iris information. Table 3.4 presents the main competitions held in recent years and the best results achieved, while Table 3.5 details the methodologies that obtained the best results in these competitions.

Table 3.4: Best results achieved in ocular biometric competitions. Extracted from [13].

Competition	Year	Database	Wavelength	Best Result	Traits
NICE.II [52]	2010	portion of UBIRIS v2	VIS	DI = 2.57 [119]	Iris + Periocular
NICE.II [52]	2010	portion of UBIRIS v2	VIS	DI = 1.82 [212]	Iris
MICHE-II [27]	2016	MICHE-I and MICHE-II	VIS	AVG = 1.00 [121, 122]	Iris + Periocular
MICHE-II [27]	2016	MICHE-I and MICHE-II	VIS	AVG = 0.86 [213]	Iris
MIR [214]	2016	MIR-Train and MIR-Test	NIR	FNMR4 = 2.24%, EER = 1.41% e DI = 3.33 [214]	Iris
VISOB 1.0 [14]	2016	VISOB	VIS	EER = 0.06% - 0.20% [110]	Periocular
CROSS-EYED [18]	2016	CROSS-EYED	CROSS	GF2 = 0.00% and EER = 0.29% (HH_1) [18]	Periocular
CROSS-EYED [18]	2016	CROSS-EYED	CROSS	GF2 = 3.31% and EER = 2.78% ($NTNU_6$) [18]	Iris
2^{nd} CROSS-EYED [19]	2017	CROSS-EYED	CROSS	GF2 = 0.00% and EER = 0.05% ($NTNU_4$) [19]	Iris
2^{nd} CROSS-EYED [19]	2017	CROSS-EYED	CROSS	GF2 = 0.74% and EER = 0.82% (HH_1) [19]	Periocular
VISOB 2.0 [28]	2020	VISOB 2.0	VIS	EER = 5.25% and AUC = 98.8% [24]	Periocular

Table 3.5: Best methodologies in ocular biometric competitions. Extracted from [13].

Contest/Author	Periocular Features	Iris Features	Periocular Matching	Iris Matching	Fusion Technique
NICE.II [119]	Texton histogram and Semantic information	Ordinal measures and color histogram	chi-square distance and exclusive or	SOBoost and diffusion distance	Sum rule
NICE.II [212]	-	2D Gabor	-	AdaBoost learning	-
MICHE-II [121, 122]	Multi-Block Transitional Local Binary Pattern (MB-TLBP)	1D Log-Gabor filter	chi-square distance	Hamming distance	Weighted sum of scores
MICHE-II [213]	-	Deep sparse filters	-	Maximized likelihood in a collaborative subspace representation	-
MIR [214]	-	Gabor wavelet	-	Cosine distance and hamming distance	-
VISOB 1.0 [110]	Maximum Response (MR) filters	-	DNN based on deeply coupled autoencoders	-	-
CROSS-EYED HH_1 [18]	SAFE, GABOR, SIFT, LBP and HOG	-	Probabilistic bayesian	-	-
CROSS-EYED $NTNU_6$ [18]	-	M-BSIF	-	chi-square distance and SVM	-
2^{nd} CROSS-EYED $NTNU_4$ [19]	-	M-BSIF	-	chi-square distance	-
2^{nd} CROSS-EYED HH_1 [19]	SAFE, GABOR, SIFT, LBP and HOG	-	Probabilistic bayesian	-	-
VISOB 2.0 [24]	ResNet-50	-	Cosine Distance	-	-

3.3.1 NICE - Noisy Iris Challenge Evaluation

The Noisy Iris Challenge Evaluation (NICE) competition contains two different contests. In the first one (NICE.I), held in 2008, the goal was the evaluation of methods for iris segmentation to remove noise factors such as specular reflections and occlusions. Regarding the evaluation of encoding and matching methods, the second competition (NICE.II), was carried out in 2010. The databases used in both competitions are subsets of UBIRIS.v2 [1], which contains VIS ocular images captured under uncontrolled environments.

Described by Proença and Alexandre [52], the first competition aimed to answer: "is it possible to automatically segment a small target as the iris in unconstrained data (obtained in a non-cooperative environment)?" In total, 97 research laboratories from 22 countries participate in the competition. The training set consisted of 500 images, and their respective manually generated binary iris masks. The committee evaluated the proposed approaches using another 500 images through a pixel-to-pixel comparison between the original and the generated segmentation masks. As a metric, the organizers choose the following error rate based on pixel-level:

$$E_j = \frac{1}{nwh} \sum_{i=1}^{n} \sum_{r=1}^{h} \sum_{c=1}^{w} P_i(r,c) \otimes G_i(r,c)\,, \tag{3.1}$$

where n refers to the number of test images, w and h are respectively the width and height of these images, $P_i(r,c)$ means the intensity of the pixel on row r and column c of the ith segmentation mask, $G_i(r,c)$ is the actual pixel value and $\otimes$ is the or-exclusive operator.

According to the values of E_j, NICE.I's best results are the following: 0.0131 [183], 0.0162 [215], 0.0180 [216], 0.0224 [217], 0.0282 [218], 0.0297 [219], 0.0301 [220], 0.0305 [221].

The second competition (NICE.II) evaluated only the feature extraction and matching results. Therefore, all the participants used the same segmented images, which were generated by the winner methodology in the NICE.I contest [52], proposed by Tan et al. [183]. The main goal was to investigate the impact of noise presented inside the iris region in the biometric recognition process. As described in both competitions [52], these noise factors have different sources, e.g., specular reflection and occlusion, caused by the uncontrolled environment where the images were taken. This competition received algorithms sent by 67 participants from 30 countries. The training set consists of 1,000 images and their respective binary masks. The proposed methods had to receive a pair of images followed by their masks as input and generate an output file containing the dissimilarity scores (d) of which pairwise comparison with the following conditions:

1. $d(I,I) = 0$
2. $d(I_1,I_2) = 0 \Rightarrow I_1 = I_2$
3. $d(I_1,I_2) + d(I_2,I_3) \geq d(I_1,I_3)$.

The submitted approaches were evaluated using a new set of 1,000 images with their binary masks. Consider $IM = \{I_1, ..., I_n\}$ as a collection of iris images, $MA = \{M_1, ..., M_n\}$ as their respective masks, and $id(.)$ representing a function that identifies an image. The comparison protocol one-against-all returns a match set $D^I = \{d_1^i, ..., d_m^i\}$ and a non-match set $D^E = d_1^e, ..., d_k^e\}$ of dissimilarity scores, where $id(I_i) = id(I_j)$ and $id(I_i) \neq id(I_j)$, respectively.

The best results of NICE.II ranked by their d' scores are as follows: 2.5748 [119], 1.8213 [212], 1.7786 [222], 1.6398 [223], 1.4758 [224], 1.2565 [225], 1.1892 [226], 1.0931 [227].

The winner method, proposed by Tan et al. [119], achieved a decidability value of 2.5748 by fusing iris and periocular features. The fusion process was performed at the score level by the sum rule method. Therefore, for iris and periocular images, different features and matching techniques were used. The iris features were extracted with ordinal measures and color histogram and for the periocular ones, texton histogram, and semantic information. To compute the matching scores, the authors employed the following metrics: SOBoost learning, diffusion distance, chi-square distance, and exclusive OR operator.

Wang et al. [212] proposed a method using only iris information. Their approach was ranked second in the competition, achieving a decidability value of 1.8213. The algorithm performed the segmentation and normalization of iris using the Daugman technique [57]. Features were extracted by applying the Gabor filters from different patches generated from the normalized image. The AdaBoost algorithm computed a selection of features and the similarity.

The main contribution of NICE competitions was the evaluation of iris segmentation and recognition methods independently, as several iris segmentation methodologies were evaluated in the first competition and the best one was applied to generate the binary masks used in the second one, in which the recognition task was evaluated. Hence, the approaches described in both competitions can be fairly compared since they employed the same images for training and testing.

Although NICE.II was intended to evaluate iris recognition systems, some approaches using information from the periocular region were also included in the final ranking. The winning method fused iris and periocular information, however, it should be noted that some approaches that also fused these two traits achieved lower results than methodologies that used only iris features. Moreover, it would be interesting to analyze the best performing approaches in the NICE.II competition in larger databases to verify the scalability of the proposed methodologies, as the database used in these competitions was not composed of a large number of images/classes.

Some recent works applying deep CNN models have achieved state-of-the-art results in the NICE.II database using information from the iris [23], periocular region [46] and fusing iris/periocular traits [48] with decidability values of 2.25, 3.47, 3.45, respectively.

3.3.2 MICHE - Mobile Iris Challenge Evaluation

In order to assess the performance that can be reached in iris recognition without the use of special equipment, the Mobile Iris CHallenge Evaluation II, or simply MICHE-II competition, was held [27]. The MICHE-I database, introduced by De Marsico et al. [15] has 3,732 images taken by mobile devices and was made available to the participants to train their algorithms, while other images obtained in the same way were employed for the evaluation.

Similarly to NICE.I and NICE.II, MICHE is also divided into two phases. MICHE.I and MICHE.II focused on iris segmentation and recognition, respectively. Ensuring a fair assessment and targeting only the recognition step, all MICHE.II participants used the segmentation algorithm proposed by Haindl and Krupicka [228], which achieved the best performance on MICHE.I.

The performance of each algorithm was evaluated through dissimilarity. Assuming I as a set of the MICHE.II database and that $I_a, I_b \in I$, the dissimilarity function D is defined by:

$$D(I_a, I_b) \Rightarrow [0, 1] \subset \mathbb{R}, \tag{3.2}$$

satisfying the following properties:

1. $D(I_a, I_a) = 0$
2. $D(I_a, I_b) = 0 \Rightarrow I_a = I_b$

3. $D(I_a, I_b) = D(I_b, I_a)$.

Two metrics were employed to assess the algorithms. The first, called RR, was used to evaluate the performance in the identification problem (1:N), while the second, called AUC, was applied to evaluate the performance in the verification problem (1:1). In addition, the methodologies were evaluated in two different setups: first comparing only images acquired by the same device and then using images obtained by two different devices (cross-sensor). The algorithms were ranked by the average performance of RR and AUC. The best results are listed in Table 3.6.

Table 3.6: Results of the MICHE.II competition. Average between RR and AUC. Adapted from [27].

Authors	All×All	GS4×GS4	Ip5×Ip5	Average
Ahmed et al. [121, 122]	0.99	1.00	1.00	1.00
Ahuja et al. [112, 229]	0.89	0.89	0.96	0.91
Raja et al. [213]	0.82	0.95	0.83	0.86
Abate et al. [230, 231]	0.79	0.82	0.88	0.83
Galdi and Dugelay [232, 233]	0.77	0.78	0.92	0.82
Aginako et al. [234, 235]	0.78	0.80	0.78	0.79
Aginako et al. [236, 237]	0.75	0.72	0.77	0.75

Ahmed et al. [121, 122] proposed the algorithm that achieved the best result. Their methodology performs the matching of the iris and the periocular region separately and combines the final score values of each approach. For the iris, they used the rubber sheet model normalization proposed by Daugman [57]. Then, the iris codes were generated from the normalized images with the 1-D Log-Gabor filter. The matching was computed with the Hamming distance. Using only iris information, an EER of 2.12% was reached. Features from the periocular region were extracted with Multi-Block Transitional Local Binary Patterns and the matching was computed with the chi-square distance. With features from the periocular region, an EER of 2.74% was reported. The outputs of both modalities (iris and periocular) were normalized with z-score and combined with weighted scores. The weights used for the fusion were 0.55 for the iris and 0.45 for the periocular region, yielding an EER of 1.22% and an average between RR and AUC of 1.00.

The best performing approach using only iris information was proposed by Raja et al. [213]. In their method, the iris region was first located through a segmentation method proposed by Raja et al. [194] and then normalized using the rubber sheet expansion model [4]. Each image band (red, green and blue) was divided into several blocks. The features were extracted from these blocks, as well as from the entire image, using a set of deep sparse filters, resulting in deep sparse histograms. The histograms of each block and each band were concatenated with the histogram of the entire image, forming the vector of iris features. The features extracted were used to learn a collaborative subspace, which was employed for matching. This algorithm achieved the third place in the competition, with an average between RR and AUC of 0.86 and with EER values of 0% in the images obtained by the iPhone 5S and 6.55% in the images obtained by Samsung S4.

This competition was the first to evaluate iris recognition using images captured by mobile devices and also to evaluate methodologies applied to the cross-sensor problem, i.e., to recognize images acquired by different sensors.

As in the NICE.II competition, one issue is the scalability evaluation of the evaluated approaches. Although the reported results are very promising, we have to consider them as preliminary since the test set used for the assessment is very small, containing only 120 images.

As expected, the best results were attained using iris and periocular region information, however, some approaches that used only iris information achieved better results than others that fused iris and periocular region information.

3.3.3 MIR - Competition on Mobile Iris Recognition

The BTAS Competition on Mobile Iris Recognition (MIR2016) was proposed to raise the state of the art of iris recognition algorithms on mobile devices under NIR illumination [214]. Five algorithms, submitted by two participants, were eligible for the evaluation.

A database (MIR-Train) was made available for training the algorithms and a second database (MIR-Test) was used for the evaluation. Both databases were collected under NIR illumination. The images of the two irises were collected simultaneously under an indoor environment. Three sets of images were obtained, with distances of 20 cm, 25 cm, and 30 cm, and 10 images for each distance. The images from both databases were collected in the same session. The MIR-Train database is composed of 4,500 images from 150 subjects, while MIR-Test has 12,000 images from 400 subjects. All images are grayscale with a resolution of 1968×1024 pixels. The main sources of intra-class variation in the images are due to variations in lighting, eyeglasses and specular reflections, defocus, distance changes, and others. Differently from NICE.II, the segmentation masks were not provided in MIR2016, thus, the methodologies submitted included iris detection, segmentation, feature extraction, and matching.

For the evaluation, the organizing committee considered that the left and right irises belong to the same class; thus, a fusion of the matching scores of both irises was performed. All possible intra-class comparisons (i.e., irises from the same subjects) were implemented to compute the False Non-Match Rate (FNMR). From each iris class, two samples were randomly selected to calculate the False Match Rate (FMR). In total, 174,000 intra-class and 319,200 inter-class matches were used. In cases where intra- or inter-class comparisons could not be performed due to failure enrollment or failure match, a random value between 0 and 1 was assigned to the score. The classification of the participants was performed using the FNMR4 metric, but the EER and DI metrics were also reported. The FNMR4 metric reports the FNMR value when the FMR equals to 0.0001. The EER is the value when FNMR is equal to the FMR, and the DI value is the decidability index, as explained previously.

The best result was from the Beijing Bata Technology Co. Ltd. reporting FNMR4 = 2.24%, EER = 1.41% and DI = 3.33. The methodology, described in [214], includes four steps: iris detection, preprocessing, feature extraction, and matching. For iris detection, the face is found using the AdaBoost algorithm [238] and eye positions are found by using SVM. Next, to lessen the effect of light reflections, the irises and pupils are detected by the modified Daugmans Integro-Differential operator [4]. In pre-processing, reflection regions are located and then removed using a threshold and shape information. Afterward, the iris region is normalized using the method proposed by Daugman [57]. Eyelashes are also detected and removed using a threshold. An improvement in image quality is achieved through histogram equalization. The features were extracted with Gabor wavelet, while Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) were applied for dimensionality reduction. The matching was performed using the cosine and Hamming distances, and the results combined.

The second place was achieved by TigerIT Bangladesh Ltd. with FNMR4 = 7.07%, EER = 1.29% and DI = 3.94. The proposed approach also made improvements in image quality through histogram equalization and smoothing. After pre-processing, the iris was normalized using the rubber sheet model [117]. Features were then extracted with 2D Gabor wavelets, while the matching was performed employing the Hamming distance. This methodology was classified

in second place since it obtained a higher FNMR4 value than the first one, but the EER and DI values were better than those reported by the winning algorithm of the competition.

The MIR2016's main contribution is to be the first competition using NIR images acquired by mobile modules, in addition to the construction of a new database containing images from both eyes of each individual. Unfortunately, the competition did not have many participants and the proposed methodologies consist only of classical literature techniques.

3.3.4 VISOB - Competition on Mobile Ocular Biometric Recognition

The VISOB 1.0 competition was designed to evaluate ocular biometric recognition methodologies using images obtained from mobile devices in visible light on a large-scale database. The database created and used for the competition was VISOB (VISOB Database ICIP2016 Challenge Version) [14]. This database has 158,136 images from 550 subjects, and is the database of images obtained from mobile devices with the largest number of subjects. The images were captured by 3 different devices (iPhone 5S, Oppo N1 and Samsung Note 4) under 3 different lighting classes: 'daylight', 'office', and 'dim light'. Four different research groups participated in the competition and 5 algorithms were submitted. The metric used to assess the performance of the algorithms was EER.

In almost all competitions, participants submit an algorithm already trained and the evaluation is performed on an unknown portion of the database. On the other hand, VISOB 1.0 competitors submitted an algorithm that was trained and tested on an unknown portion of the database. Two different evaluations were carried out. In the first one (see Table 3.7), the algorithms were trained (enrollment) and tested for each device and type of illumination.

Table 3.7: EER (%) rank by device and lighting condition. Adapted from [14].

Day light			
Method	iPhone 5S	Oppo N1	Samsung Note 4
NTNU-1 [110]	0.06	0.10	0.07
NTNU-2 [111]	0.40	0.43	0.33
ANU	7.67	7.91	8.42
IIITG [112]	18.98	18.12	15.98
Anonymous	38.09	38.29	62.23
Office			
NTNU-1 [110]	0.06	0.04	0.05
NTNU-2 [111]	0.48	0.63	0.49
ANU	10.36	16.01	9.10
IIITG [112]	19.29	19.79	18.65
Anonymous	35.26	31.69	72.84
Dim light			
NTNU-1 [110]	0.06	0.07	0.07
NTNU-2 [111]	0.45	0.16	0.16
ANU	8.44	9.02	11.89
IIITG [112]	17.54	19.49	23.25
Anonymous	31.06	34.00	67.20

In the second evaluation, the algorithms were trained only with the images from the 'office' lighting class for each of the 3 devices. To assess the effect of illumination on ocular recognition, the tests were performed with the 3 types of illumination for each device. The results are shown in Table 3.8.

Raghavendra and Busch [110] achieved an EER between 0.06% and 0.20% in all assessments, obtaining the best result of the competition. The proposed approach extracted

Table 3.8: EER (%) rank by device and lighting condition. The algorithms were trained only with the 'office' lighting class (O) and tested on all the others. Table adapted from [14].

iPhone 5S			
Method	O-O	O-Day	O-Dim
NTNU-1 [110]	0.06	0.13	0.20
NTNU-2 [111]	0.48	1.82	1.45
ANU	10.36	11.03	16.64
IIITG [112]	19.29	32.93	45.34
Anonymous	35.26	28.67	42.29
Oppo N1			
NTNU-1 [110]	0.04	0.10	0.09
NTNU-2 [111]	0.63	1.90	3.34
ANU	16.01	14.75	18.24
IIITG [112]	19.79	38.24	42.59
Anonymous	31.69	31.21	37.17
Samsung Note 4			
NTNU-1 [110]	0.05	0.13	0.10
NTNU-2 [111]	0.49	2.50	4.25
ANU	9.10	13.69	19.57
IIITG [112]	18.65	34.29	40.21
Anonymous	27.73	24.33	50.74

periocular features using Maximum Response (MR) filters from a bank containing 38 filters, and a deep neural network learned with a regularized stacked autoencoders [110]. For noise removal, the authors applied a Gaussian filter and performed histogram equalization and image resizing. Finally, the classification was performed through a deep neural network based on deeply coupled autoencoders.

All participants explored features based on the texture of the eye images, extracted from the periocular region. None of the submitted algorithms extracted features only from the iris. The organizing committee compared the performance of the algorithms using images obtained only by the same devices, that is, the algorithms were not trained and tested on images from different devices (cross-sensor). Thus, the main contributions of this competition were a large database containing images from different sensors and environments, along with the assessments on these different setups.

The second edition of this competition, called VISOB 2.0, was carried out at IEEE WCCI in 2020 [28]. A new VISOB's subset with eye images from 250 subjects captured by two mobile devices: Samsung Note 4 and Oppo N1, was employed to compare the submitted approaches. This competition evaluated ocular biometrics recognition methods using stacks of five images in the open-world (subject-independent) protocol in different lighting conditions: Dark, Office, and Daylight. In the development (training) stage, the competitors were provided with stacks of images from 150 subjects. Regarding the subject-independent evaluation, the comparison of the submitted methods was performed employing samples from other 100 subjects that were not available in the training stage. The main idea of using multi-frame (stacks) captures for ocular biometrics is to avoid degradation in the images caused by variations in illumination, noise, blur, and user to camera distance. Two participants submitted algorithms based on deep representations and one based on hand-crafted features. Table 3.9 presents the results.

The rank 1 algorithm proposed by Zanlorensi et al. [24] (UFPR) consists of an ensemble of ResNet-50 models (5 models, one for each image in the stack) pre-trained for face-recognition using the VGG-Face database. The authors had previously employed this method for cross-

Table 3.9: EER (%) rank by device and lighting condition: Dark (DK), Daylight (DL), and Office (O). Table adapted from [28].

Samsung Note 4									
Method	DK-DK	DK-DL	DK-O	DL-DK	DL-DL	DL-O	O-DK	O-DL	O-O
UFPR [24]	7.46	10.03	6.66	11.46	7.76	6.72	12.10	8.06	5.26
Bennett University	35.01	40.47	42.15	41.45	30.68	34.40	43.65	34.31	27.05
Anonymous	42.07	44.69	43.44	44.41	40.69	42.51	46.09	42.69	39.77
Oppo N1									
	DK-DK	DK-DL	DK-O	DL-DK	DL-DL	DL-O	O-DK	O-DL	O-O
UFPR [24]	6.39	9.40	8.08	8.28	8.11	6.67	9.76	8.65	6.49
Bennett University	34.33	40.36	40.90	41.99	29.70	31.91	42.95	31.79	26.21
Anonymous	40.30	44.94	43.71	45.41	42.46	45.14	46.68	45.70	42.05

spectral ocular recognition achieving state-of-the-art results on the CROSS-EYED and the PolyU Cross-Spectral databases using iris and periocular traits. In this method, each ResNet-50 model was fine-tuned using the periocular images from VISOB 2.0. The only modification in the model was the addition of a fully connected layer containing 256 neurons at the top to reduce the feature dimensionality. The training was computed in the identification mode using a Softmax cross-entropy loss function as a prediction layer. Then, in the evaluation, the prediction layer was removed, and the final combined feature vector with a size of 1280 (5×256) was used to match samples by computing the cosine distance similarity. This algorithm's best result was 5.26% of EER using images in the Office vs. Office lighting condition.

The second-place method (*Bennet University*) used Directional Threshold Local Binary Pattern (DTLBP), and wavelet transform for feature extraction (handcrafted features). Then, the Chi-square distance was employed to compute the similarity between the stack of images. This method's best result was 26.21% of EER in the Office vs. Office lighting condition. Finally, the third approach employed the GoogleNet model pre-trained in the ImageNet database for feature extraction and euclidean distance to compute the similarity between the pairs of images. A Long Short Term Memory (LSTM) model using the euclidean distance scores as input was used to predict whether the pair of images is from the same subject or not. This method's best result was 39.77 of EER in the Office vs. Office lighting condition.

To the best of our knowledge, VISOB 2.0 was the first competition to use multi-frame ocular recognition. The results show that comparison across different illumination was the most difficult for all methods. The open-world (subject-independent) protocol is a realistic scenario for applications in environments without restriction and prior knowledge of the subjects. Finally, the submitted algorithms' performance shows that there is still room for improvement in this area.

3.3.5 Cross-Eyed - Cross-Spectral Iris/Periocular Competition

The first Cross-Eyed competition was held in 2016 at the 8th IEEE International Conference on Biometrics: Theory, Applications, and Systems (BTAS). The aim of the competition was the evaluation of iris and periocular recognition algorithms using images captured at different wavelengths. The CROSS-EYED database [18, 19], employed in the competition, has iris and periocular images obtained simultaneously at the VIS and NIR wavelengths.

Iris and periocular recognition were evaluated separately. To avoid the use of iris information in the periocular evaluation, a mask excluding the entire iris region was applied. Six algorithms submitted by 2 participants, named **HH** from Halmstad University and **NTNU** from

Norway Biometrics Laboratory, qualified. The final evaluation was carried out with another set of images, containing 632 images from 80 subjects for periocular recognition and 1,280 images from 160 subjects for iris recognition.

The evaluation consisted of enrollment and template matching of intra-class (all NIR against all VIS images) and inter-class comparisons (3 NIR against 3 VIS images – per class). A metric based on Generalized False Accept Rate (GFAR) and Generalized False Reject Rate (GFFR) was used to verify the performance of the submitted algorithms. These metrics generalize the FMR and the FNMR, including Failure-to-enroll (FTE) and Failure-to-acquire (FTA). Finally, to compare the algorithms, the GF2 metric (GFRR@GFAR = 0.01) was employed.

Halmstad University (HH) team submitted 3 algorithms. The approaches consist of fusing features extracted with Symmetry Patterns (SAFE), Gabor Spectral Decomposition (GABOR), Scale-Invariant Feature Transform (SIFT), Local Binary Patterns (LBP) and Histogram of Oriented Gradients (HOG). These fusions were evaluated combining scores from images obtained by the same sensors and also by different sensors. The evaluated algorithms differ by the fusion of different features: HH_1 fusing all the features; HH_2 fusing SAFE, GABOR, LBP and HOG; and HH_3 fusing GABOR, LBP and HOG. The algorithms were applied only to periocular recognition, and the best performance was achieved by HH_1, which achieved an EER of 0.29% and GF2 of 0.00%. More details can be found in [18].

The Norwegian Biometrics Laboratory (NTNU) also submitted 3 algorithms, which applied the same approaches for feature extraction from iris and periocular traits. The iris region was located using a technique based on the approach proposed by Raja et. al. [194], and features were extracted through histograms resulting from the multi-scale BSIF, a bank of independent binarized statistical filters. These histograms were compared using the Chi-Square distance metric. Lastly, an SVM was employed to obtain the fusion and scores corresponding to each filter. The best approach achieved EER of 4.84% and GF2 of 14.43% in periocular matching, and EER of 2.78% and GF2 of 3.31% in iris matching.

In 2017, the second edition of this competition was held [19]. Similarly to the first competition, the submitted approaches were ranked by EER and GF2 values. Comparisons in periocular images were made separately for each eye, i. e., the left eyes were compared only with left eyes, and the same for the right eyes. The main difference was in the database used, as the training set consisted of the CROSS-EYED database and the test set was made with 55 subjects. As in the first competition, the matching protocol consisted of intra- and inter-class comparisons, in which all intra-class comparisons were performed and only 3 random images per class were applied in the inter-class comparisons. Results and methodologies of 4 participants were reported, being 4 participants with 11 algorithms for periocular recognition, and 1 participant with 4 algorithms for iris recognition. Two of these participants took part in the first competition, Halmstad University (HH) and Norwegian Biometrics Laboratory (NTNU). The other three competitors were IDIAP from Switzerland, IIT Indore from India, and an anonymous.

The best method using periocular information was submitted by HH_1, which fused features based on SAFE, GABOR, SIFT, LBP and HOG. Their approach, similar to the one proposed in the first competition, reached EER and GF2 values of 0.82% and 0.74%, respectively. For iris recognition, the best results were attained by $NTNU_4$, which was based on BSIF features and reported EER and GF2 values of 0.05% and 0.00%, respectively.

We point out two main contributions of these competitions: (i) the release of a new cross-spectral database, and (ii) the evaluation of several approaches using iris and periocular traits with some promising strategies that can be applied for cross-spectral ocular recognition. Nevertheless, we also highlight some problems in their evaluation protocols. First, the periocular evaluation in the second competition only matches left eyes against left eyes and right eyes against

right eyes using prior knowledge of the database. Another problem is the comparison protocol, which uses only 3 images per class in inter-class comparisons instead of all images without specifically reporting which ones were used. There is also no information on code availability, and details of the methodologies are lacking, limiting the reproducibility.

3.4 DEEP LEARNING APPROACHES FOR OCULAR RECOGNITION

Recently, deep learning approaches have won many machine learning competitions, even achieving superhuman visual results in some domains [123]. Therefore, in this section, we describe recent works that applied deep learning-based techniques to ocular biometrics including iris, periocular and sclera recognition, gender and age classification, and subject-independent recognition.

3.4.1 Iris Recognition

Liu et al. [90] presented one of the first works applying deep learning to iris recognition. Their approach, called DeepIris, was created for recognizing heterogeneous irises captured by different sensors. The proposed method was based on a CNN model with a bank of Pairwise filters, which learns the similarity between a pair of images. The evaluation in verification protocol was carried out in the Q-FIRE and CASIA cross-sensor databases and reported promising results with EER of 0.15% and 0.31%, respectively.

Gangwar and Joshi [91] also developed a deep learning method for iris verification on the cross-sensor scenario, called DeepIrisNet. They presented two CNN architectures for extracting iris representations and evaluated them using images from the ND-IRIS-0405 and ND Cross-Sensor-Iris-2013 databases. The first model was composed of 8 standard convolutional, 8 normalization, and 2 dropout layers. The second one, on the other hand, has inception layers [153] and consists of 5 convolutional layers, 7 normalization layers, 2 inception layers, and 2 dropout layers. Compared to the baselines, their methodology reported better robustness on different factors such as the quality of segmentation, rotation, and input, training, and network sizes.

To demonstrate that generic descriptors can generate discriminant iris features, Nguyen et al. [93] applied distinct deep learning architectures to NIR databases obtained in controlled environments. They evaluated the following CNN models pre-trained using images from the ImageNet database [131]: AlexNet, VGG, Inception, ResNet and DenseNet. Iris representations were extracted from normalized images at different depths of each CNN architecture, and a multi-class SVM classifier was employed for the identification task. Although no fine-tuning process was performed, interesting results were reported in the LG2200 (ND Cross-Sensor-Iris-2013) and CASIA-IrisV4-Thousand databases. In their experiments, the representations extracted from intermediate layers of the networks reported better results than the representations from deeper layers.

The method proposed by Al-Waisy et al. [92] used left and right irises information for the identification task. In this approach, each iris was first detected and normalized, and then features were extracted and matched. Finally, the left and right irises matching scores were fused. Several CNN configurations and architectures were evaluated during the training phase and, based on a validation set, the best one was chosen. The authors also evaluated other training strategies such as dropout and data augmentation. Experiments carried out on three databases (i.e., SDUMLA-HMT, CASIA-IrisV3-Interval, and IIT Delhi Iris) reported a 100% rank-1 recognition rate in all of them.

Generally, an iris recognition system has several preprocessing steps, including segmentation and normalization (using Daugman's approach [57]). In this context, Zanlorensi et

al. [23] analyzed the impact of these steps when extracting deep representations from iris images. Applying deep representations extracted from an iris bounding box without both segmentation and normalization processes, they reported better results compared to those obtained using normalized and segmented images. The authors also fine-tuned two pre-trained models for face recognition (i.e., VGG-16 and ResNet50) and proposed a data augmentation technique by rotating the iris bounding boxes. In their experiments, using only iris information, an EER of 13.98% (i.e., state-of-the-art results) was reached in the NICE.II database.

As the performance of many iris recognition systems is related to the quality of detection and segmentation of the iris, Proença and Neves [34] proposed a robust method for inaccurately segmented images. Their approach consisted of corresponding iris patches between pairs of images using a CNN model, which estimates the probability that two patches belong to the same biological region. According to the authors, the comparison of these patches can also be performed in cases of bad segmentation and non-linear deformations caused by pupil constriction/dilation. The following databases were used in the experiments: CASIA-IrisV3-Lamp, CASIA-IrisV4-Lamp, CASIA-IrisV4-Thousand, and WVU. The authors reported results using good quality data as well as data with severe segmentation errors. Using accurately segmented data, they achieved EER values of 0.6% (CASIA-IrisV3-Lamp), 2.6% (CASIA-IrisV4-Lamp), 3.0% (CASIA-IrisV4-Thousand) and 4.2% (WVU).

The methodology proposed in [94] does not require preprocessing steps, such as iris segmentation and normalization, for iris verification. In this approach, which is based on deep learning models, the authors used biologically corresponding patches to discriminate genuine and impostor comparisons in pairs of iris images, similarly to IRINA [34]. These patches were learned in the normalized iris images and then remapped into a polar coordinate system. In this way, only a detected/cropped iris bounding box is required in the matching stage. State-of-the-art results were reported in three NIR databases, achieving EER values of 0.6%, 3.0%, and 6.3% in the CASIA-Iris-V4-Lamp, CASIA-IrisV4-Thousand, and WVU, respectively.

In [29], Wang and Kumar claimed that iris features extracted from CNN models are generally sparse and can be used for template compression. In the cross-spectral scenario, the authors evaluated several hashing algorithms to reduce the size of iris templates, reporting that the supervised discrete hashing was the most effective in terms of size and matching. Features were extracted from normalized iris images with some deep learning architectures, e.g., CNN with softmax cross-entropy loss, Siamese network, and Triplet network. Promising results were reported by incorporating supervised discrete hashing on the deep representations extracted with a CNN model trained with a softmax cross-entropy loss. The proposed methodology was evaluated on a cross-spectral scenario and achieved EER values of 12.41% and 6.34% on the PolyU Cross-Spectral and CROSS-EYED databases, respectively.

Zanlorensi et al. [24] performed extensive experiments in the cross-spectral scenario applying two CNN models: ResNet-50 [85] and VGG16 [84]. Both models were first pre-trained for face recognition and then fine-tuned using periocular and iris images. The results of the experiments, carried out in two databases: CROSS-EYED and PolyU Cross-Spectral, indicated that it is possible to apply a single CNN model to extract discriminant features from images captured at both NIR and VIS wavelengths. The authors also evaluate the impact of representation extraction at different depths from the ResNet-50 model and the use of different weights for fusing iris and periocular features. For the verification task, their approach achieved state-of-the-art results in both databases on intra- and cross-spectral scenarios using iris, periocular, and fused features.

3.4.2 Periocular Recognition

Luz et al. [46] designed a biometric system for the periocular region employing the VGG-16 model [84]. Promising results were reported by performing transfer learning from the face recognition domain and fine-tuning the system for periocular images. This model was compared to a model trained from scratch, showing that the proposed transfer learning and fine-tuning processes were crucial for obtaining state-of-the-art results. The evaluation was performed in the NICE.II and MobBIO databases, reporting EER values of 5.92% and 5.42%, respectively.

Using a similar methodology, Silva et al. [48] fused deep representations from iris and periocular regions by applying the Particle Swarm Optimization (PSO) to reduce the feature vector dimensionality. The experiments were performed in the NICE.II database and promising results were reported using only iris information and also fusing iris and periocular traits, reaching EER values of 14.56% and 5.55%, respectively.

Proença and Neves [51] demonstrated that periocular recognition performance can be optimized by first removing the iris and sclera regions. The proposed approach, called Deep-PRWIS, consists of a CNNs model that automatically defines the regions of interest in the periocular input image. The input images were generated by cropping the ocular region (iris and sclera) belonging to an individual and pasting the ocular area from another individual in this same region. They obtained state-of-the-art results (closed-world protocol) in the UBIRIS.v2 and FRGC databases, with EER values of 1.9% and 1.1%, respectively.

Zhao and Kumar [88] developed a CNN-based method for periocular verification. This method first detects eyebrow and eye regions using a Fully Convolutional Network (FCN) and then uses these traits as key regions of interest to extract features from the periocular images. The authors also developed a verification oriented loss function (Distance-driven Sigmoid Cross-entropy loss (DSC)). Promising results were reported on six databases both in closed- and open-world protocols, achieving EER values of 2.26% (UBIPr), 8.59% (FRGC), 7.68% (FOCS), 4.90% (CASIA-IrisV4-Distance), 0.14% (UBIRIS.v2) and 1.47% (VISOB).

Using NIR images acquired by mobile devices, Zhang et al. [89] developed a method based on CNN models to generate iris and periocular region features. A weighted concatenation fused these features. These weights and also the parameters of convolution filters were learned simultaneously. In this sense, the joint representation of both traits was optimized. They performed experiments in a subset of the CASIA-Iris-Mobile-V1.0 database reporting EER values of 1.13% (Periocular), 0.96% (Iris) and 0.60% (Fusion).

3.4.3 Sclera, Age, and Gender Recognition

In ocular biometrics using the sclera region, deep learning techniques are generally applied in the segmentation stage [43, 239, 240, 241], helping the recognition system by locating traits as the sclera itself and the iris. As described by Vitek et al. [242], the recognition is often performed using the segmented sclera vasculature by employing key-point and dense-grid descriptors as SIFT, SURF, ORB, and Dense SIFT. As the sclera is a relatively new ocular biometric trait, there are currently few deep learning-based approaches to perform person recognition [243, 244].

Regarding segmentation methods, Lucio et al. [43] proposed two approaches based on FCN and GAN to segment the sclera region. Experiments performed on two ocular databases demonstrated that the FCN model achieved better results on a single-sensor configuration. In contrast, for the cross-sensor scenario, the GAN model reached higher scores. Wang et al. [241] presented the ScleraSegNet, which is based on the U-Net model. The authors also proposed and compared different embed attention modules in the U-Net model regarding learning discriminative features. Extensive experiments using three ocular databases showed that the

channel-wise attention module was the most effective for performing the segmentation and that data augmentation techniques improved the generalization ability. Naqvi and Loh [240] proposed a model for sclera segmentation employing a residual encoder and decoder network, called Sclera-Net. The authors also addressed sclera segmentation in images acquired by different sensors achieving promising results in this work. Recent competitions on sclera segmentation [239, 245] demonstrated that deep learning-based methods achieved the highest results, mainly models based on the U-Net and FCN architectures. The results reached in these competitions show that sclera segmentation is still an open and challenging problem.

Regarding the sclera recognition task based on deep learning methods, one of the first approaches found in the literature is the ScleraNET [243]. In this work, the authors proposed a multi-task CNN model combining losses from the identity and gaze direction recognition. This model extracts vasculature descriptors and uses them to infer the identity of the subject. Promising results were achieved and compared with handcrafted-based methods. Maheshan et al. [244] also proposed a method based on CNN for sclera recognition. The model comprises four convolutional layers, followed by a max-pooling layer and a fully connected layer at the top. The proposed model was evaluated and compared with the top 2 ranked algorithms in the SSRBC 2016 Sclera Segmentation and Recognition Competition [246] reaching the higher scores.

Soft biometrics, such as gender and age classification, using ocular traits are tasks that have gained attention in research in recent years [53, 54, 55, 56, 17]. It can be used as primary biometric information to improve the accuracy of biometric systems [54]. A few works in the literature employ ocular traits (iris and periocular region) using VIS images for gender and age estimation/classification based on deep learning techniques [106, 55, 56, 107, 17].

Kuehlkanp and Bowyer [56] performed extensive experiments using hand-crafted and deep-representations with iris and periocular traits for gender classification. The results sustain that gender prediction using periocular images is at least 17% more accurate than normalized iris images, regardless of the classifier (hand-crafted or deep representations). Krishnan et al. [53] investigated the fairness of ocular biometrics methods using mobile images across gender. The evaluation employing the ResNet, LightCNN, and MobileNet models for periocular biometrics presented an equivalent verification performance for males and females. However, in gender classification, males outperformed females by a difference of 22.58%.

Rattani et al. [106] investigated age classification using VIS ocular images acquired by mobile devices. The proposed method consists of a 6-layer CNN model comprising convolution, max-pooling, batch-normalization, and fully connected layers. Ages were grouped into 8 ranges, and a soft-max activation was employed to compute each group's probability. Experiments conducted on a 5-fold cross-validation protocol using only the ocular region (both eyes, eyebrows, and periocular region) reported closer and promising results than full-face methods for age estimation, achieving an accuracy (%) of 46.97 ± 2.9 against 49.5 ± 4.4, respectively. Angeloni et al. [107] proposed a multi-stream CNN model using facial parts for age classification. The model consists of 4 streams, each one for the following traits: eyebrows, eyes, nose, and mouth. The proposed approach reached better results in accuracy than methods employing images from the entire face. Furthermore, an ablation study on the method reported that the eyes region was the most important trait to improve the entire approach accuracy.

In a recent work [17], the authors proposed a multi-task learning network for periocular recognition using VIS images acquired by mobile devices. The CNN architecture was composed of a MobileNetV2 as a base model and 5 fully connected layers followed by soft-max layers for the following soft biometrics tasks: identity, age, gender, eye side, and smartphone model classification. The proposed multi-task model reached better results than several CNN architectures for verification and identification tasks on experiments conducted on closed- and open-world

(subject-independent) protocols. Moreover, performing an ablation study, the authors stated that age, gender, and mobile device classification were critical components regarding the accuracy of the method for the identification task.

3.4.4 Final Remarks

Regarding the works described in this section, we point out that some deep learning-based approaches for iris recognition aim to develop end-to-end systems by removing preprocessing steps (e.g., segmentation and normalization) since a failure in such processes would probably affect recognition systems [23, 34, 94]. Several works [46, 48, 51, 88, 89] show that the periocular region contains discriminant features and can be used, or fused with iris and sclera information, to improve the performance of biometric systems. Furthermore, recent works on soft-biometrics for periocular recognition [53, 54, 55, 56, 17] reported promising results and stated that this kind of information can be used to improve the accuracy of the biometric system. Finally, biometric systems evaluated in the open-world setting are still a challenging task since it is highly affected by the intra- and inter-class variability, especially in VIS images collected in unconstrained scenarios. Some works [29, 17] evaluated the most employed CNN architectures for the verification task in the open-world setting. These approaches are generally based on Pairwise filters, Siamese, and Triplet networks. Regarding only these kinds of architectures, in [29], the Siamese model achieved better results than the Triplet network. On the other hand, in [17] the Pairwise filters network reached better results than the Siamese network. It is important to note that in both works [29, 17], even in the open-world setting, the best results for the verification task were achieved employing CNN models using a soft-max layer in the training stage.

For completeness, there are several works and applications with ocular images using deep learning frameworks, such as: spoofing and liveness detection [36, 37], left and right iris images recognition [38], contact lens detection [40], iris detection [42], sclera and iris segmentation [43, 44], iris and periocular region detection [41], gender classification [174, 39], iris/periocular biometrics by in-set analysis [35], iris recognition using capsule networks [47], and sensor model identification [45].

4 PROPOSED METHODS

This chapter describes the proposed methods organized into the following subjects: Iris recognition without preprocessing, Cross-spectral ocular biometrics, Attribute normalization for the periocular region, and the newest collected periocular database (UFPR-Periocular). We also detail all the experimental protocols and the databases employed for each experiment.

4.1 IRIS RECOGNITION WITHOUT PREPROCESSING

As described in the previous sections, a typical iris recognition system comprises the following stages: image acquisition, feature extraction, and matching. Before feature extraction and recognition, a preprocessing technique is applied, such as iris and pupil detection, segmentation for noise removal, and normalization using the rubber sheet model [57], as shown in Figure 4.1.

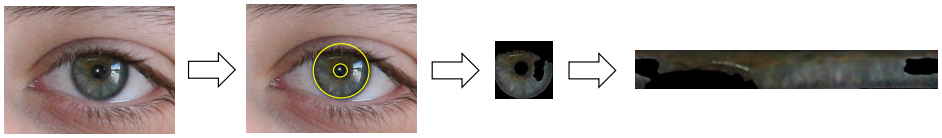


Figure 4.1: Preprocessing steps: (a) original image, (b) iris and pupil delineation, (c) iris segmentation for noise removal and (d) normalization.

The first step – the iris detection – is of paramount importance because an error in this stage will be propagated to the next steps (feature extraction and matching), degrading the biometric system's accuracy. Several works in the literature propose methodologies and approaches for iris detection [33, 4, 247, 248, 249, 250, 251, 252]. Considering that many applications use normalized iris images as input for feature representation, most iris detection methodologies are based on finding a circle to delimit iris and pupil regions. These methodologies usually have some parameters, such as min/max iris and pupil boundaries, which need to be tuned specifically for each database. Recently we have found methodologies developed for the detection of the iris bounding box, i.e., the smallest square/rectangle bounding box that encompasses the entire region of the iris [42].

The second step, i.e., the iris segmentation, consists of extracting the visible iris portion of the image. This process uses a variety of boundary and region detection, and active contour techniques [30]. The delineation process's output image contains only the annulus region (delimited by iris and pupil circles) of the iris. This image may have occlusion by eyelids and eyelashes and noises caused by reflections, glasses, angle, among others. In order to totally or partially remove these noises/occlusions, a segmentation approach is applied. Commonly, the approaches for NIR and VIS iris images segmentation are different. There are several works with different techniques in the literature, usually applied to segment non-ideal iris images [183, 228, 91]. An extensive and detailed survey on iris segmentation and detection is presented by Jan [253].

Finally, the delineated and segmented iris image is normalized using the rubber sheet model [57, 4, 30]. This normalization process consists of transforming the iris's circular region from the Cartesian space into a polar coordinate system resulting in a rectangular region.

Regarding image scale and iris deformation caused by pupil dilation or constriction (considering the uniform iris elasticity), normalization processes are applied to reduce these issues. After the preprocessing steps, features are extracted from the resultant image. Then, the final biometric task uses these features (e.g., matching, classification, verification). Nonetheless, errors may occur in the preprocessing, and these can degrade and compromise the effectiveness of a biometric system.

In our first investigation, we study the use of iris images without preprocessing to extract deep representations. For this, it was necessary to evaluate the impact of the preprocessing steps such as segmentation for noise removal and normalization in the recognition system's performance. The iris recognition system employed consists of three main stages: image preprocessing, feature extraction, and matching, as shown in Fig. 4.2.

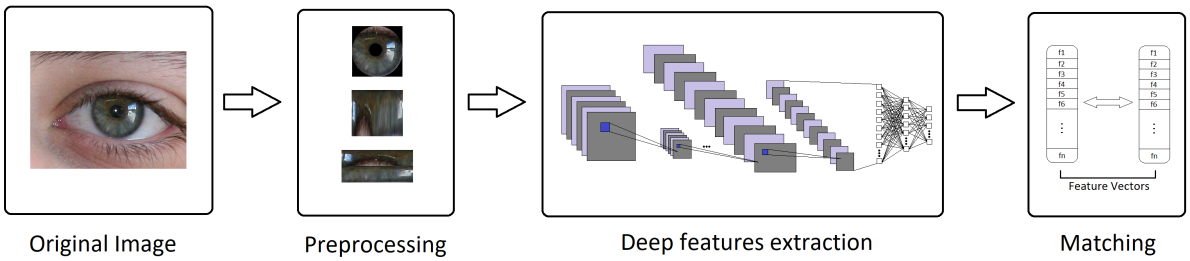


Figure 4.2: Ocular biometric system employed to evaluate the impact of iris preprocessing. Extracted from [23].

In the preprocessing stage, we applied segmentation and normalization techniques. We also employed a data augmentation technique to increase the number of training samples. The features were extracted using two CNN models, which were fine-tuned using the original images, and the images generated through data augmentation. Finally, the matching was performed using the cosine distance.

We performed the experiments using images from two databases: Nice.II [52] and CASIA-IrisV3-Interval [2]. These databases were chosen because they can represent the best and worst-case concerning the images' capture since the CASIA-IrisV3-Interval has images obtained in a controlled environment and the Nice.II was obtained in an uncontrolled environment.

For the Nice.II the official competition protocol was used, being $1,000$ images from 171 classes for training and $1,000$ images from 150 classes for the test. As the CASIA-IrisV3-Interval database does not have an official protocol, we split the database into two subsets based on the number of classes. In this way, the protocol consists of $1,383$ images from 197 classes for training (first 197 classes of the database) and $1,256$ images from 196 classes for the test. It is important to note that each class corresponds to one eye of the individual, i.e., the right and left eyes of the same individual corresponds to two distinct classes.

4.1.1 Image preprocessing

To fine-tuning CNN models trained in other domains without excluding any layers and using them as feature extractors, it was necessary to resize the input images to the CNN default input size. Depending on the aspect ratio of the image, the resizing process can generate distortions. Considering that the input size of the most common CNN models has a $1:1$ aspect ratio, normalized images are the ones with the most significant distortions. The distortion caused by resizing in normalized images is shown in Fig. 4.3.

We generated six different inputs from the original iris images to analyze the impact of the preprocessing. In the first input image scheme, irises were normalized with the standard

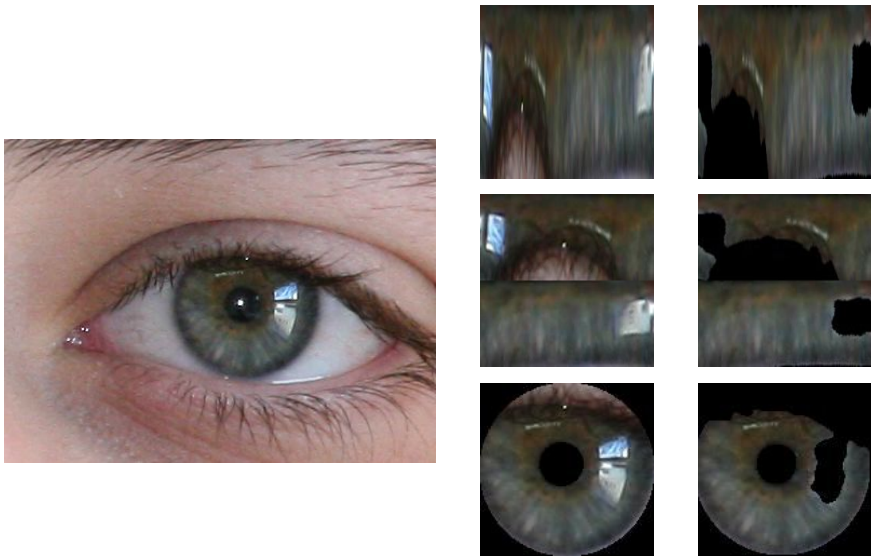


Figure 4.3: (a) Non-segmented and (b) segmented images for noise removal from Nice.II database. From top to bottom, it is shown 8:1, 4:2 aspect ratios and non-normalized images. Adapted from [23].

rubber sheet model [57] using an 8:1 aspect ratio. The second one was also normalized. However, from the standard 8:1 ratio, they were rearranged in a 4:2 ratio to employ less interpolation. In the third and last one, no normalization was performed, applying only the original iris image as input to the models. We also evaluated the impact of the segmentation technique for noise removal in all representations. When fine-tuning was applied, all the iris images used as inputs for the feature representation models, once normalized, were resized to 224×224 pixels.

The normalization through the rubber sheet model [57] aims to obtain invariance regarding size and pupil dilatation. In the NICE.II database, the main problem is the difference of the iris size due to distances in the image capture. On the other hand, the images from the CASIA-IrisV3-Interval database were obtained in a controlled environment. Thus there is no variation in distance, pupil dilation/contraction, or noises in the iris region.

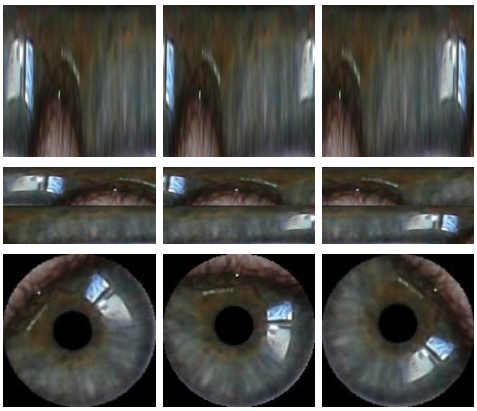
The segmentation process for noise removal was performed using the methodology proposed by Tan et al. [183], winner of NICE.I [52] and Gangwar et al. [254], for Nice.II and CASIA-IrisV3-Interval databases, respectively. It is important to note that in non-normalized images, an arc delimitation preprocessing (i.e., two circles, an outer and an inner), based on the iris mask, was used.

4.1.2 Data Augmentation

Since the training subset has only $1,000$ images belonging to 171 classes in Nice.II database and $1,383$ images from 197 classes in the CASIA-IrisV3-Interval database, it is essential to apply data augmentation techniques to increase the number of training samples. The fine-tuning process can result in a better generalization of the models with more images. In this sense, we rotate the original images at specific angles since we noted some slightly rotated images in the dataset.

The ranges of angles used were: $-15°$ to $15°$, $-30°$ to $30°$, $-45°$ to $45°$, $-60°$ to $60°$, $-90°$ to $90°$ and $-120°$ to $120°$. The rotation angles were proportional for each range, generating 4, 6, and 8 images for each original image, respectively. For example, considering the range $-60°$ to $60°$ with 6 angles, for each original image, another six were generated rotating $-60°$, $-40°$,

$-20°$, $20°$, $40°$ and $60°$. Performing the validation of all these data augmentation methods on all input images, we determined (based on accuracy and loss) that the best range was $-60°$ to $60°$ with 6 rotation angles. These parameters were applied to perform the data augmentation in the training set, totaling $7,000$ images in Nice.II database and $9,681$ images in CASIA-IrisV3-Interval database. Some samples generated by data augmentation can be seen in Fig. 4.4.



(a) (b) (c)

Figure 4.4: Data augmentation samples in Nice.II database: (a) $-45°$ rotated images, (b) original images and (c) $45°$ rotated images. Extracted from [23].

4.1.3 Feature Extraction and Matching

For feature extraction, the fine-tuning of VGG16 [134] and ResNet-50 [6] CNN models (architectures) trained for face recognition (VGG16 [84] and ResNet-50 [85]) were applied. The VGG16 model has an architecture composed of convolution, activation (ReLu), pooling, and fully connected layers. The Resnet-50 model has the same operations as VGG16, with the difference being deeper and considering residual information between the layers. We choose these models based on the promising results reported by Luz et al. [46] employing the VGG16 architecture for the periocular recognition and the capability of better generalization of the ResNet models compared to the VGG ones, as detailed in Section 2.2.1.

For both models, we employed the same architecture modifications and parameters described in [46]: in the training stage, we removed the last layer (used to predict) and added two new layers. The new last layer, used for classification, is composed of 171 neurons for Nice.II database and 197 neurons for CASIA-IrisV3-Interval database, where each neuron corresponds to a class in the training set and has a softmax-loss function. The layer before that one is a fully-connected layer with 256 neurons used to reduce feature dimensionality.

We split the training set into two subsets with 80% of the data for training and 20% for validation. We defined two learning rates for 30 epochs: 0.001 for the first 10 epochs and 0.0005 for the remaining 20. Other parameters include momentum = 0.9 and batch size = 48. The number of epochs used for training was chosen based on the experiments carried out in the validation set (highest accuracy and lowest loss). Similarly to [255, 46], we do not freeze the weights of any layer during training to perform the fine-tuning. The last layer of each model was removed, and the features were extracted on the new last layer.

We evaluated the models adopting the verification protocol reporting EER and Decidability metrics. For this, the all against all approach was applied in the test set, generating 4,634 intra-class pairs and 494,866 inter-class pairs.

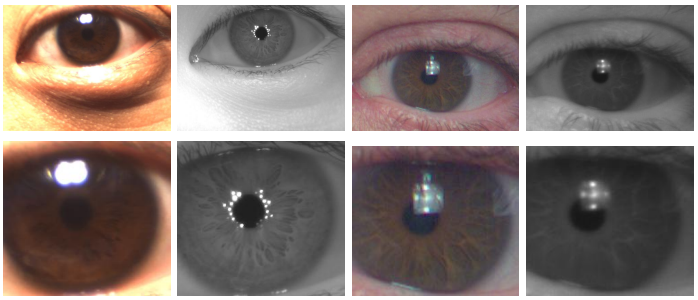
The cosine metric, which measures the cosine of the angle between two vectors, was applied to compute the difference between feature vectors. This metric can be used for information retrieval [256] due to its invariance to scalar transformation. The cosine distance metric is represented by

$$d_c(A, B) = 1 - \frac{\sum_{j=1}^{N} A_j B_j}{\sqrt{\sum_{j=1}^{N} A_j^2}\sqrt{\sum_{j=1}^{N} B_j^2}} \tag{4.1}$$

where A and B stand for the features vectors. We also employed other distances such as Euclidean, Mahalanobis, Jaccard, and Manhattan. However, due to its best performance, only the cosine distance was reported.

4.2 CROSS-SPECTRAL OCULAR BIOMETRICS

In this research, we analyzed the use of deep representations from the eye regions (iris and periocular) on the cross-spectral scenario, i.e., obtaining models able to match VIS against NIR wavelength images. Particularly, we evaluated and combined deep representations extracted from two modalities (traits): the iris and periocular regions. In the periocular modality, features were extracted from the entire image (considering the iris, sclera, skin, eyelids, and eyelashes components). On the other way, the iris features were extracted from a bounding box, i.e., a cropped image that contains only the iris region, as described by Zanlorensi et al. [23] (detailed in Section 4.1). These bounding boxes were generated manually by coarse annotations and are publicly available to the research community[1] and appear in [41]. Samples of the periocular and iris images used in our experiments are shown in Figure 4.5.



(a) (b) (c) (d)

Figure 4.5: VIS (a,c) and NIR (b,d) samples from the PolyU Cross-Spectral (a,b) and CROSS-EYED (c,d) databases. First and second rows show periocular and iris images, respectively. Extracted from [24].

Deep representations from the periocular and iris regions were extracted using a similar approach proposed in [23]. In this way, the VGG16 [84] and ResNet-50 [85] CNN models trained for face recognition were fine-tuned to each modality. We choose these models because they reported promising results in recent works applied in ocular recognition [46, 23, 48, 29]. The architecture modifications for both models consist of removing the last layer and adding two new layers. The first one is a fully-connected layer with 256 neurons that will be used as the feature representation and aim to reduce the feature dimensionality, since originally VGG16 and ResNet-50 have 4096 and 2048 features/outputs, respectively. The other layer added has a softmax cross-entropy loss function, and it is used only in the training phase in an identification mode. We chose a feature vector of 256 features based on the results reported by Luz et al. [46], where

[1] https://web.inf.ufpr.br/vri/databases/iris-periocular-coarse-annotations/

the authors evaluated different feature vector sizes and stated that vector with such size (256) showed the best trade-off regarding matching time, amount of memory required and matching effectiveness. The strategy applied to extract features from NIR and VIS images is detailed in Figure 4.6.

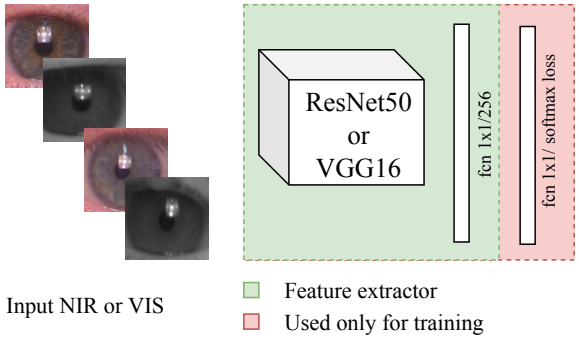


Figure 4.6: The cross-/intra-spectral ocular recognition strategy. A single model (ResNet50 or VGG16) is used to learn features from both spectra: NIR and VIS. Extracted from [24].

The number of epochs used for training was chosen based on a validation subset composed of 20% of the training set images. After defining the number of epochs, the CNN models were trained using the entire training set. The training was performed with the Stochastic Gradient Descent (SGD) optimizer without freezing any weights of the pre-trained layers. As previously mentioned, the last layer of each model was removed in the test phase, and the features were extracted from the first new last layer, composed of 256 neurons.

The all-against-all matching was performed using the cosine distance metric, which measures the cosine of the angle between two vectors. Regarding the similarity of biometrics features/representations, it is known that orientation is more important than the magnitude coefficient. Thus, the cosine distance often outperforms other distance measures.

The iris and periocular region representations were combined, applying the score-level fusion technique. Similar to approaches that used score-level fusion for iris and periocular region traits [121, 122, 20] and also based on the individual performance of each trait in our experiments, we chose to use weights of 0.6 and 0.4 for the periocular region and iris representations, respectively. To perform fusion at the score-level, first, we compute the matching for each trait independently. Then we calculated the weighted arithmetic mean between the cosine distances computed for the iris and periocular modalities.

It is important to note that, in the model learning process, all images (NIR and VIS) were used to feed the CNN models, making a single model to learn discriminant features of images captured in both spectra. To the best of our knowledge, this procedure is similar to the adopted in [29] for the CNN architecture. In the test phase the features are extracted for all images NIR or VIS images. However, note that only images acquired under different wavelengths are paired to match for evaluating the cross-spectral scenario.

4.2.1 Database, Metrics and Protocol

To evaluate the proposed method, we employed two well-known databases: the PolyU Cross-Spectral, and the CROSS-EYED. The PolyU Cross-Spectral database is composed of images obtained simultaneously under both NIR and VIS wavelengths. The entire database has 12,540 images with a resolution of 640×480 pixels. For every spectrum, there are 15 samples of each eye (left and right) from 209 subjects (418 classes) [20]. The CROSS-EYED iris database has

3,840 images from 120 subjects (240 classes). There are 8 samples from each of the classes for every spectrum. The resolution of the images is 400×300 pixels. All images were obtained at a distance of 1.5 meters, in an uncontrolled indoor environment, with a wide variation of ethnicity and eye colors, and lightning reflexes [18].

In all experiments, the *verification* setting was the unique considered, in which pairs of images are compared in order to determine whether a subject is whom he claims to be or not. For this, following a *one-against-all* pairwise matching strategy, all pairs of genuine and impostor comparisons were generated.

For a fair comparison with the state-of-the-art methods, the test protocol used in this work follows the procedures given in[20, 29], which consists of a *closed-world* protocol, where different instances of the same class are distributed in the training and test sets. In the PolyU Cross-Spectral database, the first ten instances from every subject were used for training, and the remainder (five) were employed for the matching. In the CROSS-EYED database, the first five instances from every subject are used for training, and the remaining three instances were employed for the matching.

To perform the experiments, we considered that in both databases, the NIR and VIS images were obtained synchronously. Thus, here in the intra-class comparison in the cross-spectral scenario, images of the same index were not matched because they represent the same image but in different spectra. Note that in work by Wang and Kumar [29], the authors considered that in the CROSS-EYED database, non-synchronously spectrum images were obtained (based on the numbers of intra- and inter-class comparisons), so they matched NIR against VIS images of the same index in the intra-class comparison. Then for a fair comparison with the state-of-the-art method [29], in the closed-world protocol, we also report results considering that the NIR and VIS images were obtained non-synchronously in the CROSS-EYED database.

To evaluate the robustness of the proposed methodology, we also evaluated and reported results on the *open-world* protocol, in which the training and test sets have images from different classes. In other words, there are no images from the same subject in the training and testing. For both databases, we use the first half of the subject images for training and the second half for testing in this protocol.

The images and class distributions for the training and test sets, as well as the number of genuine and impostors pairs generated in the test phase for both databases and protocols, are detailed in Table 4.1.

Table 4.1: Genuine and impostor matches for the Closed-world (CW) and Open-world (OW) protocols on Cross- and Intra-spectral scenarios. *The comparison with the state-of-the-art methods was performed using the closed-world protocol. Adapted from [24].

Database	Protocol	Scenario	Train/Test Images(Classes)	Gen./Imp. pairs
PolyU Cross-Spectral	CW	Cross	8,360(418)/4,180(418)	4,180/4,357,650
		Intra	8,360(418)/2,090(418)	4,180/2,178,825
	OW	Cross	6,270(209)/6,270(209)	21,945/9,781,200
		Intra	6,270(209)/3,135(209)	21,945/4,890,600
CROSS-EYED	CW	Cross	2,400(240)/1,440(240)	720/516,240
		Intra	2,400(240)/720(240)	720/258,120
	OW	Cross	1,920(120)/1,920(120)	3,360/913,920
		Intra	1,920(120)/960(120)	3,360/456,960

For evaluating the algorithms, we choose the EER metric and the decidability index d' [257]. The mean and standard deviation of 30 repetitions for the EER and decidability figures obtained by the proposed methodology are shown.

4.3 ATTRIBUTE NORMALIZATION

The development of ocular biometric systems operating under unconstrained environments is challenging since the collected data (images) may present some problems caused by noise, blur, motion blur, occlusion, eye gaze, off-angle, eyeglasses, contact lenses, makeup, among others. These problems generate high intra-class variability, degrading the uniqueness of the features extracted from the biometric trait.

With the recent advancement of deep learning techniques, several approaches applying Convolutional Neural Networks (CNN) to periocular recognition have been developed [46, 51, 88, 258, 24, 259]. An advantage of applications based on deep learning is that, unlike the handcrafted features, there is a process of representation learning. This process can produce feature extractor models invariant for some intra-class factors, depending on the training set's image samples. Nevertheless, new approaches are still being developed using handcrafted features and achieving top-ranked results in ocular recognition competitions [19, 27, 122, 13]. The main advantage of these approaches is the computational cost compared with methods based on deep learning techniques.

Even though CNN approaches can handle intra-class variability, several factors are present in images captured under unconstrained environments, which affect periocular recognition in biometric systems based on deep learning and mainly on handcrafted features. Regarding these kinds of problems, we proposed an image preprocessing method to normalize the most common image attributes that can decrease the recognition effectiveness in periocular biometric systems. The proposed attribute normalization prepossessing consists of remove or correct attributes that are different in a pairwise image comparison using deep models for image editing, as shown in Fig. 4.7. For example, in a database containing images from the same subject wearing and not wearing eyeglasses, the proposed preprocess will normalize all the images by removing the eyeglasses.

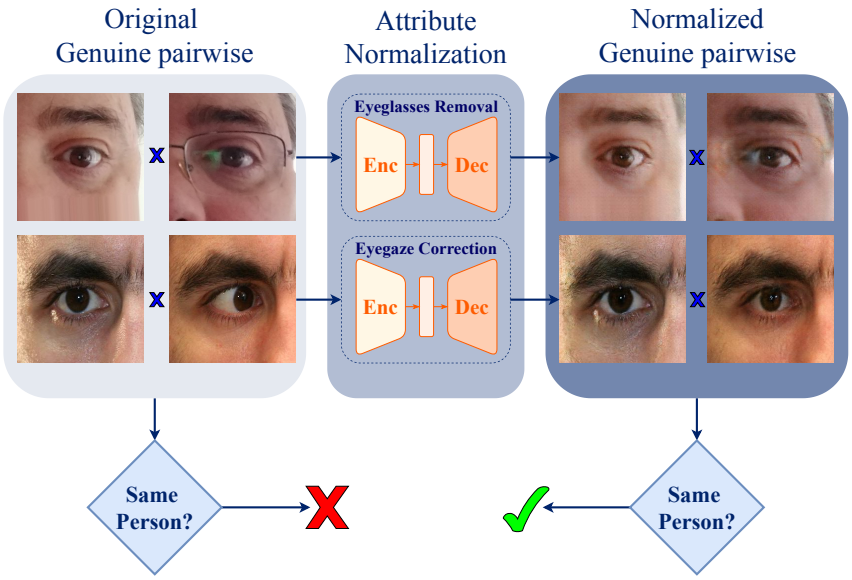


Figure 4.7: Cohesive perspective of the proposed attribute normalization scheme: images feed an encoder/decoder deep model for automatic image editing, removing the eyeglasses and correcting deviated gazes before the recognition step. This contributes for reducing the intra-class variability without significantly reducing the discriminability between classes, which is the key for the observed improvements in performance. Extracted from [25].

The proposed attribute normalization preprocess consists of applying generative deep models for image attribute editing to a pair of ocular images aiming for the correction/removal of

different attributes. Regarding the intra-class variability in periocular images caused by different aspects such as eyeglasses and eye gaze, the hypothesis that we considered in this work is that it is possible to decrease this variability by an attribute normalization preprocess.

To perform such a normalization process, we employed the AttGAN model [26] since its results compared with other state-of-the-art methods demonstrated a better capacity in changing facial attributes keeping the subject identity information as can be seen in Fig. 4.8, which is a crucial factor for a biometric system.

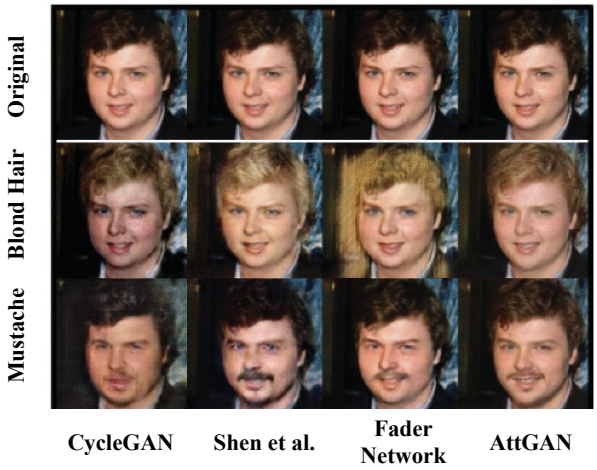


Figure 4.8: Comparison of state-of-the-art methods for facial attribute editing results. Adapted from [26].

The AttGAN [26] is a deep model based on an encoder/decoder architecture. Compared with other facial attribute editing models, its main difference is an attribute classification constraint, which requires the correct attribute manipulation in the generated images. Regarding information loss, the architecture has a reconstruction learning, used to preserve the other attribute details, i.e., changing only the required attribute. The model training is performed using three learning components: reconstruction, attribute classification, and adversarial learning. These components guarantee the visual and reconstruction quality of the generated images with the correct attribute manipulation. Due to all these features and mainly regarding reducing the information loss, we choose the AttGAN network to perform the proposed attribute normalization. As the generative model receives as input an image and the attributes to be changed, we performed the attribute normalization by feeding the model with the images and requesting removing the eyeglasses and correcting the eye gaze.

The AttGAN can handle multiple attribute editing, i.e., changing more than one attribute with a single model. However, as we had to use different databases for each attribute normalization in our experiments, we trained two models, one for each attribute. We validate our proposed normalization by comparing the results of biometric systems based on handcrafted features and deep learning approaches using the original and normalized images.

4.3.1 Databases and Baseline Methods

We performed the experiments in two databases: the UFPR-Eyeglasses, collected for this research and used for the eyeglasses attribute normalization, and the UBIPr [16] for the eye gaze normalization. These databases were detailed below.

The UFPR-Eyeglasses is a new challenging database to evaluate the occlusion effect caused by eyeglasses in periocular recognition using images captured by mobile devices under real uncontrolled environments. The database has 2,270 periocular images (containing both

eyes) from 83 subjects (166 classes), all taken by the subject himself/herself using his/her smartphone at the visible wavelength in 3 distinct sessions. This database comprises images captured by mobile devices from subjects wearing and not wearing eyeglasses. We manually annotated each image's iris bounding box to perform the image normalization regarding rotation and scale and to crop the periocular region of each eye to 256×256 pixels. The intra-class variations are mainly caused by different aspects of the images, such as illumination, occlusions, distances, reflection, eyeglasses, and image quality. The UFPR-Eyeglasses database (images and annotations) is available (under author request) to the research community at [`https://web.inf.ufpr.br/vri/databases/ufpr-eyeglasses/`].

The UBIPr database [16] is composed of 10,250 ocular images from 344 subjects. These images were captured under an uncontrolled environment by a Canon OS 5D camera with a 400mm focal length at visible wavelength. The main challenge of this database includes several variability factors in the images, such as different distances, scales, occlusions, poses, eye gazes, and eyeglasses. Unlike the UFPR-eyeglasses, this database does not contain images from the same subject with and without eyeglasses. Instead, there are images from the same subject with and without eye gaze. Thus, we used this database to evaluate eye gaze normalization.

We evaluated the proposed ocular normalization scheme using handcrafted features [115, 122], and deep representations based on approaches that recently reported state-of-the-art performances in the periocular and iris recognition [46, 23]. These methods are detailed below.

For the evaluation of the handcrafted features-based methods, we employed three approaches. The first is one of the first periocular recognition methods found in the literature, proposed by Park et al. [115]. This approach combined Local Binary Patterns (LBP) [260, 261], Histogram of Oriented Gradients (HOG) [262], and Scale-Invariant Feature Transform (SIFT) [263] features. The second one is the winner approach in the Miche-II contest [27, 122], which also has an iris recognition scheme. However, in our experiments, we used only the periocular recognition module, which was performed using Multi-Block Transitional Local Binary Patters (MB-TLBP) features [122]. Finally, we combine the following features by a score-level fusion: LBP, Local Phase Quantization (LPQ) [264], HOG and SIFT. All the features were extracted from a gray representation of the images extracted by the intensity channel. The normalized LBP and LPQ features were extracted from 16 patches with a size of 64×64 pixels cropped from each image. Then, each patch's features were concatenated, generating feature vectors with a size of 944 and 4096 for the LBP and LPQ, respectively. The HOG features were extracted from the entire image producing a feature vector with 72,900 of size.

Recent works reported promising results in developing biometric systems based on deep representations of the periocular region [46, 51, 258, 24, 259]. These approaches generally consist of a CNN model with a Softmax layer at the top, and it is trained using the cross-entropy loss function. After the training stage, the Softmax layer is removed, and then the deep representations can be extracted at the newest last layer. To evaluate the attribute normalization using these kinds of models, we employed two state-of-the-art methods to extract deep representations [46, 24]. These methods are based on the VGG16 and ResNet50 architectures pre-trained for face recognition [85]. Both methods generated a feature vector with a size of 256 for each image. We reported results from 5 runs (repetitions) for each model.

4.4 UFPR-PERIOCULAR DATABASE AND SOFT-BIOMETRICS

Regarding the existing ocular databases, it is difficult to assess the scalability performance of biometric applications, i.e., if an approach can produce discriminative features even in a large database in terms of the number of subjects. As we can see in Table 4.2, the databases in the

literature do not present a large number of subjects and have few sensors and session captures. As described in some previous works [24, 25], one common problem in ocular biometric systems is the intra-class variability, which is generally affected by noises and attributes present in the same individual images. A robust biometric system must handle images obtained from different sensors, extracting distinctive representations regardless of the source and environment. In this sense, samples from the same subject obtained in different sessions are of paramount importance to capture the intra-class variation caused by various noise factors.

Table 4.2: Comparison of the available ocular databases containing VIS images with our database (UFPR-Periocular).

database	Subjects	Images	Sessions	Sensors
VSSIRIS [194]	28	560	1	2
CSIP [198]	50	2,004	N/A	7
QUT [193]	53	212	N/A	2
IIITD [192]	62	1,240	N/A	3
UPOL [195]	64	384	N/A	1
UTIRIS [191]	79	1,540	2	2
MICHE-I [15]	92	3,732	2	3
CROSS-EYED [18, 19]	120	3,840	N/A	2
PolyU Cross-Spectral [20]	209	12,540	2	2
UBIRIS.v1 [190]	241	1,877	2	1
UBIRIS.v2 [1]	261	11,102	2	1
UBIPr [16]	261	10,950	2	1
VISOB [14]	550	**158,136**	2	3
UFPR-Periocular	**1,122**	33,660	**3**	**196**

Considering the above discussion, we created a new periocular database, called UFPR-Periocular. The subjects themselves collected the images that compose our database through a mobile application (app). In this way, the images were captured in unconstrained environments, with a minimum of cooperation from the participant, and have real noises caused by poor lighting, occlusion, specular reflection, blur, and motion blur. Fig. 4.9 shows some samples from the UFPR-Periocular. We also performed an extensive benchmark, employing several state-of-the-art architectures of CNN models that have been explored to develop ocular biometric systems.

Figure 4.9: Sample images from the UFPR-Periocular database. Observe that there is great diversity in terms of lighting conditions, age, gender, eyeglasses, specular reflection, occlusion, resolution, eye gaze, and ethnic diversity.

Note that our database is the largest one in terms of the number of subjects, sessions, and sensors, as shown in Table 4.2. It also has more images than all databases except VISOB. Another key feature is that the proposed database has images captured by 196 different mobile devices. The samples captured with less cooperation of the participant in unconstrained environments have several variations on the ocular images since they are obtained during three different sessions. To the best of our knowledge, this is the first ocular database with more than 1,000 subject samples and the largest one in different sensors in the literature. Thus, we believe that it can provide a new benchmark to evaluate and develop new robust ocular biometric approaches.

4.4.1 Database Information

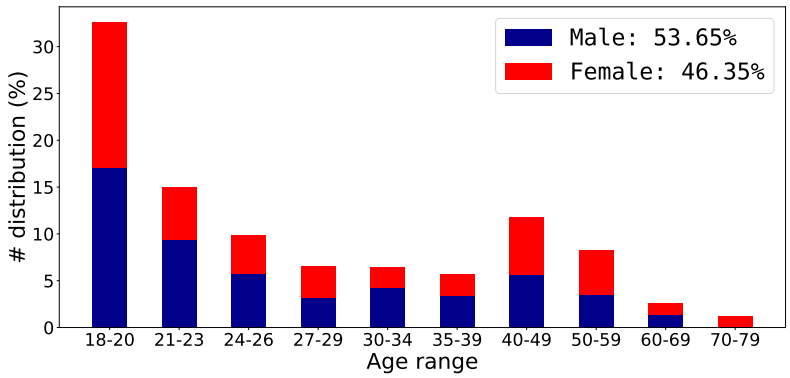
The UFPR-Periocular database was created to obtain images in unconstrained scenarios that contain realistic noises caused by occlusion, blur, and variations in lighting, distance, and angles. To this end, we developed a mobile application (app) enabling the participants to collect their pictures using their smartphones[2]. The single instruction to the participants is to place their eyes on a region of interest marked by a rectangle drawn in the app, as illustrated in "Picture" in Fig. 4.11. We also restricted the images to be captured in 3 sessions, with 5 images per session and a minimum interval of 8 hours between sessions. In this way, we guarantee that the database has samples of the same subject with different noises, mainly due to different lighting and environments. Furthermore, imposing this minimum time interval between sessions, it is possible to collect different attributes in the same subject's periocular region, e.g., subjects wearing and not wearing glasses and makeup. Another attractive feature of this database is that all participants are Brazilian. As Brazil has great ethnic diversity, there are images of subjects from different races, making this one of the first periocular databases with such cultural diversity.

The images were collected from June 2019 to January 2020. The gender distribution of the subjects is (53,65%) male and (46,35%) female, and approximately 66% of the subjects are under 31 years old. In total, the database has images captured from 196 different mobile devices – the five most used device models were: *Apple iPhone 8* (4.1%), *Apple iPhone 9* (3.1%), *Xiaomi Mi 8 Lite* (3.0%), *Apple iPhone 7* (3.0%), and *Samsung Galaxy J7 Prime* (2.7%). We remark that each subject captured all of their images using the same device model. The distribution of age, gender, and image resolutions present in our database is shown in Fig. 4.10.

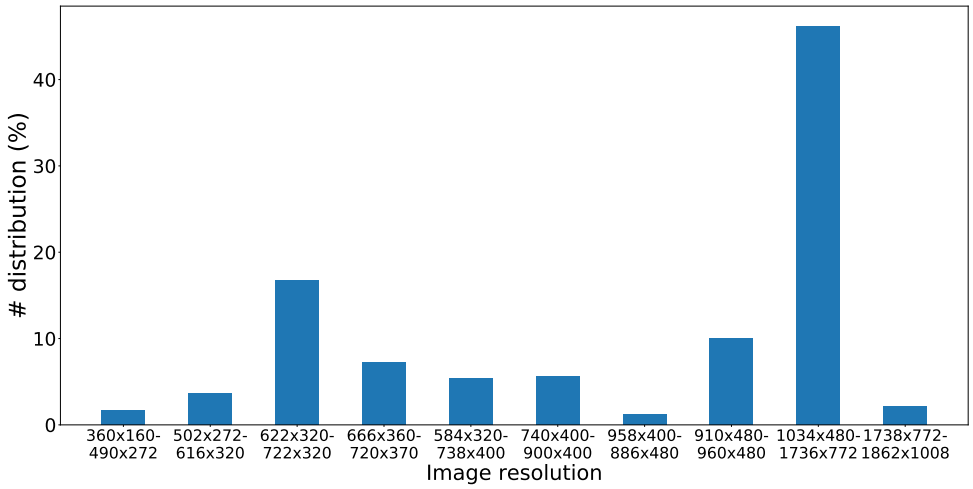
The database has 16,830 images of both eyes from 1,122 subjects. Image resolutions vary from 360×160 to 1862×1008 pixels – depending on the mobile device used to capture the image. We cropped/separated the right and left eyes' periocular regions to perform the benchmark, assigning a unique class to each side. Note that, once the image is cropped, the remainder image region is discarded as claimed in our project request to the Ethics Committee Board to preserve at maximum the identity of the participants. We manually annotated the eye corners with 4 points per image (inside and outside eye corners) and used these points to normalize the periocular region regarding scale and rotation. This process is detailed in Fig. 4.11. All the original and cropped periocular images, along with the eye corner annotations, are publicly available for the research community (upon request) at *https://web.inf.ufpr.br/vri/databases/ufpr-periocular/*.

Using the center point of each eye (average corners point), the images were rotated and scaled to normalize the eye positions in size of 512×512 pixels. Then, the images were split into 2 patches to create the left and right eye sides, generating 33,660 periocular images from

[2]Project approved by the Ethics Committee Board from the Health Science Sector of the Federal University of Paraná, Brazil – Process CAAE 02166918.2.0000.0102, registered in the *Plataforma Brazil* system – `https://plataformabrasil.saude.gov.br/`



(a) gender distribution among the age ranges



(b) image resolutions grouped into 10 intervals

Figure 4.10: Age, gender, and image resolution distributions in the UFPR-Periocular database. (a) note that gender has a balanced distribution, but the age range is concentrated under 30 years old (64% of the subjects). (b) more than 45% of the images have a resolution between 1034×480 and 1736×772 pixels, and more than 65% of the images have resolution higher than 740×400 pixels.

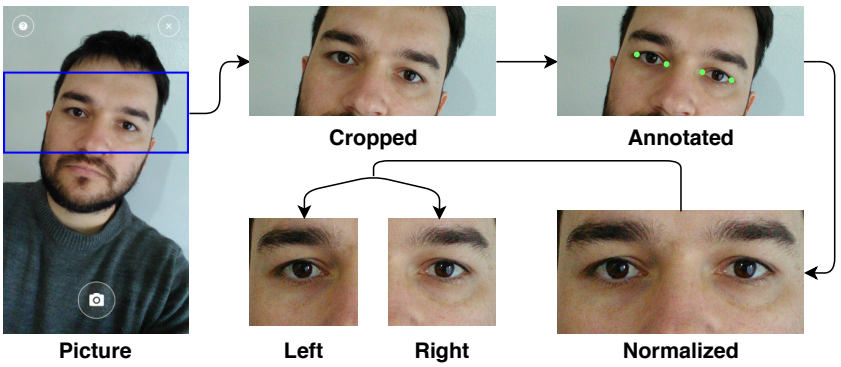


Figure 4.11: Image acquisition and normalization process. After the subject takes the shot, the rectangular region (outlined in blue) is cropped and stored. Then, the images are normalized in terms of rotation and scale using the manual annotations of the eyes' corners. Finally, the normalized images are cropped, generating the periocular regions of the left and right eyes.

2,244 classes. This database's intra- and inter-class variability are mainly caused by lighting,

occlusion, specular reflection, blur, motion blur, eyeglasses, off-angle, eye-gaze, makeup, and facial expression.

4.4.2 Experimental Protocols

We proposed protocols for the two most common tasks in biometric systems: identification (1:N) and verification (1:1). The identification task consists of determining a subject sample identity (probe) within a known database or a cluster (gallery). The probe is compared against all the gallery samples, considering the closest match as the subject's identity. Furthermore, probabilistic models can be employed/trained using the gallery data to determine the probe subject's identity based on the highest confidence output. The verification task refers to verifying whether a subject is whom she/he claims to be. If two samples match sufficiently, the identity is verified; otherwise, it is rejected [32]. Verification is usually used for positive recognition, where the goal is to prevent multiple people from using the same identity. The identification is a critical component in negative recognition, where the goal is to prevent a single person from using multiple identities [3]. Furthermore, the proposed protocol also encompasses two different scenarios: closed-world and open-world. In the closed-world protocol, the database is split through different samples from the same subject, i.e., training and test sets have samples of the same subjects. In the open-world protocol, there are different subjects both in the training and test sets. The identification task is performed in the closed-world protocol, while the verification task can be performed in both closed and open-world protocols.

In the open-world protocol, we also proposed two different splits regarding the training and validation sets. Note that we do not change the test set, keeping it in the open-world protocol, and only vary the training protocols. The first split uses the closed-world protocol, in which the training and validation sets have samples from the same subjects. On the other hand, the second split has different subjects in the training and validation sets, i.e., in an open-world protocol. With these two training/validation splits, it is possible to use multi-class networks (classification/identification) and also models based on the similarity of two distinct inputs (verification task): Siamese networks, triplet networks, and pairwise filters. Although models built for the verification task can be trained through the closed-world protocol, the design can be better improved using the open-world protocol to split the training and validation sets, as it is a more realistic scenario regarding the test set. Table 4.3 summarizes the proposed protocols.

Table 4.3: Images, Classes, and Pairwise comparison distributions for the closed-world (CW) and open-world (OW) protocols. Values for each fold (3 folds).

Protocol	Train/Val	Images / Classes			Genuine pairs / Impostor pairs		
		Train	Validation	Test	Train	Validation	Test
CW	CW/CW	13,464/2,244	8,976/2,244	11,220/2,244	33,660/ 90,599,256	13,464/40,266,336	22,440/12,583,230
OW	OW/CW	13,464/1,496	8,976/1,496	11,220/ 748	53,856/ 90,579,060	22,440/40,257,360	78,540/ 4,190,670
OW	OW/OW	15,000/1,000	7,440/ 496	11,220/ 748	105,000/112,387,500	52,080/27,621,000	78,540/ 4,190,670

We defined 3 folds with a stratified split into training, validation, and test sets for both biometric tasks (identification and verification) for all protocols. The test set comprises all against all comparisons for genuine pairs and aiming to reduce the pairwise comparisons only impostor pairs using the images of all subjects with the same sequence index, i.e., the i-th images of each subject are combined two-at-a-time to generate all impostor pairs, for $1 \leq i \leq n$, where $n = 3$ sessions $\times$ 5 images. As the UFPR-Periocular database has images captured under 3 sessions, we designated one session as a test set for each fold in the *closed-world protocol*. Thus, we have images from sessions 1 and 2, 2 and 3, 3 and 1 for training/validation, and sessions 3, 1,

and 2 for testing, respectively for each of the three folds. To evaluate the ability of the models to recognize subjects samples at different environments, for all folds, we employed samples of both sessions in the training and validation sets to fed the models with images from the same subject varying the capture conditions. For each subject, we employed the first 3 images of each session for training and the remaining 2 for validation (60%/40% for training/validation splits). The test set contains new images from the subjects present in the training/validation sets with different noises caused by the environment, lighting, occlusion, and facial attributes.

For the *open-world protocol* we generated the training, validation, and test sets by splitting the database through different subjects. Thus, for each fold, the test set has samples of subjects not present in the training/validation set. Splitting sequentially by the subject index for each fold, we have samples of 748 subjects for training/validation and 374 subjects for testing. Moreover, we proposed two different splits for the training/validation splits, the first one containing images of the same subject in the training and validation sets (closed-world validation). The second one contains samples from different subjects in the training and validation sets (open-world validation). Both training/validation protocols have pros and cons. The advantage of using the closed-world validation is that the training has samples of more subjects than the open-world validation protocol. However, in this scenario, the models can only learn distinctive features for the gallery samples and may not extract distinctive features for subjects not present in the training process. On the other hand, the open-world validation has samples of fewer subjects than the closed-world validation protocol, presenting a more realistic scenario since samples of subjects not known in the training stage are present in the validation set. In the closed-world validation protocol, for each one of the 748 subjects in the training set, we used the first 3 images of each session for training and the remaining 2 for validation (60%/40% for training/validation splits). In the open-world validation protocol, we employed samples of the first 700 subjects for training and samples of the remaining 48 subjects to validate each fold. The number of the generated pairwise comparison for all protocols are detailed in Table 4.3. The files determining all splits and setups detailed in this section are available along with the UFPR-Periocular database.

4.4.3 Benchmark and Experimental Setup

To carry out an extensive benchmark, we employed different models and strategies based on deep learning that achieved promising results in the ImageNet database/contest [131] and were applied in recent ocular works recognition [23, 48, 46, 29, 24]. These methods differ from each other in network architecture, loss function, and training strategies. We employed the following CNN architectures: Multi-class classification, Multi-task learning, Siamese networks, and Pairwise filters networks.

Inspired by several recent works [46, 23, 265, 29, 266, 24, 149, 50, 49], we performed the benchmark employing pre-trained models on ImageNet and also for face recognition (VGG16-Face and ResNet50-Face). Afterward, we fine-tuned these models using the UFPR-Periocular database.

Regarding Multi-class models, we evaluated the following CNN architectures that achieved expressive results in the ImageNet database/contest [131]: VGG16 [134], VGG16-Face [84], ResNet50 [6], ResNet50V2 [150], ResNet50-Face [85], InceptionResNet [7], MobileNetV2 [138], DenseNet121 [136], and Xception [135]. In summary, these models' architecture has several convolutional, pooling, activation, and fully-connected layers.

Regarding soft-biometric information, we created a Multi-task network sharing all convolutional layers and some dense layers. The model has exclusive dense layers for each task (soft-biometric prediction), followed by the prediction layers, using the softmax cross-entropy as function loss. Based on the multi-class classification results, we employed the MobileNetV2

as the base model on the multi-task approach. Furthermore, as detailed in Table 4.4, we build our multi-task model with hard parameter sharing for the following 5 tasks: (i) class prediction, (ii) age rate, (iii) gender, (iv) eye side, and (v) smartphone model.

Table 4.4: Multi-task architecture in the closed-world protocol.

#	Layer	Connected to	Input	Output
0	MobileNetV2 (88 layers)	–	$224 \times 224 \times 3$	1280
1	dense (classes)	#0	1280	256
2	dense (age)	#0	1280	256
3	dense (gender)	#0	1280	256
4	dense (eye side)	#0	1280	256
5	dense (smartphone model)	#0	1280	256
6	predict (classes)	#1	256	2244
7	predict (age)	#2	256	10
8	predict (gender)	#3	256	2
9	predict (eye side)	#4	256	2
10	predict (smartphone model)	#5	256	196

For the age estimation task, we generated the classes by grouping ages into the following 10 ranges: 18-20, 21-23, 24-26, 27-29, 30-34, 35-39, 40-49, 50-59, 60-69, and 70-79. The gender and eye side prediction tasks have only 2 classes, while the smartphone model prediction has 196 classes. Note that Multi-task learning networks can use the weighted loss for the tasks, penalizing the wrong classification of some tasks more than others. For simplicity, we do not use weighted losses in our experiments in this research, giving equal importance to all tasks.

Inspired by [90], which is one of the first works applying deep learning for iris verification, we also evaluated the performance of the pairwise filters network. This kind of model directly learns the similarity between a pair of images through pairwise filters. The Pairwise Filters Network is a Multi-class classification model that contains one or two outputs informing whether the input pairs are from the same class or different classes. The difference is that the network input is a pair of images instead of a single image. As this model requires a pair of images as input, different concatenation strategies can be employed. Following Liu et al. [90], we generated the input pairs by concatenating the images at the depth level. Let two RGB images with shapes of $224 \times 224 \times 3$, concatenating both images by their channels; the resulting input image will have a shape of $224 \times 224 \times 6$. The output of our model has two neurons and uses a softmax cross-entropy loss. As the verification problem has only two classes, this model's output can have only one neuron using a binary cross-entropy loss function. As in the Multi-task network, we employed the MobileNetV2 as a base model for our Pairwise Filters Network.

We also evaluated the Siamese Network, which learns similarities between a pair of images by a twin branch architecture sharing its parameters. As detailed in Table 4.5 we employ the MobileNetV2 as a base model for each branch and compute the similarity between the input pair images using the contrastive loss [158, 159].

Table 4.5: Siamese network architecture description.

#	Layer	Connected to	Input	Output
0	branch_a (MobileNetV2 (88 layers))	–	$224 \times 224 \times 3$	256
1	branch_b (MobileNetV2 (88 layers))	–	$224 \times 224 \times 3$	256
2	dense	#0 and #1	512	256
3	Euclidean dist. / Contrastive loss	#2	256	1

Similar to recent works on ocular recognition [46, 23, 48, 13], we modify all models by adding a fully convolutional layer before the last layer (softmax) to generate a feature vector

with a size of 256 for each image. The models' default input size is $224 \times 224 \times 3$, except for the InceptionResNet and Xception models, which have an input size of $299 \times 299 \times 3$. Note that the input dimensions are different because we used pre-trained models, and our fine-tuning process should respect the input size of the original architectures.

For all methods, the training was performed during 60 epochs with a learning rate of 10^{-3} for the first 15 epochs and 5×10^{-4} for the remaining epochs using the Stochastic Gradient Descent (SGD) optimizer. Then, we used the epoch's weights that achieve the lower loss in the validation set to perform the evaluation.

We employed Rank 1 and Rank 5 accuracy for the identification task and the Area Under the Curve (AUC), Equal Error Rate (EER), and Decidability (DEC) metrics for verification. Furthermore, to generate the verification scores, we computed the cosine distance between the deep representations generated by each CNN model.

Regarding the models explicitly developed for the verification tasks, i.e., the Siamese network and the Pairwise Filters network, as this task has unbalanced samples of genuine and impostors pairs, selecting the best samples to perform the training is challenging. Thus, trying to fit the models by feeding them as diverse samples as possible, we employed all genuine pairs and randomly selected the same number from the impostor pairs for each epoch. Hence, each epoch may have different impostor samples. However, for a fair comparison, we generated the random impostor pairs only once for each epoch and fold and used the same samples for training both models. The reported results are from 5 repetitions for each fold, except for the Siamese and Pairwise filter networks, in which we ran only 3 repetitions due to the high computational cost.

4.4.4 Final Remarks

This Chapter described and detailed the proposed methods, experimental protocols, and databases employed to evaluate the raised hypothesis. We started by our method to analyze the need for segmentation and normalization process in iris recognition when using deep representations. Then, we described a method using a single CNN model to directly learn representations from ocular images (iris and periocular) captured at NIR and VIS wavelengths. Regarding the intra-class variability due to non-inherent subject attributes, we described our proposed approach employing a GAN model to perform an attribute normalization. Finally, we presented our new collected periocular database (UFPR-Periocular), describing how the images were collected, the database stats, and the proposed experimental protocol.

5 RESULTS AND DISCUSSION

In this chapter, we report and discuss the results achieved for our approaches organized by the same subjects detailed in Chapter 4 (Proposed Methods).

5.1 THE IMPACT OF PREPROCESSING ON DEEP REPRESENTATIONS FOR IRIS RECOGNITION

In our first investigation, we performed an ablation study of preprocessing steps on deep representations for iris recognition analyzing the impact of the data augmentation techniques using non-segmented iris images. Then, the impact of both iris segmentation and iris delineation. Finally, the best results obtained by the proposed approaches are compared with state-of-the-art in the Nice.II database. In all subsections, the impact of normalization is also analyzed. Note that in all experiments, the mean and standard deviation values from 30 runs are reported. For analyzing the different results, we perform statistical paired t-tests at significance level $\alpha = 0.05$. The table rows with the results that have no statistical difference were painted with the same color.

5.1.1 Data Augmentation

In the first analysis, we evaluated the impact of the data augmentation. For ease of analysis, all iris images employed in this initial experiment may contain noise in the iris region, i.e., no segmentation preprocessing was applied. As shown in Table 5.1 and Table 5.2, in all cases where data augmentation was employed, the decidability and EER values improved with statistical difference. Note that the models trained with data augmentation reported smaller standard deviations. In general, it is also observed that non-normalization yielded better results than 8 : 1 and 4 : 2 normalization schemes for both trained models, i.e., VGG16 and ResNet-50.

Table 5.1: Impact of the data augmentation (DA) on the effectiveness obtained with VGG16 and ResNet-50 in Non-Seg. experiments in Nice.II database

Network	Norm.	DA	EER (%)	Decidability
VGG16	8 : 1		26.19 ± 1.95	1.3140 ± 0.1246
VGG16	8 : 1	✓	23.63 ± 1.33	1.4712 ± 0.0881
ResNet-50	8 : 1		24.38 ± 1.41	1.4297 ± 0.0916
ResNet-50	8 : 1	✓	19.18 ± 0.75	1.7988 ± 0.0552
VGG16	4 : 2		24.77 ± 1.42	1.4127 ± 0.1001
VGG16	4 : 2	✓	18.74 ± 0.89	1.8527 ± 0.0712
ResNet-50	4 : 2		22.78 ± 1.22	1.5307 ± 0.0853
ResNet-50	4 : 2	✓	17.11 ± 0.53	1.9822 ± 0.0482
VGG16	Non-Norm		23.32 ± 1.10	1.4891 ± 0.0740
VGG16	**Non-Norm**	✓	**17.49 ± 0.90**	**1.9529 ± 0.0760**
ResNet-50	Non-Norm		21.51 ± 0.97	1.6119 ± 0.0677
ResNet-50	**Non-Norm**	✓	**13.98 ± 0.55**	**2.2480 ± 0.0528**

It is worth noting that the largest differences occurred in the non-normalized inputs in both databases, with greater impact specifically in the ResNet-50 model in Nice.II database, where the mean EER dropped 7.53% and the decidability improved 0.6361 when applying data augmentation.

Table 5.2: Impact of the data augmentation (DA) on the effectiveness obtained with VGG16 and ResNet-50 in Non-Seg. experiments in CASIA-Interval database

Network	Norm.	DA	EER (%)	Decidability
VGG16	8 : 1		11.67 ± 1.47	2.5992 ± 0.1845
VGG16	8 : 1	✓	10.86 ± 0.86	2.7117 ± 0.1089
ResNet-50	8 : 1		8.30 ± 0.90	2.9443 ± 0.1257
ResNet-50	8 : 1	✓	6.95 ± 0.66	3.2183 ± 0.1220
VGG16	4 : 2		12.18 ± 1.13	2.5552 ± 0.1300
VGG16	4 : 2	✓	11.37 ± 0.73	2.6376 ± 0.0978
ResNet-50	4 : 2		10.43 ± 0.77	2.6682 ± 0.1015
ResNet-50	4 : 2	✓	9.01 ± 0.95	2.9009 ± 0.1322
VGG16	Non-Norm		9.85 ± 0.79	2.8412 ± 0.1104
VGG16	**Non-Norm**	✓	**7.42 ± 0.50**	**3.2700 ± 0.0798**
ResNet-50	Non-Norm		7.06 ± 0.65	3.1861 ± 0.1003
ResNet-50	**Non-Norm**	✓	**5.50 ± 0.37**	**3.5350 ± 0.0781**

5.1.2 Segmentation

In the second analysis, we investigated the impact of the segmentation for noise removal. For such an aim, the CNN models were trained (fine-tuned): using segmented and non-segmented images, all with data augmentation.

In Nice.II database, as can be seen in Table 5.3, for the VGG model segmentation has improved the results. On the other hand, for the ResNet-50 model, the non-segmented images have presented better results. For both models, statistical difference is achieved in two situations, and in another one (rows painted with the same color), there is no statistical difference.

Table 5.3: Impact of the segmentation (Seg.) on the effectiveness of iris verification for VGG16 and ResNet-50 networks in Nice.II database. Same color rows do not present statistical significance.

Network	Norm.	Seg.	EER(%)	Decidability
VGG16	8 : 1	✓	22.58 ± 1.07	1.5437 ± 0.0697
VGG16	8 : 1		23.63 ± 1.33	1.4712 ± 0.0881
ResNet-50	8 : 1	✓	20.68 ± 1.39	1.6801 ± 0.1071
ResNet-50	8 : 1		19.18 ± 0.75	1.7988 ± 0.0552
VGG16	4 : 2	✓	18.00 ± 0.93	1.9055 ± 0.0750
VGG16	4 : 2		18.74 ± 0.89	1.8527 ± 0.0712
ResNet-50	4 : 2	✓	17.44 ± 0.85	1.9450 ± 0.0803
ResNet-50	4 : 2		17.11 ± 0.53	1.9822 ± 0.0482
VGG16	Non-Norm	✓	**17.48 ± 0.68**	1.9439 ± 0.0589
VGG16	Non-Norm		17.49 ± 0.90	1.9529 ± 0.0760
ResNet-50	Non-Norm	✓	14.89 ± 0.78	2.1781 ± 0.0794
ResNet-50	**Non-Norm**		**13.98 ± 0.55**	**2.2480 ± 0.0528**

The results of the CASIA-Interval database (Table 5.4) show that in the two cases where there is a statistical difference, the best performance occurred using segmented images, with the lowest impact on ResNet-50. In the normalized images in aspect ratio 8 : 1, there is no statistical difference using or not a segmentation technique. The results of non-normalized and non-segmented images are not presented because the methodology used for segmentation does not provide these images. However, unlike the Nice.II database, in this one, the images do not have a specular reflection, and the segmentation process only removes eyelid and eyelash.

Table 5.4: Impact of the segmentation (Seg.) on the effectiveness of iris verification for VGG16 and ResNet-50 networks in CASIA-Interval database. Same color rows do not present statistical significance.

Network	Norm.	Seg.	EER(%)	Decidability
VGG16	8 : 1	✓	10.69 ± 0.95	2.7085 ± 0.1132
VGG16	8 : 1		10.86 ± 0.86	2.7117 ± 0.1089
ResNet-50	8 : 1	✓	6.84 ± 0.45	3.2607 ± 0.0853
ResNet-50	8 : 1		6.95 ± 0.66	3.2183 ± 0.1220
VGG16	4 : 2	✓	10.19 ± 0.96	2.7869 ± 0.1195
VGG16	4 : 2		11.37 ± 0.73	2.6376 ± 0.0978
ResNet-50	4 : 2	✓	8.27 ± 0.82	3.0154 ± 0.1251
ResNet-50	4 : 2		9.01 ± 0.95	2.9009 ± 0.1322
VGG16	Non-Norm	✓	**7.42 ± 0.50**	3.2700 ± 0.0798
VGG16	Non-Norm		* * *	* * *
ResNet-50	**Non-Norm**	✓	**5.50 ± 0.37**	**3.5350 ± 0.0781**
ResNet-50	Non-Norm		* * *	* * *

Regarding the better results achieved by the ResNet-50 models when using non-segmented images, we hypothesized that this might be related to the fact that the ResNet-50 architecture uses residual information and is deeper than VGG. Thus, some layers of ResNet-50 might be responsible for extracting discriminant patterns present in regions that were occluded in the segmented images but not in non-segmented ones. Moreover, in segmented images, black regions (zero values) were employed for representing noise regions, and no special treatment was given for those regions.

It is noteworthy that segmentation is a complex process and might impact positively or negatively. However, as the best results were achieved by the ResNet-50 models using non-segmented images, we state that the segmentation preprocessing can be disregarded using the suitable representation model. Once again, non-normalization showed better results in all scenarios, being more expressive than in the data augmentation analysis.

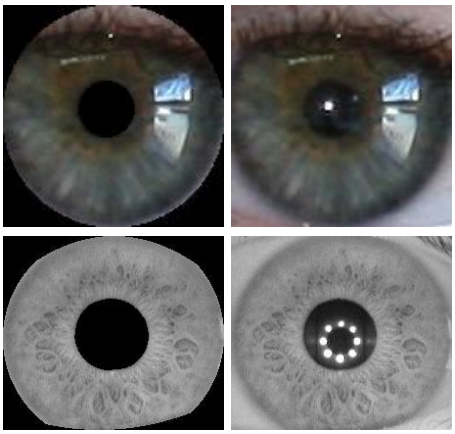
5.1.3 Delineation

We also evaluated the impact on recognizing using a usual delineated iris image and a non-delineated iris image, i.e., applying only the *squared* iris bounding box as input to the deep feature extractor. In both situations, non-normalized and non-segmented images are used. A delineated iris image and its corresponding bounding box (or non-delineated) from Nice.II and CASIA-Interval databases are shown in Fig. 5.1.

The comparison of the results of this analysis is shown in Table 5.5 and 5.6. Although the results reported by delineated iris images are better in some cases, there is a statistical difference only using the VGG16 model in the CASIA-Interval database. From this result, we state that the iris bounding box can be used as input for deep representation without the iris delineating (a.k.a. detection) preprocessing.

Table 5.5: Comparison of delineated and non-delineated iris images in Nice.II database. Both with no segmentation (for noise removal), normalization and data augmentation.

Method	Delineated	EER (%)	Decidability
VGG16	✓	17.49 ± 0.90	1.9529 ± 0.0760
VGG16		17.52 ± 0.98	1.9652 ± 0.0790
Resnet-50	✓	13.98 ± 0.55	2.2480 ± 0.0528
Resnet-50		14.26 ± 0.47	2.2304 ± 0.0542



(a) (b)

Figure 5.1: Input images: (a) delineated iris and (b) non-delineated iris / bounding box version from the NICE.II (top row) and CASIA-IrisV3-Interval database.

Table 5.6: Comparison of delineated and non-delineated iris images in CASIA-Interval database. Both with no normalization and data augmentation.

Method	Delineated	EER (%)	Decidability
VGG16	✓	7.42 ± 0.50	3.2700 ± 0.0798
VGG16		7.02 ± 0.48	3.3583 ± 0.0973
Resnet-50	✓	5.50 ± 0.37	3.5350 ± 0.0781
Resnet-50		5.42 ± 0.43	3.5532 ± 0.0900

Considering that the bounding box is not pure iris, it is important to verify if this modality can still be considered iris recognition since there may be discriminant patterns that have been extracted from regions outside the iris. Therefore, the proposed methodology was compared with state-of-the-art methods using delineated iris images.

5.1.4 Final Considerations

Finally, the results attained with our models using non-normalized, non-segmented, and delineated iris images are compared with the state-of-the-art approaches in the Nice.II database and it is shown in Table 5.7.

These experiments showed that the representations learned using deep models perform better the iris verification task on the Nice.II competition when the preprocessing steps of normalization and segmentation (for noise removal) are removed, outperforming the state-of-the-art method, which uses preprocessed images.

Table 5.7: Results on the NICE.II contest database. Comparison of the state of the art with the results achieved by our proposed approaches using non-normalized, non-segmented, and delineated iris images.

Method	EER (%)	Decidability
Wang et al.[212]	19.00	1.8213
Silva et al.[48]	14.56	2.2200
Proposed ResNet-50 [23]	13.98	2.2480

It is important to note that the methodology proposed by Silva et al. [48] reports the result of the CNN model that achieved better values of EER and decidability, so we also report

the results of our best CNN model. However, for a fair comparison, we encourage using the models' mean value since there may be variations in the results.

5.2 DEEP REPRESENTATIONS FOR CROSS-SPECTRAL OCULAR BIOMETRICS

This section presents and discusses the results observed for the intra- and cross-spectral scenarios for both iris and periocular recognition. We started by providing the results using the closed-world protocol to establish a baseline concerning the state-of-the-art methods. We also investigated the impact of the feature vector size and the weights used to merge information from the periocular region and iris traits. The results using the open-world protocol are then presented to perceive how robust deep representations can be obtained. Using the ResNet-50 model, a comparison of the verification effectiveness using features extracted from various network depths is performed. Lastly, we performed a subjective analysis of the pairwise errors.

In a complementary setting, we explored the advantages of fusing representations from the periocular and iris traits to improve the recognition performance. Similar to previous works [20, 121, 122] that applied higher weights in the most discriminating traits, and also considering that in all our experiments, the periocular region reported better results compared to the iris, we decided to use constant weights of 0.6 and 0.4 respectively for the periocular and iris representations when obtaining the fused score by linear combination.

5.2.1 Closed-world protocol

At first, Table 5.8 and Table 5.9 report the results observed for verification mode, in the cross-spectral and intra-spectral scenarios (NIR against NIR and VIS against VIS) and using the closed-world protocol. In a way similar to Nalla and Kumar [20] and also to guarantee a fair comparison to their method, the fusion of two spectra on the PolyU Cross-Spectral database was carried out by linear combination, using weights of 0.6 and 0.4, respectively, to the NIR and VIS images. However, based on the individual spectral results, on the CROSS-EYED database, we used weights of 0.6 and 0.4 for the VIS and NIR representations, respectively. Also, on the CROSS-EYED database, we can perceive that the spectral fusion using iris representations extracted by the VGG16 model reported lower results than the only VIS spectral information. The results show that the representations obtained from NIR images presented a higher EER value, which penalized the fusion of spectra. Therefore, lower weight for NIR representations may improve the fusion result. The results of those fusions are shown in Table 5.8 and Table 5.9 (VIS and NIR Fusion section).

Anyway, it can be seen that - for both databases - the proposed approach achieved better results than the state-of-the-art methods, both in the cross-spectral and in the intra-spectral scenarios even that the protocol used in our experiments is more challenging. For example, in the PolyU Cross-Spectral database, we used images from all 209 subjects in the experiments, while the approaches proposed by Wang and Kumar [29], and Nalla and Kumar [20] used images from only 140 subjects. In the CROSS-EYED database, based on the number of pairs of intra-class comparisons reported in the experiments by Wang and Kumar [29], the authors considered that the database has images obtained non-synchronously. Images from the CROSS-EYED database were obtained using a dual sensor with a beam splitter, so the NIR and VIS images are acquired simultaneously. However, we visually verified that the same index images, i.e., those that should be the same one in the NIR and VIS, have a random shift in each spectrum. Thus, we reported the results using both protocols for a fair comparison with the state-of-the-art approaches, considering the images obtained synchronously and non-synchronously. Note that

Table 5.8: Results - closed-world protocol on the PolyU Cross-Spectral database. *Using only 140 subjects from a total of 209. Extracted from [24].

Approach	Modality	EER (%)	Decidability
Cross-Spectral			
CNN with SDH [29]*	Iris	5.39	2.13
CNN with SDH [29]	Iris	12.41	–
VGG16 with SDH [29]*	Iris	4.85	–
Proposed VGG16	Iris	2.16 ± 0.16	5.23 ± 0.08
ResNet50 with SDH [29]*	Iris	7.17	–
Proposed ResNet50	**Iris**	**1.13 ± 0.14**	**5.17 ± 0.08**
Proposed VGG16	Periocular	1.80 ± 0.21	6.03 ± 0.20
Proposed ResNet50	**Periocular**	**0.78 ± 0.09**	**5.97 ± 0.08**
Proposed VGG16	Fusion	0.93 ± 0.10	6.97 ± 0.13
Proposed ResNet50	**Fusion**	**0.49 ± 0.06**	**6.75 ± 0.08**
VIS vs VIS			
Nalla and Kumar [20]*	Iris	6.56	–
Proposed VGG16	Iris	1.53 ± 0.12	6.27 ± 0.08
Proposed ResNet50	**Iris**	**0.78 ± 0.08**	**5.91 ± 0.07**
Proposed VGG16	Periocular	1.50 ± 0.16	6.63 ± 0.21
Proposed ResNet50	**Periocular**	**0.61 ± 0.11**	**6.57 ± 0.08**
Proposed VGG16	Fusion	0.76 ± 0.10	7.73 ± 0.14
Proposed ResNet50	**Fusion**	**0.35 ± 0.06**	**7.44 ± 0.10**
NIR vs NIR			
Nalla and Kumar [20]*	Iris	3.97	–
Proposed VGG16	Iris	1.21 ± 0.13	6.61 ± 0.10
Proposed ResNet50	**Iris**	**0.68 ± 0.07**	**6.05 ± 0.07**
Proposed VGG16	Periocular	1.56 ± 0.19	6.58 ± 0.21
Proposed ResNet50	**Periocular**	**0.68 ± 0.10**	**6.59 ± 0.07**
Proposed VGG16	Fusion	0.70 ± 0.11	7.86 ± 0.17
Proposed ResNet50	**Fusion**	**0.40 ± 0.06**	**7.54 ± 0.09**
VIS and NIR Fusion			
Nalla and Kumar [20]*	Iris	2.86	–
Proposed VGG16	Iris	1.01 ± 0.09	6.81 ± 0.08
Proposed ResNet50	**Iris**	**0.59 ± 0.08**	**6.29 ± 0.07**
Proposed VGG16	Periocular	1.36 ± 0.15	6.79 ± 0.21
Proposed ResNet50	**Periocular**	**0.56 ± 0.10**	**6.82 ± 0.08**
Proposed VGG16	Fusion	0.63 ± 0.10	8.05 ± 0.16
Proposed ResNet50	**Fusion**	**0.35 ± 0.05**	**7.75 ± 0.10**

we collected the state-of-the-art results from the original papers [20, 29], i.e., we did not have implemented any approach from these works.

In terms of the CNN architectures, the ResNet-50 model reported lower EER values compared to the VGG16 model in all cases. However, in some cases, specifically in the PolyU Cross-Spectral database, the representations extracted with the VGG16 model obtained a better separation of intra- and inter-class distributions, as shown in their Decidability index.

The results show that in CROSS-EYED, the periocular modality achieved better results than the iris one. However, in the PolyU Cross-Spectral database, there is no significant difference between iris and periocular representations, mainly in the intra-spectral experiments. From a visual inspection analysis of the pairwise comparison errors (some examples are shown in Subsection 5.2.5), we perceived that in the PolyU Cross-Spectral database, some uncontrolled conditions present in the images such as pose, eye gaze, and rotation might penalize the quality

Table 5.9: Results - closed-world protocol on the CROSS-EYED database. *Same protocol used by Wang and Kumar [29]. Extracted from [24].

Approach	Modality	EER (%)	Decidability
Cross-spectral			
CNN with SDH [29]	Iris	6.34	2.54
VGG16 with SDH [29]	Iris	3.13	–
Proposed VGG16*	Iris	5.58 ± 0.59	3.87 ± 0.16
Proposed VGG16	Iris	6.76 ± 0.56	3.58 ± 0.14
ResNet50 with SDH [29]	Iris	6.11	–
Proposed ResNet50*	**Iris**	**2.45 ± 0.25**	**4.73 ± 0.09**
Proposed ResNet50	**Iris**	**3.07 ± 0.38**	**4.49 ± 0.09**
Proposed VGG16*	Periocular	2.35 ± 0.28	5.61 ± 0.20
Proposed VGG16	Periocular	3.18 ± 0.42	5.19 ± 0.21
Proposed ResNet50*	**Periocular**	**1.45 ± 0.24**	**4.73 ± 0.09**
Proposed ResNet50	**Periocular**	**1.95 ± 0.35**	**5.34 ± 0.12**
Proposed VGG16*	Fusion	1.86 ± 0.19	5.78 ± 0.11
Proposed VGG16	Fusion	2.66 ± 0.29	5.31 ± 0.12
Proposed ResNet50*	**Fusion**	**1.06 ± 0.15**	**6.29 ± 0.11**
Proposed ResNet50	**Fusion**	**1.40 ± 0.26**	**5.93 ± 0.12**
VIS vs VIS			
Proposed VGG16	Iris	3.66 ± 0.39	4.85 ± 0.16
Proposed ResNet50	**Iris**	**2.47 ± 0.42**	**5.12 ± 0.13**
Proposed VGG16	Periocular	2.60 ± 0.40	5.57 ± 0.21
Proposed ResNet50	**Periocular**	**1.70 ± 0.37**	**5.66 ± 0.13**
Proposed VGG16	Fusion	1.94 ± 0.29	6.15 ± 0.16
Proposed ResNet50	**Fusion**	**1.17 ± 0.25**	**6.39 ± 0.13**
NIR vs NIR			
Proposed VGG16	Iris	7.31 ± 0.91	3.46 ± 0.18
Proposed ResNet50	**Iris**	**2.74 ± 0.34**	**4.72 ± 0.08**
Proposed VGG16	Periocular	2.97 ± 0.46	5.36 ± 0.23
Proposed ResNet50	**Periocular**	**1.78 ± 0.39**	**5.54 ± 0.13**
Proposed VGG16	Fusion	2.40 ± 0.35	5.36 ± 0.12
Proposed ResNet50	**Fusion**	**1.31 ± 0.24**	**6.14 ± 0.12**
VIS and NIR Fusion			
Proposed VGG16	Iris	3.69 ± 0.39	4.65 ± 0.15
Proposed ResNet50	**Iris**	**2.18 ± 0.31**	**5.25 ± 0.10**
Proposed VGG16	Periocular	2.44 ± 0.43	5.70 ± 0.22
Proposed ResNet50	**Periocular**	**1.54 ± 0.30**	**5.76 ± 0.13**
Proposed VGG16	Fusion	1.92 ± 0.29	6.09 ± 0.14
Proposed ResNet50	**Fusion**	**1.11 ± 0.20**	**6.47 ± 0.12**

of the periocular representations. These conditions are more controlled in the CROSS-EYED images. Also, CROSS-EYED images are smaller than PolyU Cross-Spectral images, so the iris region is even smaller. The periocular images are better centralized based on the iris region in the CROSS-EYED and not in the PolyU Cross-Spectral database. Nevertheless, CROSS-EYED images present a more significant difference in color and illumination among classes, which makes them more distinct and may explain the better results in VIS against VIS comparisons than NIR against NIR.

5.2.2 Feature size and fusion weights analyses

This section analyzes and discusses the impact of feature vector size and the weights used for the fusion of the iris and periocular region representations. We choose the feature size of 256 based on the experiments and results reported in [46]. Therefore, we also performed some experiments creating new models with different sizes in the last layer before the Softmax one, i.e., the layer used to extract the features (representations). The results of the fusion of iris and periocular representations extracted with these models are presented in Table 5.10. Luz et al. [46] stated that for the cosine distance metric, high dimensional vectors resulted in better performance. Conversely, our results show that representations extracted with the ResNet50 model achieve lower values of EER when the feature vector is smaller. The same occurs in the VGG16 model features in the PolyU Cross-Spectral database. Regarding the decidability index, the size of the feature vector does not show to have much impact. These results may be related to the fact that both models can generate sparse feature vectors, as stated by Wang and Kumar [29]. Thus a bigger feature vector will not always improve the performance of the biometric system. Here, we decided to keep a feature vector size of 256 because it keeps a trade-off between EER and Decidability.

Table 5.10: Feature vector size results fusing iris and periocular region traits on Cross-spectral scenario. Extracted from [24].

Model	Feat. Size	PolyU Cross-Spectral		CROSS-EYED	
		EER (%)	Decidability	EER (%)	Decidability
ResNet50	1024	0.54 ± 0.09	6.76 ± 0.10	1.61 ± 0.25	5.93 ± 0.13
	512	0.56 ± 0.06	6.73 ± 0.08	1.35 ± 0.22	6.00 ± 0.11
	256	0.49 ± 0.06	6.75 ± 0.08	1.40 ± 0.26	5.93 ± 0.12
	128	0.43 ± 0.05	6.70 ± 0.08	1.35 ± 0.30	5.99 ± 0.13
	64	0.37 ± 0.07	6.50 ± 0.08	1.26 ± 0.22	5.93 ± 0.15
	32	0.30 ± 0.05	6.05 ± 0.15	1.41 ± 0.27	5.65 ± 0.16
VGG16	1024	0.99 ± 0.10	6.85 ± 0.08	2.68 ± 0.28	5.29 ± 0.11
	512	0.92 ± 0.12	6.94 ± 0.11	2.53 ± 0.38	5.35 ± 0.14
	256	0.93 ± 0.10	6.97 ± 0.13	2.66 ± 0.29	5.31 ± 0.12
	128	0.80 ± 0.12	7.03 ± 0.10	2.78 ± 0.33	5.28 ± 0.10
	64	0.73 ± 0.11	6.93 ± 0.11	2.67 ± 0.37	5.23 ± 0.15
	32	0.69 ± 0.10	6.46 ± 0.07	2.79 ± 0.47	4.98 ± 0.17

Similar to some approaches [121, 122, 20] and based on the individual performance in our experiments, we choose weights of 0.6 and 0.4 for the periocular and iris fusion, respectively. Nevertheless, we evaluated the impact of different iris and periocular weights on the trait representations fusion in the cross-spectral scenario for both models. Indeed, we impose $w_p \in [0, 1]$, such that $w_i + w_p = 1$, where w_p and w_i stand for the periocular and iris weights, respectively. The results are reported in Figure 5.2.

Even though the values of EER are lower using features extracted with the ResNet50 model, we can observe a similar behavior regarding the weight difference in both databases for both models. That is, when the weights are appropriately combined, the best results were achieved. We can also observe that the periocular trait has more impact on the CROSS-EYED database than on the PolyU Cross-Spectral database. We also note that on the PolyU Cross-Spectral database, in some cases, fusion with a higher iris weight ($w_i = 0.6$ and $w_p = 0.4$ using VGG16 features) may achieve a lower value of EER.

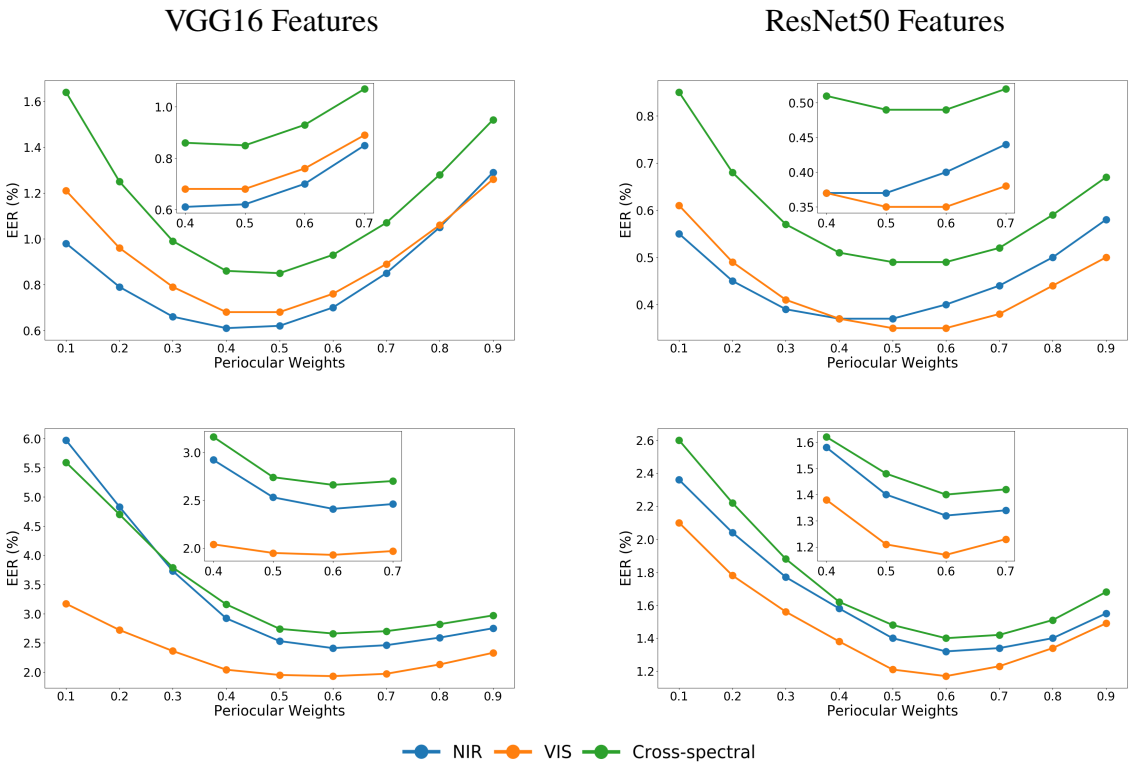


Figure 5.2: Periocular weights impact on the traits fusion in the cross-spectral scenario on the PolyU Cross-Spectral (top row) and CROSS-EYED (bottom row) databases. Extracted from [24].

5.2.3 Open-world protocol

The experimental results observed for the open-world scenario are presented in Table 5.11 and Table 5.12 for the PolyU Cross-Spectral and CROSS-EYED databases, respectively. Notice that this protocol is more challenging since there is no sample of the test classes in the training set. Another factor that makes it more difficult is that fewer images are available for model training compared to the closed-world protocol, and there are more images on the test set, increasing the pair of genuine and imposter comparisons.

Table 5.11: Verification in the open-world protocol on the PolyU Cross-Spectral database. Extracted from [24].

Approach	Modality	EER (%)	Decidability
Cross-spectral			
Proposed ResNet50	Iris	12.01 ± 0.78	2.44 ± 0.08
Proposed ResNet50	Periocular	8.02 ± 0.65	3.00 ± 0.11
Proposed ResNet50	Fusion	6.01 ± 0.39	3.35 ± 0.08
VIS vs VIS			
Proposed ResNet50	Iris	4.30 ± 0.24	3.86 ± 0.07
Proposed ResNet50	Periocular	3.94 ± 0.27	4.14 ± 0.09
Proposed ResNet50	Fusion	2.61 ± 0.11	4.71 ± 0.06
NIR vs NIR			
Proposed ResNet50	Iris	4.00 ± 0.24	3.88 ± 0.08
Proposed ResNet50	Periocular	4.00 ± 0.26	4.10 ± 0.10
Proposed ResNet50	Fusion	2.55 ± 0.17	4.68 ± 0.10

Table 5.12: Results - open-world protocol on the CROSS-EYED database. Extracted from [24].

Approach	Modality	EER (%)	Decidability
Cross-spectral			
Proposed ResNet50	Iris	8.87 ± 0.77	2.85 ± 0.11
Proposed ResNet50	Periocular	4.39 ± 0.44	3.85 ± 0.11
Proposed ResNet50	Fusion	3.51 ± 0.32	4.17 ± 0.07
VIS vs VIS			
Proposed ResNet50	Iris	4.25 ± 0.35	4.01 ± 0.10
Proposed ResNet50	Periocular	3.41 ± 0.38	4.41 ± 0.11
Proposed ResNet50	Fusion	2.57 ± 0.26	4.97 ± 0.09
NIR vs NIR			
Proposed ResNet50	Iris	5.04 ± 0.43	3.63 ± 0.12
Proposed ResNet50	Periocular	3.51 ± 0.40	4.38 ± 0.12
Proposed ResNet50	Fusion	2.75 ± 0.28	4.83 ± 0.10

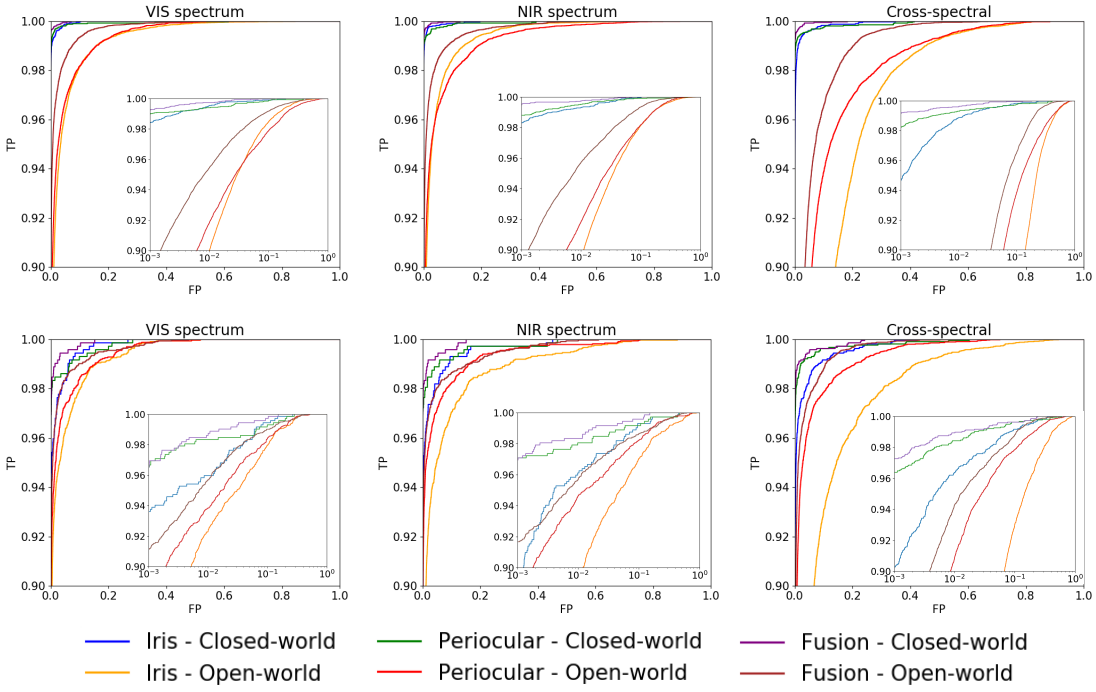


Figure 5.3: ROC curves comparing the closed- and open-world protocols on the PolyU Cross-Spectral (top row) and CROSS-EYED (bottom row) databases. Extracted from [24].

To perceive the differences in performance, a comparison of the results using closed- and open-world is shown with the ROC curve in Figure 5.3. Even though a fully fair comparison between closed- and open-world protocols is not feasible because the number of subjects used for learning is different, it is noticeable that the open-world protocol reported worse performance in all modes than the closed-world protocol. Nevertheless, we conclude that fusing the ocular and iris representations also leads to promising results in the open-world protocol, given that the observed decidability was higher than three for both databases considered.

5.2.4 ResNet-50: Performance vs. Network Depth

Having concluded that the ResNet-50 yields the optimal results in terms of EER in our experiments, our next goal was to perceive how the verification performance varies concerning the depth of the layer from where representations are taken. In this experiment, we considered all the convolution layers with stride equal to 2, resulting in four different depths to be tested: 12, 24, 42, and 50 layers. For each of the four possibilities (depths), the same modifications described in the methodology section were made, adding a fully-connected layer with 256 neurons and a layer with a softmax cross-entropy loss function. The verification results using the different depths are reported in Table 5.13 for the PolyU Cross-Spectral and CROSS-EYED databases.

Table 5.13: EER values observed for different depths (trainable parameters) of ResNet-50 architecture, using the closed-world protocol. Extracted from [24].

Spec.	Trait	12 layers (26M)	24 layers (14.5M)	42 layers (15.6M)	50 layers (24.1M)
PolyU Cross-Spectral					
VIS	Iris	3.21 ± 0.16	2.29 ± 0.15	1.60 ± 0.10	0.78 ± 0.08
	Perioc.	3.84 ± 0.14	3.17 ± 0.18	2.17 ± 0.12	0.61 ± 0.11
	Fusion	1.66 ± 0.06	1.41 ± 0.07	1.06 ± 0.11	0.35 ± 0.06
NIR	Iris	3.55 ± 0.18	2.36 ± 0.11	1.46 ± 0.10	0.68 ± 0.07
	Perioc.	4.16 ± 0.17	3.39 ± 0.18	2.27 ± 0.14	0.68 ± 0.10
	Fusion	2.13 ± 0.08	1.56 ± 0.08	1.09 ± 0.10	0.40 ± 0.06
Cross	Iris	6.39 ± 0.41	4.50 ± 0.23	3.09 ± 0.19	1.13 ± 0.14
	Perioc.	5.38 ± 0.20	4.04 ± 0.17	2.71 ± 0.14	0.78 ± 0.09
	Fusion	2.95 ± 0.15	2.07 ± 0.13	1.41 ± 0.09	0.49 ± 0.06
CROSS-EYED					
VIS	Iris	4.77 ± 0.38	3.29 ± 0.26	2.16 ± 0.34	2.47 ± 0.42
	Perioc.	6.34 ± 0.36	3.70 ± 0.35	1.90 ± 0.23	1.70 ± 0.37
	Fusion	3.78 ± 0.22	1.94 ± 0.16	1.25 ± 0.18	1.17 ± 0.25
NIR	Iris	20.24 ± 0.70	16.28 ± 0.66	8.78 ± 0.56	2.74 ± 0.34
	Perioc.	7.28 ± 0.35	4.08 ± 0.32	1.88 ± 0.23	1.78 ± 0.39
	Fusion	7.78 ± 0.30	4.90 ± 0.33	2.03 ± 0.23	1.31 ± 0.24
Cross	Iris	20.88 ± 0.74	15.91 ± 0.60	8.12 ± 0.63	3.07 ± 0.38
	Perioc.	7.53 ± 0.38	4.17 ± 0.38	2.31 ± 0.31	1.95 ± 0.35
	Fusion	8.29 ± 0.46	4.43 ± 0.29	2.14 ± 0.24	1.40 ± 0.26

It can be observed that the largest degradation of the results occurred when using shallow models occurs in the CROSS-EYED database. In all cases, the VIS against VIS comparison reports the best results, and it is the scenario where it presents the lowest degradation of the response in the different depths of the model.

As shown in the NIR against NIR and Cross-spectral results in the CROSS-EYED database, some EER values in the fusion of traits is higher than the ones using information from the periocular region only. This behavior is due to the weight used in the fusion of features where the low discrimination of the iris region penalizes and degrades the fused matching score, as we discuss in Section 5.2.2.

The experiments performed by Nguyen et al. [93] show that features extracted from intermediate layers of the networks achieved better results compared to deep layer representations. However, our results report lower EER rates using features extracted from deeper layers. It is important to point out that in [93] the ResNet152 model (i.e., a deeper model than ResNet50, used in our work) was employed. The same behavior can be observed in work by Henandez-Diaz et al. [259], where the authors stated that features extracted from the intermediate layers of the

ResNet-101 model reported the best results. Thus, the deepest layer reported in this work is approximately at the same depth as the intermediate layer reported by Nguyen et al. [93] and by Hernandez-Diaz et al. [259]. In another work, Hernandez-Diaz et al. [258] reported that using the ResNet50 model, representations from the intermediate layers achieved better results in the UBIPr Periocular database [16]. Oppositely, in this work, periocular representations extracted from the last layer of the ResNet50 model achieved the best results. Notice the UBIPr database has some larger images (from 501×401 pixels (8m) to 1001×801 (4m)) than PolyU Cross-Spectral and CROSS-EYED databases and also the periocular region is more extensive, containing eyebrows information, which can explain why a shallow model can extract more discriminant features from the intermediate layers, in this case.

As described in [29], a disadvantage of the VGG16 model, when compared to ResNet, is its larger number of trainable parameters (98.6M, when compared to their CNN with SD methodology 0.6M). As before stated, in our case, the best responses were observed when using the ResNet50 model, which after the modifications has 24.1M (four times lower compared to VGG16). As shown in Table 5.13, smaller networks in terms of depth lead to increasingly high losses in performance, however also decreasing nearly 10M training parameters, which can be an interesting solution for embedded systems and other cases where the computational complexity might be a concern. The ResNet with 12 layers has more trainable parameters than the other models since it considers an input image of 28×28 pixels and 128 filters. Besides, its convolutional part is connected with a fully connected layer containing 256 neurons added to reduce feature dimensionality.

5.2.5 Subjective evaluation

To provide some insight into the weaknesses of the solutions proposed in our work and a basis for subsequent improvements in the technology, this section highlights some notable cases of image pairwise comparisons that led to the best/worst performance (using the closed-world protocol). Results are shown in Figure 5.4, grouped into the worst genuine (when the system rejected a true matching) and the best impostors (when the system accepted a false matching) comparisons.

Although Figure 5.4 only shows VIS images, we noticed that pose and gaze are factors that can lead to matching errors also in NIR against NIR and cross-spectral scenarios. We observed that there were also confusions in images of the same subject but from different classes (left and right eyes) no matter the spectral scenario. Thus, we believe that it is possible to improve the recognition system accuracy using information based on the angle of the periocular region images and performing a preprocessing to determine the left and right eyes (i.e., a soft biometrics process). Also, based on the pairwise comparison errors, we can state that another factor that may improve system accuracy is the process of centralization/resizing of the periocular image based on the iris region size and location, similar to the method proposed by Hernande-Diaz et al. [258].

5.2.6 Final Considerations

The experiments showed that the models learned on the ResNet-50 architecture reported best results in terms of EER than its VGG counterpart, both in the PolyU Cross-Spectral and CROSS-EYED databases. Interestingly, we note that even this simple processing chain was observed to advance the state-of-the-art results in both databases.

Overall, in most of the experiments, features taken from the periocular region were observed to provide better performance than iris features. However, the fusion of these two traits reported better EER and decidability index than the best individual trait.

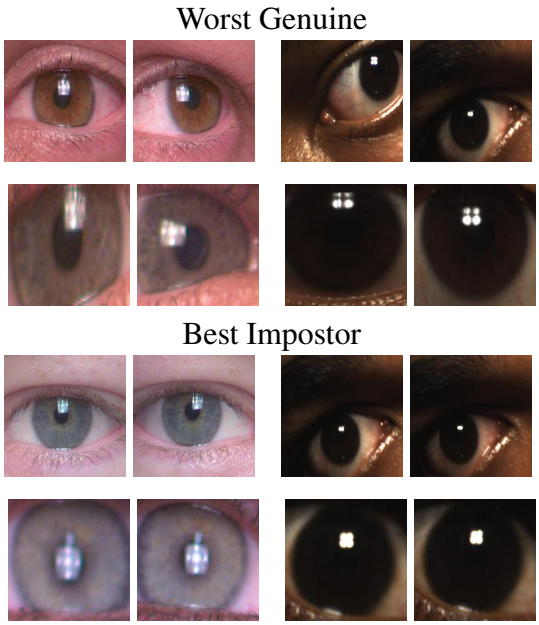


Figure 5.4: Pairwise comparison errors in the VIS against VIS scenario on CROSS-EYED (left) and PolyU Cross-Spectral (right) databases. Periocular and iris matching modalities are presented at Top and Bottom rows, respectively. Extracted from [24].

Finally, our subjective analysis of the best/worst false genuine and true impostors image pairwise comparisons showed that factors such as the angle of image capture might interfere with the recognition system's accuracy. In this direction, it is interesting to investigate how to build representation taking into account eye gaze and pose.

5.3 ATTRIBUTE NORMALIZATION FOR UNCONSTRAINED PERIOCULAR RECOGNITION

This section presents the evaluation of our proposed attribute normalization process to reduce the intra-class variability in periocular images captured under unconstrained environments.

The first step in our normalization strategy was training the AttGAN model for ocular attribute editing using periocular images. For the eyeglasses normalization (removal), we employed the entire UBIPr database in the training stage. Then, we normalized all the images from the UFPR-eyeglasses database by removing the eyeglasses. For the eye gaze normalization, we trained the Att-GAN using images from the first half of the subjects from the UBIPr database and normalized all images from the second half of the subjects by correcting the eye gaze. The Deep learning-based approaches were trained using the first half of the subjects for both databases. The second half of the subjects were used to evaluate and compare handcrafted features and deep learning approaches using original and normalized images. Some qualitative results of the attribute normalization using the AttGAN model are shown in Fig. 5.5.

For the recognition performance evaluation, according to the conclusions we previously drew about distance measures in ocular representations [46, 24], we chose to use the cosine distance metric to match both deep learning-based and handcrafted approaches. Regarding the SIFT features matching, we used the ratio test proposed by Lowe [263].

We started by generating pairwise comparisons considering only images with different attributes, i.e., pairs with eyeglasses/no-eyeglasses in the UFPR-Eyeglasses database and pairs with different gazes, the case of the UBIPr database. Using the second half of the subjects for

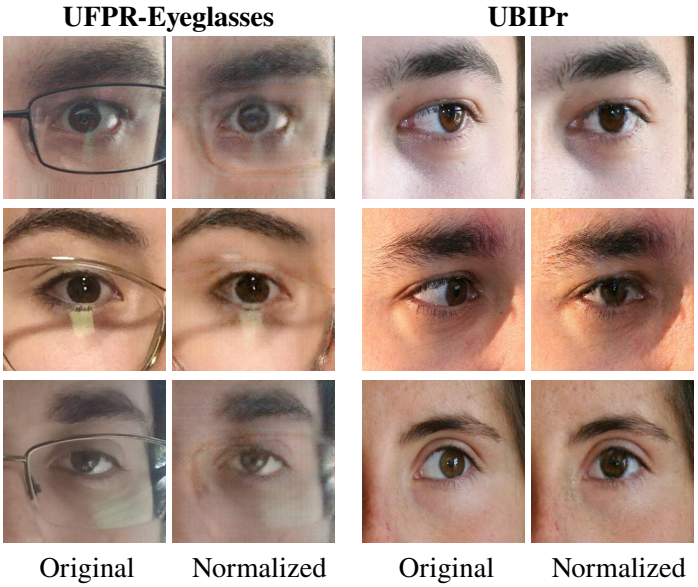


Figure 5.5: Examples of original and normalized images from the UFPR-Eyeglasses (Eyeglasses removal) and UBIPr (Eyegaze correction) databases. Extracted from [25].

each database, we applied the all-against-all protocol, generating 3,072 genuine and 274,464 impostor pairs for the UFPR-Eyeglasses database and 22,012 genuine / 6,246,232 impostors pairs for the UBIPr database.

Considering a verification task, we used the Decidability index and the Area Under the Curve (AUC) as metrics to evaluate the methods. As the proposed normalization aims to decrease the intra-class variability, we considered Decidability as the primary metric. The results achieved with the proposed attribute normalization are shown in Table 5.14 for the UFPR-Eyeglasses and UBIPr databases. Note that we compared the results of the methods using the original and normalized images to evaluate better the improvements in performance concerning the solution described in this work.

Table 5.14: Comparison of results using original and normalized images in the UFPR-Eyeglasses and UBIPr databases. Adapted from [25].

Method - Features	Att. Normalization	UFPR-Eyeglasses		UBIPr	
		AUC (%)	Decidability	AUC (%)	Decidability
Ahmed et al. [122]		73.0	0.77	84.9	1.16
	✓	**73.2**	**0.79**	**85.2**	**1.17**
Park et al. [115]		78.8	1.11	**89.6**	**1.73**
	✓	**85.2**	**1.43**	87.8	1.62
LBP + LPQ + HOG + SIFT		75.9	0.92	**90.2**	1.71
	✓	**87.2**	**1.58**	90.0	**1.77**
Luz et al. [46]		85.9	1.57	**98.3**	**3.64**
	✓	**89.0**	**1.81**	98.1	3.50
Zanlorensi et al. [24]		92.2	2.09	99.2	4.00
	✓	**92.9**	**2.16**	**99.4**	**4.14**

The results showed that the proposed normalization preprocessing consistently improve the verification results in the UFPR-Eyeglasses database, increasing the Decidability by 28% (i.e., 1.4261/1.1093) and 71% (i.e., 1.5764/0.9206), respectively using the features from the method

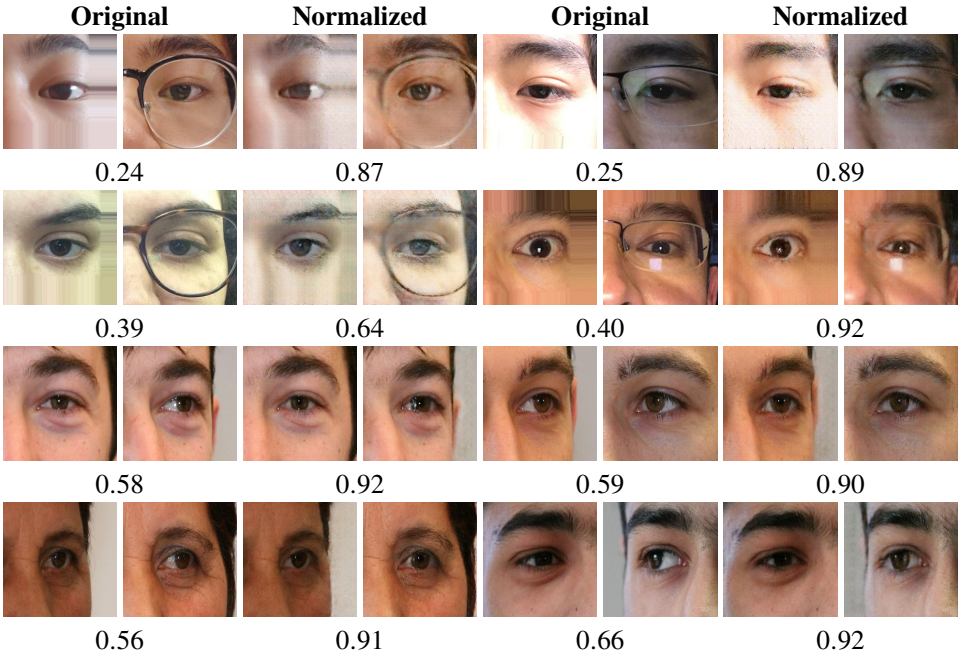


Figure 5.6: Genuine scores comparison from original and normalized images. Higher scores mean that the periocular image pairwise is more likely to be genuine. Extracted from [25].

proposed by Park et al. [115] and from the proposed handcrafted features fusion. Using the deep learning-based approaches, the attribute normalization improved the Decidability by 15% and 4% for the methods proposed by Luz et al. [46] and Zanlorensi et al. [24], respectively. Unlike the experiments performed using the UFPR-Eyeglasses database, in the UBIPr one, the attribute normalization process consists of the eye gaze correction. Since this process is computed in a small portion of the periocular image (only in the eyeball region), in general, we can observe that the impact of applying the attribute normalization is smaller than the ones obtained in the UFPR-Eyeglasses images. Nevertheless, the highest Decidability index in the UBIPr database using handcrafted features and Deep learning-based models was achieved by employing the normalized images. Fig. 5.6 shows some qualitative results where wrong genuine matching between original images was corrected using the proposed attribute normalization.

One can also observe that in the UFPR-Eyeglasses database, even when the eyeglasses were not entirely removed, the generative model was able to smooth them, such that the biometric system was able to correctly classified a pair as genuine. Investigating other wrong genuine matches, we stated that the pose and illumination aspect is one of the most significant factors that penalize the intra-class variability in the UBIPr database.

5.3.1 Final Considerations

The idea of this investigation was to employ a state-of-the-art generative model that normalizes specific factors of all samples before being used by the recognition algorithm. The proposed solution is fully agnostic to the recognition method used, and our proof-of-concept was conducted in two databases and five different baseline methods. We compare the performance levels attained by the recognition methods when using the raw data and when receiving the images preprocessed by our solution. The observed results corroborated our hypothesis that the proposed attribute normalization is highly effective in reducing the intra-class variabilities without compromising the discriminability between classes, which is the root for the observed improvements in performance.

5.4 UFPR-PERIOCULAR DATABASE AND SOFT-BIOMETRICS

This section presents the results obtained by each approach in the closed-world and open-world protocols and an ablation study on the Multi-task learning network to evaluate each task's influence in the identification mode. First, we show in Table 5.15 the size and the number of trainable parameters of each CNN model used as a benchmark. This information is from the models that we used on the closed-world protocol since they have more neurons on the last layer than the open-world protocol models.

Table 5.15: Size (MB) and number of trainable parameters of the CNN models used in the benchmark.

Model	Size (MB)	Trainable parameters
VGG16	1088	135,886,084
VGG16-Face	1088	135,886,084
InceptionResNet	445	55,246,372
ResNet50V2	400	49,786,436
ResNet50	198	24,609,284
ResNet50-Face	198	24,609,284
Xception	176	21,908,204
DenseNet121	64	7,792,964
MobileNetV2	26	3,128,516
Multi-task	37	4,494,230
Siamese	21	2,551,808
Pairwise	20	2,349,479

As can be seen, the benchmark has a great diversity of models with different sizes and parameters due to their difference in structure, depth, concept, and architectures.

5.4.1 Closed-world protocol

In the closed-world protocol, we perform the benchmark for both the identification and verification tasks. All results are presented in Table 5.16. As can be seen, although MobileNetV2 is the smallest model in terms of size and trainable parameters, it achieved the best results for both identification and verification tasks. Hence, we used MobileNetV2 as the base model for the Multi-task, Siamese, and Pairwise Filters networks.

In general, the Multi-task model achieved the best results in terms of Rank 1, Rank 5, AUC, and EER. We highlight that we only explored the other tasks – age, gender, eye side, and mobile device model – at this model's training stage. For the evaluation, we extracted the representations for the classification task and used them for identification (using the softmax layer) and verification (using the cosine distance) tasks. The Siamese network obtained the worst results in the benchmark. In contrast, the Pairwise Filters network reached the higher Decidability index, indicating that it was best to separate genuine and impostors distributions. However, it did not achieve the best results in terms of AUC and EER.

As stated in some previous works [46, 266], the models pre-trained for face recognition generally achieve the best results than those pre-trained on the ImageNet database.

5.4.2 Open-world protocol

The main idea of employing the open-world protocol was to evaluate the methods to extract discriminant features from samples of classes that are not present in the training stage. Thus, for

Table 5.16: Benchmark results in the closed-world protocol for the identification and verification tasks.

Model	Identification (1:*N*)		Verification (1:1)		
	Rank 1 (%)	Rank 5 (%)	AUC (%)	EER (%)	Decidability
VGG16	50.56 ± 3.30	68.73 ± 3.01	99.41 ± 0.11	3.59 ± 0.32	4.4544 ± 0.1502
VGG16-Face	56.29 ± 1.62	73.84 ± 1.48	99.43 ± 0.08	3.44 ± 0.28	4.5069 ± 0.1379
Xception	57.43 ± 1.43	75.88 ± 1.52	99.77 ± 0.04	2.19 ± 0.18	4.2470 ± 0.0538
ResNet50V2	63.18 ± 2.14	77.79 ± 1.81	99.74 ± 0.04	2.24 ± 0.18	4.9382 ± 0.1184
InceptionResNet	65.16 ± 2.45	81.53 ± 1.99	99.78 ± 0.15	1.85 ± 0.40	4.5561 ± 0.1183
ResNet50	71.06 ± 1.14	85.22 ± 0.82	99.89 ± 0.02	1.41 ± 0.10	5.1242 ± 0.0634
ResNet50-Face	73.76 ± 1.43	86.86 ± 1.02	99.83 ± 0.03	1.74 ± 0.12	5.2400 ± 0.0837
DenseNet121	75.54 ± 1.36	88.53 ± 0.97	99.93 ± 0.02	1.11 ± 0.09	5.1730 ± 0.0497
MobileNetV2	77.98 ± 1.08	90.19 ± 0.79	99.93 ± 0.01	1.13 ± 0.07	5.2477 ± 0.0650
Multi-task	**84.32 ± 0.71**	**94.55 ± 0.58**	**99.96 ± 0.01**	**0.81 ± 0.06**	5.1978 ± 0.0340
Siamese	–	–	98.94 ± 0.22	4.86 ± 0.44	3.0005 ± 0.1871
Pairwise	–	–	99.44 ± 0.66	3.06 ± 1.84	**6.4503 ± 1.2270**

this protocol, we performed a benchmark only for the verification task. The results are shown in Table 5.17.

Table 5.17: Benchmark results in the open-world protocol for the verification task.

Model	Validation	Verification (1:1)		
		AUC (%)	EER (%)	Decidability
VGG16	Closed-World	97.38 ± 0.53	8.52 ± 0.92	2.9599 ± 0.1572
VGG16-Face	Closed-World	97.70 ± 0.42	7.78 ± 0.75	3.0327 ± 0.1428
ResNet50	Closed-World	98.60 ± 0.28	5.98 ± 0.67	3.3702 ± 0.1413
ResNet50V2	Closed-World	98.73 ± 0.28	5.69 ± 0.64	3.4312 ± 0.1459
Xception	Closed-World	98.93 ± 0.16	5.23 ± 0.42	3.3493 ± 0.0712
InceptionResNet	Closed-World	99.10 ± 0.24	4.61 ± 0.65	3.4982 ± 0.1208
ResNet50-Face	Closed-World	99.18 ± 0.16	4.38 ± 0.47	3.8319 ± 0.1239
DenseNet121	Closed-World	99.51 ± 0.12	3.39 ± 0.46	3.8646 ± 0.1215
MobileNet	Closed-World	99.56 ± 0.08	3.17 ± 0.33	3.9868 ± 0.1067
Multi-task	**Closed-World**	**99.67 ± 0.08**	**2.81 ± 0.39**	3.9263 ± 0.0921
Siamese	Closed-World	97.27 ± 0.64	8.10 ± 1.01	2.6678 ± 0.2433
Pairwise	Closed-World	98.62 ± 0.72	5.77 ± 1.57	**4.4404 ± 0.5834**
Siamese	Open-World	96.85 ± 0.70	8.87 ± 1.14	2.6218 ± 0.1514
Pairwise	Open-World	97.80 ± 2.03	7.11 ± 3.66	4.1977 ± 1.0663

Like the closed-world protocol, the Multi-task model achieved the best results in Rank 1, Rank 5, AUC, and EER, and the Pairwise network achieved the best Decidability index. The Siamese and Pairwise Filters networks trained using the closed-world validation split reached better results than when trained using the open-world validation split. We believe this occurred because the open-world validation split's training has samples of fewer classes than in the closed-world validation split.

Although the open-world validation split corresponds to a more realistic scenario regarding the test set, the networks trained with samples from a larger number of classes can reach a higher capability of generalization, producing discriminative representations even for samples from classes that are not present in the training stage.

5.4.3 Multi-task Learning

The Multi-task model achieved the best results both in the closed- and open-world protocols. As this network simultaneously learns different tasks, we perform an ablation study by running some experiments with 4 new models created by removing one of the tasks at a time. The experiments were carried out in the closed-world protocol to evaluate the performance of both identification and verification. We also evaluated the results achieved by all models in each task.

Table 5.18: Results (%) from several Multi-task models trained to predict different tasks.

Model	Rank 1	Rank 5	Device Model	Age	Gender	Eye Side
Multi-task (no model)	80.76 ± 0.94	91.96 ± 0.51	–	82.14 ± 0.83	97.72 ± 0.17	**99.99 ± 0.01**
Multi-task (no age)	81.93 ± 0.99	93.51 ± 0.69	87.20 ± 0.63	–	97.65 ± 0.20	**99.99 ± 0.01**
Multi-task (no gender)	82.48 ± 0.64	93.55 ± 0.52	86.71 ± 0.54	83.17 ± 0.54	–	**99.99 ± 0.01**
Multi-task (no side)	83.72 ± 0.61	94.07 ± 0.54	87.22 ± 0.79	83.75 ± 0.53	97.70 ± 0.20	–
Multi-task	**84.32 ± 0.71**	**94.55 ± 0.58**	**87.42 ± 0.65**	**84.34 ± 0.71**	**97.80 ± 0.21**	99.98 ± 0.02

According to Table 5.18, the Multi-task network without the prediction of the mobile device model was the most penalized for the identification task, followed by the network variations without age, gender, and eye side estimation, respectively. The gender and eye side classification tasks were handled well by all models, while the device model and age range classification tasks proved to be more challenging. One problem in the device model and age range classification is the unbalanced number of samples per class, which can generate a bias during the training stage.

Note that in both closed-world and open-world protocols, we only explored the class prediction for the matching. However, as shown in Table 5.18, the multi-task architecture also achieved promising results in the other tasks. In this sense, it may be possible to further improve the recognition results by adopting heuristic rules based on the scores of the other tasks.

5.4.4 Subjective evaluation

In this section, we perform a subjective evaluation through visual inspection on the pairs of images erroneously classified by the Multi-task model, which achieved the best result in the verification task in the closed-world protocol. The best impostors (impostors classified as genuine) and the worst genuines (genuine classified as impostors) pairs are presented in Fig. 5.7.

Performing a visual analysis of all pairwise errors, it is clear that hair occlusion, age, eyeglasses, and eye shape were the most influential factors that led the model to the wrong classification of genuine pairs (intra-class comparison). In pairs wrongly classified as impostors (inter-class comparison), we saw that lighting, blur, eyeglasses, off-angle, eye-gaze, reflection, and facial expression caused the main difference between the images. We hypothesize that some errors caused by lightning, blur, reflection, and occlusion can be reduced by employing some data augmentation techniques in the training stage. Attribute normalization [25] can also reduce the errors caused by attributes present in the periocular region such as eyeglasses, eye gaze, makeup, and some types of occlusion. Although some methods can be applied to reduce the matching errors, there are still several characteristics in these images that make the mobile periocular recognition a challenging task, mainly to the high intra-class variations.

5.4.5 Final Considerations

This research aimed to create a database with real-world images regarding lighting, noises, and attributes in the periocular region. To the best of our knowledge, in the literature, this is the first

Figure 5.7: Pairwise images wrongly classified by the model that obtained the best result in the verification task in the open-world protocol. Higher scores mean that the pair of periocular images is more likely to be genuine.

periocular database with more than 1,000 subject samples and the largest one in the number of different sensors (196).

We presented an extensive benchmark with several CNN models and architectures employed in recent works for ocular recognition. These architectures consist of models for Multi-class classification and Multi-task Learning, in addition to Siamese and Pairwise Filters networks. We evaluated the methods in the closed-world and open-world protocols for the identification and verification tasks. For both protocols and tasks, the Multi-task model achieved the best results. Thus, we conducted an ablation study on this model to understand which tasks had the most significant influence on the results. We stated that the mobile device model identification task was the most important, followed by age range, gender, and eye side classification. The model trained using all these tasks reported the best result for the identification and verification in the closed- and open-world protocols.

In a complementary way, we performed a subjective analysis of the best/worst false genuine and true impostors image pairwise comparisons using the Multi-task model, which achieved the best performance for the verification task. We observed that lighting, occlusion, and image resolution were the most critical factors that led the model to wrong verification.

We believe that the UFPR-Periocular database will be of great relevance to assist in evolving ocular biometric systems using images obtained by mobile devices in unconstrained

scenarios. This database is the most extensive in terms of the number of subjects in the literature and has natural intra-class variability due to samples captured in different sessions.

The Multi-task network using the MobileNetV2 as baseline model achieved the best benchmark results for the identification and verification tasks, reaching a rank 1 of 84, 32% and an EER of 0.81% in the closed-world protocol, and an EER of 2.81% in the open-world protocol. Therefore, there is still room for improvement in both identification and verification tasks.

6 CONCLUSION

In this thesis, we explored and investigated deep representations for iris and periocular recognition. Considering the hypothesis that it is possible to achieve state-of-the-art results by employing deep learning techniques at different stages of ocular biometric systems based on periocular and iris traits, we proposed and evaluated several approaches achieving state-of-the-art results for ocular recognition in different scenarios.

First, we investigated the impact of preprocessing steps on iris recognition. Performing an ablation study on the preprocessing steps as segmentation and normalization on controlled and uncontrolled iris databases, we stated that it is possible to directly use an iris bounding box as input to CNN models to extract iris deep representations. The proposed method achieved the state-of-the-art EER in the NICE.II contest database. Regarding cross-spectral ocular recognition, we performed extensive experiments on two publicly available databases showing that CNN models can directly learn representation from NIR and VIS images for both iris and periocular traits. The proposed method reached state-of-the-art results in both databases using the iris trait and significantly decrease the EER by fusing iris and periocular region. As we stated in previous work, some subjects' noninherent attributes, e.g., eyeglasses and eye gaze, usually increase intra-class variability. Thus, we proposed an attribute normalization method to handle this problem. The attribute normalization method proved to be effective for both handcrafted features and deep representations. Finally, we collected a new periocular database comprising images from mobile devices under unconstrained environments. Employing this database, we proposed a multitask model using soft biometrics information in the training stage, improving the periocular deep representation's discriminability. We also performed an extensive benchmark of the most recent CNN architectures that have been employed to build ocular biometric systems.

Supported by our investigation, experiments, and results, we can state that deep learning techniques applied to ocular recognition for both the iris and periocular traits can achieve impressive results even in unconstrained and uncontrolled environments. However, there is still room for improvements since there are complex and open problems related to the methods' scalability, multimodal biometric fusion, multi-session (intra-class variability), cross-sensor and cross-spectral images, and different protocols (closed and open-world, and cross-database).

REFERENCES

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

[2] CASIA. Casia version 4 database. `http://biometrics.idealtest.org/dbDetailForUser.do?id=4`, 2010. Acessado em 2021-02-02.

[3] A. K. Jain and A. Ross. *Introduction to Biometrics*, pages 1–22. Springer US, Boston, MA, 2008.

[4] J. Daugman. How iris recognition works. *IEEE Trans. on Circuits and Systems for Video Technology*, 14(1):21–30, 2004.

[5] S. Albelwi and A. Mahmood. A framework for designing the architectures of deep convolutional neural networks. *Entropy*, 19(6), 2017.

[6] K. He, X. Zhang, S. Ren, and J. Sun. Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, jun 2016.

[7] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning. In *ICLR 2016 Workshop*, 2016.

[8] J. S. Doyle and K. W. Bowyer. Robust Detection of Textured Contact Lenses in Iris Recognition Using BSIF. *IEEE Access*, 3:1672–1683, 2015.

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

[10] D. Yadav, N. Kohli, J. S. Doyle, R. Singh, M. Vatsa, and K. W. Bowyer. Unraveling the effect of textured contact lenses on iris recognition. *IEEE Transactions on Information Forensics and Security*, 9(5):851–862, 2014.

[11] J. Doyle and K. Bowyer. Notre dame image database for contact lens detection in iris recognition. `https://cvrl.nd.edu/projects/data/#nd-cosmetic-contact-lenses-2013-data-set`, 2014. Acessado em 2021-02-02.

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

[13] L. A. Zanlorensi, R. Laroca, E. Luz, A. S. Britto Jr., L. S. Oliveira, and D. Menotti. Ocular recognition databases and competitions: A survey. *Artificial Intelligence Review*, pages 1–52, 2021.

[14] A. Rattani, R. Derakhshani, S. K. Saripalle, and V. Gottemukkula. ICIP 2016 competition on mobile ocular biometric recognition. In *IEEE International Conference on Image Processing (ICIP) 2016, Challenge Session on Mobile Ocular Biometric Recognition*, pages 320–324, Phoenix, AZ, USA, Sep. 2016. IEEE.

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

[16] C. N. Padole and H. Proença. Periocular recognition: Analysis of performance degradation factors. In *IAPR International Conference on Biometrics (ICB)*, pages 439–445, New Delhi, India, mar 2012. IEEE.

[17] L. A. Zanlorensi, R. Laroca, D. R. Lucio, L. R. Santos, A. S. Britto Jr., and D. Menotti. UFPR-Periocular: A periocular dataset collected by mobile devices in unconstrained scenarios. *arXiv preprint*, arXiv:2011.12427:1–12, 2020.

[18] A. Sequeira, L. Chen, P. Wild, J. Ferryman, F. Alonso-Fernandez, K. B. Raja, R. Raghavendra, C. Busch, and J. Bigun. Cross-Eyed - Cross-Spectral Iris/Periocular Recognition Database and Competition. In *2016 International Conference of the Biometrics Special Interest Group (BIOSIG)*, volume 260, pages 1–5, Darmstadt, Germany, sep 2016. IEEE.

[19] A. F. Sequeira, L. Chen, J. Ferryman, P. Wild, F. Alonso-Fernandez, J. Bigun, K. B. Raja, R. Raghavendra, C. Busch, T. de Freitas Pereira, S. Marcel, S. S. Behera, M. Gour, and V. Kanhangad. Cross-eyed 2017: Cross-spectral iris/periocular recognition competition. In *IEEE International Joint Conference on Biometrics*, pages 725–732, Denver, CO, USA, Oct 2017. IEEE.

[20] P. R. Nalla and A. Kumar. Toward more accurate iris recognition using cross-spectral matching. *IEEE Transactions on Image Processing*, 26(1):208–221, jan 2017.

[21] A. F. Sequeira, J. C. Monteiro, A. Rebelo, and H. P. Oliveira. MobBIO: A multimodal database captured with a portable handheld device. In *International Conference on Computer Vision Theory and Applications (VISAPP)*, volume 3, pages 133–139, Lisbon, Portugal, Jan 2014. IEEE.

[22] Y. Yin, L. Liu, and X. Sun. Sdumla-hmt: A multimodal biometric database. In Zhenan Sun, Jianhuang Lai, Xilin Chen, and Tieniu Tan, editors, *Biometric Recognition*, pages 260–268, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.

[23] L. A. Zanlorensi, E. Luz, R. Laroca, A. S. Britto Jr., L. S. Oliveira, and D. Menotti. The impact of preprocessing on deep representations for iris recognition on unconstrained environments. In *Conference on Graphics, Patterns and Images (SIBGRAPI)*, pages 289–296, Oct 2018.

[24] L. A. Zanlorensi, D. R. Lucio, A. S. Britto Jr., H. Proença, and D. Menotti. Deep representations for cross-spectral ocular biometrics. *IET Biometrics*, 9:68–77, 2020.

[25] L. A. Zanlorensi, H. Proença, and D. Menotti. Unconstrained periocular recognition: Using generative deep learning frameworks for attribute normalization. In *2020 International Conference on Image Processing (ICIP)*, pages 1361–1365, October 2020.

[26] Z. He, W. Zuo, M. Kan, S. Shan, and X. Chen. Attgan: Facial attribute editing by only changing what you want. *IEEE Transactions on Image Processing*, 28(11):5464–5478, Nov 2019.

[27] M. De Marsico, M. Nappi, and H. Proença. Results from MICHE II – Mobile Iris CHallenge Evaluation II. *Pattern Recognition Letters*, 91:3–10, 2017.

[28] H. Nguyen, N. Reddy, A. Rattani, and R. Derakhshani. *VISOB 2.0 - second international competition on mobile ocular biometric recognition*, pages 1–8. Springer International Publishing, 2021.

[29] K. Wang and A. Kumar. Cross-spectral iris recognition using CNN and supervised discrete hashing. *Pattern Recognition*, 86:85–98, 2019.

[30] J. Daugman. Probing the uniqueness and randomness of iriscodes: Results from 200 billion iris pair comparisons. *Proceedings of the IEEE*, 94(11):1927–1935, Nov 2006.

[31] P. J. Phillips, P. J. Flynn, J. R. Beveridge, W. T. Scruggs, A. J. O'Toole, D. Bolme, K. W. Bowyer, B. A. Draper, G. H. Givens, Y. M. Lui, H. Sahibzada, J. A. Scallan, and S. Weimer. Overview of the multiple biometrics grand challenge. In *Advances in Biometrics*, pages 705–714, Berlin, Heidelberg, 2009. Springer Berlin Heidelberg.

[32] K. W. Bowyer, K. Hollingsworth, and P. J. Flynn. Image understanding for iris biometrics: A survey. *Computer Vision and Image Understanding*, 110(2):281–307, 2008.

[33] R. P. Wildes. Iris recognition: an emerging biometric technology. *Proceedings of the IEEE*, 85(9):1348–1363, Sep 1997.

[34] H. Proença and J. C. Neves. IRINA: Iris recognition (even) in inaccurately segmented data. In *2017 IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 6747–6756, Jul 2017.

[35] H. Proença and J. C. Neves. A reminiscence of "mastermind": Iris/periocular biometrics by "in-set" CNN iterative analysis. *IEEE Transactions on Information Forensics and Security*, 14(7):1702–1712, July 2019.

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

[37] L. He, H. Li, F. Liu, N. Liu, Z. Sun, and Z. He. Multi-patch convolution neural network for iris liveness detection. In *2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pages 1–7. IEEE, sep 2016.

[38] Y. Du, T. Bourlai, and J. Dawson. Automated classification of mislabeled near-infrared left and right iris images using convolutional neural networks. In *IEEE International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pages 1–6, Sept 2016.

[39] J. Tapia and C. Aravena. Gender classification from nir iris images using deep learning. *IEEE Transactions on Information Forensics and Security*, pages 219–239, 2017.

[40] P. Silva, E. Luz, R. Baeta, H. Pedrini, A. X. Falcao, and D. Menotti. An approach to iris contact lens detection based on deep image representations. In *Conference on Graphics, Patterns and Images (SIBGRAPI)*, pages 157–164, Aug 2015.

[41] D. R. Lucio, R. Laroca, L. A. Zanlorensi, G. Moreira, and D. Menotti. Simultaneous iris and periocular region detection using coarse annotations. In *Conference on Graphics, Patterns and Images (SIBGRAPI)*, pages 178–185, Oct 2019.

[42] E. Severo, R. Laroca, C. S. Bezerra, L. A. Zanlorensi, D. Weingaertner, G. Moreira, and D. Menotti. A benchmark for iris location and a deep learning detector evaluation. In *International Joint Conference on Neural Networks (IJCNN)*, pages 1–7, Rio de Janeiro, Brazil, July 2018. IEEE.

[43] D. R. Lucio, R. Laroca, E. Severo, A. S. Britto Jr., and D. Menotti. Fully convolutional networks and generative adversarial networks applied to sclera segmentation. In *IEEE International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pages 1–7, Oct 2018.

[44] C. S. Bezerra, R. Laroca, D. R. Lucio, E. Severo, L. F. Oliveira, A. S. Britto Jr., and D. Menotti. Robust iris segmentation based on fully convolutional networks and generative adversarial networks. In *Conference on Graphics, Patterns and Images*, pages 281–288, Oct 2018.

[45] F. Marra, G. Poggi, C. Sansone, and L. Verdoliva. A deep learning approach for iris sensor model identification. *Pattern Recognition Letters*, 2017.

[46] E. Luz, G. Moreira, L. A. Zanlorensi Junior, and D. Menotti. Deep periocular representation aiming video surveillance. *Pattern Recognition Letters*, 114:2–12, 2018.

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

[48] P. H. Silva, E. Luz, L. A. Zanlorensi, D. Menotti, and G. Moreira. Multimodal feature level fusion based on particle swarm optimization with deep transfer learning. In *2018 Congress on Evolutionary Computation (CEC)*, pages 1–8, July 2018.

[49] K. Hernandez-Diaz, F. Alonso-Fernandez, and J. Bigun. Cross-spectral periocular recognition with conditional adversarial networks, 2020.

[50] K. H. Diaz, F. Alonso-Fernandez, and J. Bigun. Spectrum translation for cross-spectral ocular matching. *arXiv arXiv:2002.06228*, 2020.

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

[52] H. H. Proença and L. A. Alexandre. Toward covert iris biometric recognition: Experimental results from the NICE contests. *IEEE Transactions on Information Forensics and Security*, 7(2):798–808, apr 2012.

[53] A. Krishnan, A. Almadan, and A. Rattani. Probing fairness of mobile ocular biometrics methods across gender on visob 2.0 dataset. *arXiv preprint*, arXiv:2011.08898:1–15, 2020.

[54] A. Rattani, N. Reddy, and R. Derakhshani. Gender prediction from mobile ocular images: A feasibility study. In *2017 IEEE International Symposium on Technologies for Homeland Security (HST)*, pages 1–6, 2017.

[55] A. Rattani, N. Reddy, and R. Derakhshani. Convolutional neural networks for gender prediction from smartphone-based ocular images. *IET Biometrics*, 7(5):423–430, 2018.

[56] A. Kuehlkamp and K. Bowyer. Predicting gender from iris texture may be harder than it seems. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 904–912, 2019.

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

[58] N. Damer, J. H. Grebe, C. Chen, F. Boutros, F. Kirchbuchner, and A. Kuijper. The effect of wearing a mask on face recognition performance: an exploratory study. In *2020 International Conference of the Biometrics Special Interest Group (BIOSIG)*, pages 1–6, 2020.

[59] J. Han, D. Zhang, G. Cheng, N. Liu, and D. Xu. Advanced deep-learning techniques for salient and category-specific object detection: A survey. *IEEE Signal Processing Magazine*, 35(1):84–100, 2018.

[60] Z. Zhao, P. Zheng, S. Xu, and X. Wu. Object detection with deep learning: A review. *IEEE Transactions on Neural Networks and Learning Systems*, 30(11):3212–3232, 2019.

[61] D. V. Ruiz, B. A. Krinski, and E. Todt. Ida: Improved data augmentation applied to salient object detection. In *2020 33rd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, pages 210–217, 2020.

[62] B. A. Krinski, D. V. Ruiz, G. Z. Machado, and E. Todt. Masking salient object detection, a mask region-based convolutional neural network analysis for segmentation of salient objects. In *2019 Latin American Robotics Symposium (LARS), 2019 Brazilian Symposium on Robotics (SBR) and 2019 Workshop on Robotics in Education (WRE)*, pages 55–60, 2019.

[63] D. V. Ruiz, B. A. Krinski, and E. Todt. Anda: A novel data augmentation technique applied to salient object detection. In *2019 19th International Conference on Advanced Robotics (ICAR)*, pages 487–492, 2019.

[64] G. Hinton, L. Deng, D. Yu, G. Dahl, A. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. Sainath, and B. Kingsbury. Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups. *IEEE Signal Processing Magazine*, 29(6):82–97, nov 2012.

[65] Y. Zhang, W. Chan, and N. Jaitly. Very deep convolutional networks for end-to-end speech recognition. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4845–4849. IEEE, mar 2017.

[66] S. Kim, T. Hori, and S. Watanabe. Joint CTC-attention based end-to-end speech recognition using multi-task learning. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4835–4839. IEEE, mar 2017.

[67] Y. Tu, J. Du, and C. Lee. Speech enhancement based on teacher–student deep learning using improved speech presence probability for noise-robust speech recognition. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 27(12):2080–2091, 2019.

[68] W. Zhang, X. Cui, U. Finkler, B. Kingsbury, G. Saon, D. Kung, and M. Picheny. Distributed deep learning strategies for automatic speech recognition. In *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5706–5710, 2019.

[69] X. Glorot, A. Bordes, and Y. Bengio. Domain adaptation for large-scale sentiment classification: A deep learning approach. In *Proceedings of the 28th International Conference on International Conference on Machine Learning*, ICML'11, pages 513–520, USA, 2011. Omnipress.

[70] R. Socher, J. Pennington, E. H. Huang, A. Y. Ng, and C. D. Manning. Semi-supervised recursive autoencoders for predicting sentiment distributions. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, EMNLP '11, pages 151–161, Stroudsburg, PA, USA, 2011. Association for Computational Linguistics.

[71] J. Guo, H. He, T. He, L. Lausen, M. Li, H. Lin, X. Shi, C. Wang, J. Xie, S. Zha, A. Zhang, H. Zhang, Z. Zhang, Z. Zhang, S. Zheng, and Y. Zhu. Gluoncv and gluonnlp: Deep learning in computer vision and natural language processing. *Journal of Machine Learning Research*, 21(23):1–7, 2020.

[72] D. W. Otter, J. R. Medina, and J. K. Kalita. A survey of the usages of deep learning for natural language processing. *IEEE Transactions on Neural Networks and Learning Systems*, pages 1–21, 2020.

[73] J. H. Alves, P. M. M. Neto, and L. F. Oliveira. Extracting lungs from ct images using fully convolutional networks. In *2018 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, 2018.

[74] C. Q. Cordeiro, S. O. Ioshii, J. H. Alves, and L. F. Oliveira. An automatic patch-based approach for her-2 scoring in immunohistochemical breast cancer images using color features, 2018.

[75] R. Laroca, E. Severo, L. A. Zanlorensi, L. S. Oliveira, G. R. Gonçalves, W. R. Schwartz, and D. Menotti. A robust real-time automatic license plate recognition based on the YOLO detector. In *International Joint Conference on Neural Networks (IJCNN)*, pages 1–10, July 2018.

[76] G. R. Gonçalves, M. A. Diniz, R. Laroca, D. Menotti, and W. R. Schwartz. Real-time automatic license plate recognition through deep multi-task networks. In *Conference on Graphics, Patterns and Images (SIBGRAPI)*, pages 110–117, Oct 2018.

[77] R. Laroca, V. Barroso, M. A. Diniz, G. R. Gonçalves, W. R. Schwartz, and D. Menotti. Convolutional neural networks for automatic meter reading. *Journal of Electronic Imaging*, 28(1):013023, 2019.

[78] R. Laroca, A. B. Araujo, L. A. Zanlorensi, Eduardo C. de Almeida, and D. Menotti. Towards image-based automatic meter reading in unconstrained scenarios: A robust and efficient approach. *IEEE Access*, 9:67569–67584, 2021.

[79] R. Laroca, L. A. Zanlorensi, G. R. Gonçalves, E. Todt, W. R. Schwartz, and D. Menotti. An efficient and layout-independent automatic license plate recognition system based on the YOLO detector. *IET Intelligent Transport Systems*, 15(4):483–503, 2021.

[80] A. G. Hochuli, L. S. Oliveira, A. d. Souza Britto, and R. Sabourin. Segmentation-free approaches for handwritten numeral string recognition. In *2018 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, 2018.

[81] A. G. Hochuli, L. S. Oliveira, A. S. Britto Jr, and R. Sabourin. Handwritten digit segmentation: Is it still necessary? *Pattern Recognition*, 78:1–11, 2018.

[82] A. G. Hochuli, A. S. Britto, J. P. Barddal, R. Sabourin, and L. E. S. Oliveira. An end-to-end approach for recognition of modern and historical handwritten numeral strings. In *2020 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, 2020.

[83] A. G. Hochuli, A. S. Britto Jr, D. A. Saji, J. M. Saavedra, R. Sabourin, and L. S. Oliveira. A comprehensive comparison of end-to-end approaches for handwritten digit string recognition. *Expert Systems with Applications*, 165:114196, 2021.

[84] O. M. Parkhi, A. Vedaldi, and A. Zisserman. Deep face recognition. In *British Machine Vision Conference (BMVC)*, pages 1–12, Sept 2015.

[85] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman. VGGFace2: A dataset for recognising faces across pose and age. *CoRR*, 2017.

[86] C. N. Duong, K. G. Quach, I. Jalata, N. Le, and K. Luu. Mobiface: A lightweight deep learning face recognition on mobile devices. In *2019 IEEE 10th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pages 1–6, 2019.

[87] G. Guo and N. Zhang. A survey on deep learning based face recognition. *Computer Vision and Image Understanding*, 189:102805, 2019.

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

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

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

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

[92] A. S. Al-Waisy, R. Qahwaji, S. Ipson, S. Al-Fahdawi, and T. A. M. Nagem. A multi-biometric iris recognition system based on a deep learning approach. *Pattern Analysis and Applications*, Oct 2017.

[93] K. Nguyen, C. Fookes, A. Ross, and S. Sridharan. Iris recognition with off-the-shelf CNN features: A deep learning perspective. *IEEE Access*, 6:18848–18855, 2018.

[94] H. Proença and J. C. Neves. Segmentation-less and non-holistic deep-learning frameworks for iris recognition. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1–10, California, USA, June 2019. IEEE.

[95] M. De Marsico, A. Petrosino, and S. Ricciardi. Iris recognition through machine learning techniques: A survey. *Pattern Recognition Letters*, 82:106 – 115, 2016. An insight on eye biometrics.

[96] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems 27*, pages 2672–2680. Curran Associates, Inc., 2014.

[97] D. P. Kingma and M. Welling. Auto-encoding variational bayes. In *International Conference on Learning Representations*, 2014.

[98] A. B. L. Larsen, S. K. Sønderby, H. Larochelle, and O. Winther. Autoencoding beyond pixels using a learned similarity metric. In *ICML*, volume 48, pages 1558–1566, New York, New York, USA, Jun 2016. PMLR.

[99] G. Perarnau, Joost van de Weijer, Bogdan Raducanu, and Jose M. Álvarez. Invertible Conditional GANs for image editing. In *NIPS Workshop on Adversarial Training*, 2016.

[100] Y. Choi et al. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In *CVPR*, June 2018.

[101] G. Lample, Neil Zeghidour, Nicolas Usunier, Antoine Bordes, Ludovic DENOYER, and M. A. Ranzato. Fader networks:manipulating images by sliding attributes. In *Advances in Neural Information Processing Systems*, pages 5967–5976. Curran Associates, Inc., 2017.

[102] W. Shen and R. Liu. Learning residual images for face attribute manipulation. In *CVPR*, July 2017.

[103] J. Zhu, T. Park, P. Isola, and A. A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *ICCV*, Oct 2017.

[104] T. Karras, S. Laine, and T. Aila. A style-based generator architecture for generative adversarial networks. In *CVPR*, June 2019.

[105] T. Xiao, J. Hong, and J. Ma. Elegant: Exchanging latent encodings with gan for transferring multiple face attributes. In *ECCV*, September 2018.

[106] A. Rattani, N. Reddy, and R. Derakhshani. Convolutional neural network for age classification from smart-phone based ocular images. In *2017 IEEE International Joint Conference on Biometrics (IJCB)*, pages 756–761, 2017.

[107] M. de Assis Angeloni, R. de Freitas Pereira, and H. Pedrini. Age estimation from facial parts using compact multi-stream convolutional neural networks. In *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, pages 3039–3045, 2019.

[108] P. J. Phillips, K. W. Bowyer, P. J. Flynn, X. Liu, and W. T. Scruggs. The iris challenge evaluation 2005. In *IEEE International Conference on Biometrics: Theory, Applications and Systems*, pages 1–8, Sept 2008.

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

[110] R. Raghavendra and C. Busch. Learning deeply coupled autoencoders for smartphone based robust periocular verification. In *IEEE International Conference on Image Processing (ICIP)*, volume 1, pages 325–329, Phoenix, AZ, USA, sep 2016. IEEE.

[111] K. B. Raja, R. Raghavendra, and C. Busch. Collaborative representation of deep sparse filtered features for robust verification of smartphone periocular images. In *IEEE Intern. Conference on Image Processing*, volume 1, pages 330–334, Phoenix, AZ, USA, sep 2016. IEEE.

[112] K. Ahuja, R. Islam, F. A. Barbhuiya, and K. Dey. A preliminary study of CNNs for iris and periocular verification in the visible spectrum. In *International Conference on Pattern Recognition (ICPR)*, pages 181–186, Cancun, Mexico, dec 2016. IEEE.

[113] A. K. Jain, R. M. Bolle, and S. Pankanti. *Biometrics: Personal Identification in networked Society*. Springer US, 1 edition, 2006.

[114] U. Park, A. Ross, and A. K. Jain. Periocular biometrics in the visible spectrum: A feasibility study. In *IEEE International Conference on Biometrics: Theory, Applications, and Systems (BTAS)*, pages 1–6, Washington, DC, USA, Sep. 2009. IEEE.

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

[116] M. Uzair, A. Mahmood, A. Mian, and C. McDonald. Periocular region-based person identification in the visible, infrared and hyperspectral imagery. *Neurocomputing*, 149:854 – 867, 2015.

[117] J. Daugman. New Methods in Iris Recognition. *IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics)*, 37(5):1167–1175, oct 2007.

[118] D. M. Rankin, B. W. Scotney, P. J. Morrow, D. R. McDowell, and B. K. Pierscionek. Dynamic iris biometry: a technique for enhanced identification. *BMC Research Notes*, 3(1):182, Jul 2010.

[119] T. Tan, X. Zhang, Z. Sun, and H. Zhang. Noisy iris image matching by using multiple cues. *Pattern Recognition Letters*, 33(8):970–977, jun 2012.

[120] C. W. Tan and A. Kumar. Towards online iris and periocular recognition under relaxed imaging constraints. *IEEE Transactions on Image Processing*, 22(10):3751–3765, 2013.

[121] N. U. Ahmed, S. Cvetkovic, E. H. Siddiqi, A. Nikiforov, and I. Nikiforov. Using fusion of iris code and periocular biometric for matching visible spectrum iris images captured by smart phone cameras. In *International Conference on Pattern Recognition (ICPR)*, pages 176–180, Cancun, Mexico, dec 2016. IEEE.

[122] N. U. Ahmed, S. Cvetkovic, E. H. Siddiqi, A. Nikiforov, and I. Nikiforov. Combining iris and periocular biometric for matching visible spectrum eye images. *Pattern Recognition Letters*, 91:11–16, may 2017.

[123] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.

[124] T. S. Lee, D. Mumford, R. Romero, and V. A. Lamme. The role of the primary visual cortex in higher level vision. *Vision Research*, 38(15–16):2429–2454, 1998.

[125] D. J. Felleman and D.C. van Essen. Distributed hierarchical processing in the primate cerebral cortex. *Cereb Cortex*, 1(1):1–47, 1991.

[126] C. Cortes and V. N. Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, 1995.

[127] B. Schölkopf and A. J. Smola. *Learning with Kernels*. MIT Press, 2002.

[128] Y. LeCun, B. E. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. E. Hubbard, and L. D. Jackel. Handwritten digit recognition with a back-propagation network. In D. S. Touretzky, editor, *Advances in Neural Information Processing Systems 2*, pages 396–404. Morgan-Kaufmann, 1990.

[129] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, Nov 1998.

[130] D. Menotti, G. Chiachia, A. X. Falcão, and V. J. O. Neto. Vehicle license plate recognition with random convolutional networks. In *2014 27th SIBGRAPI Conference on Graphics, Patterns and Images*, pages 298–303, Aug 2014.

[131] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, and L. Fei-Fei. ImageNet: A large-scale hierarchical image database. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, Miami, FL, USA, June 2009. IEEE.

[132] V. Nair and G. E. Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th International Conference on International Conference on Machine Learning*, ICML'10, pages 807–814, USA, 2010. Omnipress.

[133] A. Waibel, T. Hanazawa, G. Hinton, K. Shikano, and K. J. Lang. Phoneme recognition using time-delay neural networks. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 37(3):328–339, March 1989.

[134] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In *International Conference on Learning Representations (ICLR)*, May 2015.

[135] F. Chollet. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[136] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger. Densely connected convolutional networks. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[137] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le. Learning transferable architectures for scalable image recognition. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[138] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *IEEE Conference on Computer Vision and Pattern Recognition*, June 2018.

[139] J. H. Alves and L. F. d. Oliveira. Optimizing neural architecture search using limited gpu time in a dynamic search space: A gene expression programming approach. In *2020 IEEE Congress on Evolutionary Computation (CEC)*, pages 1–8, 2020.

[140] V. Lopes, A. Gaspar, L. A. Alexandre, and J. Cordeiro. An automl-based approach to multimodal image sentiment analysis, 2021.

[141] V. Lopes, S. Alirezazadeh, and L. A. Alexandre. Epe-nas: Efficient performance estimation without training for neural architecture search, 2021.

[142] V. Lopes and L. A. Alexandre. Auto-classifier: A robust defect detector based on an automl head. In Haiqin Yang, Kitsuchart Pasupa, Andrew Chi-Sing Leung, James T. Kwok, Jonathan H. Chan, and Irwin King, editors, *Neural Information Processing*, pages 137–149, Cham, 2020. Springer International Publishing.

[143] J. C. Platt, N. Cristianini, and J. Shawe-Taylor. Large margin dags for multiclass classification. In *International Conference on Neural Information Processing Systems (NIPS)*, Nov 1999.

[144] T. Hastie, S. Rosset, J. Zhu, and H. Zou. Multi-class adaboost. *Statistics and Its Interface*, 2(3):349–360, 2009.

[145] G. Huang, H. Zhou, X. Ding, and R. Zhang. Extreme learning machine for regression and multiclass classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 42(2):513–529, 2012.

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

[147] S. S. Behera, S. S. Mishra, B. Mandal, and N. B. Puhan. Variance-guided attention-based twin deep network for cross-spectral periocular recognition. *Image and Vision Computing*, page 104016, 2020.

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

[149] F. Boutros, N. Damer, K. Raja, R. Ramachandra, F. Kirchbuchner, and A. Kuijper. Fusing iris and periocular region for user verification in head mounted displays. In *IEEE International Conference on Information Fusion (FUSION)*, pages 1–8, 2020.

[150] K. He, X. Zhang, S. Ren, and J. Sun. Identity mappings in deep residual networks. In *European Conf. on Computer Vision*, pages 630–645, 2016.

[151] Min Lin, Qiang Chen, and Shuicheng Yan. Network in network. 2014.

[152] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov. Improving neural networks by preventing co-adaptation of feature detectors. 2012.

[153] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–9, Boston, MA, USA, June 2015. IEEE.

[154] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017.

[155] R. Caruana. Multitask learning. *Machine Learning*, 28(1):41–75, July 1997.

[156] S. Ruder. An overview of multi-task learning in deep neural networks. *arXiv preprint arXiv:1706.05098*, 2017.

[157] J. Bromley, I. Guyon, Y. LeCun, E .Säckinger, and R. Shah. Signature verification using a "siamese" time delay neural network. In *Intl. Conf. on Neural Information Processing Systems*, page 737–744, 1993.

[158] S. Chopra, R. Hadsell, and Y. LeCun. Learning a similarity metric discriminatively, with application to face verification. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 539–546, 2005.

[159] R. Hadsell, S. Chopra, and Y. LeCun. Dimensionality reduction by learning an invariant mapping. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 1735–1742, 2006.

[160] R. P. Wildes, J. C. Asmuth, G. L. Green, S. C. Hsu, R. J. Kolczynski, J. R. Matey, and S. E. McBride. A machine-vision system for iris recognition. *Machine Vision and Applications*, 9:1–8, 1996.

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

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

[163] A. Rattani and R. Derakhshani. Ocular Biometrics in the Visible Spectrum: A Survey. *Image and Vision Computing*, 59:1–16, 2017.

[164] S. Shah and A. Ross. Generating Synthetic Irises by Feature Agglomeration. In *2006 International Conference on Image Processing*, pages 317–320. IEEE, 2006.

[165] J. Zuo, N. A. Schmid, and X. Chen. On generation and analysis of synthetic iris images. *IEEE Transactions on Information Forensics and Security*, 2(1):77–90, 2007.

[166] V. Ruiz-Albacete, P. Tome-Gonzalez, F. Alonso-Fernandez, J. Galbally, J. Fierrez, and J. Ortega-Garcia. Direct attacks using fake images in iris verification. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 5372 LNCS:181–190, 2008.

[167] A. Czajka. Database of Iris Printouts and its Application : Development of Liveness Detection Method for Iris Recognition. *MMAR, 18th International Conference*, pages 28–33, 2013.

[168] P. Gupta, S. Behera, M. Vatsa, and R. Singh. On Iris Spoofing Using Print Attack. In *2014 22nd International Conference on Pattern Recognition*, pages 1681–1686. IEEE, aug 2014.

[169] N. Kohli, D. Yadav, M. Vatsa, R. Singh, and A. Noore. Detecting medley of iris spoofing attacks using desist. In *2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pages 1–6, 2016.

[170] S. E. Baker, A. Hentz, K. W. Bowyer, and P. J. Flynn. Degradation of iris recognition performance due to non-cosmetic prescription contact lenses. *Computer Vision and Image Understanding*, 114(9):1030–1044, sep 2010.

[171] S. P. Fenker and K. W. Bowyer. Analysis of template aging in iris biometrics. In *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pages 45–51. IEEE, jun 2012.

[172] S. E Baker, K. W. Bowyer, P. J. Flynn, and P. J. Phillips. *Template Aging in Iris Biometrics*, pages 205–218. Springer London, London, 2013.

[173] S. S. Arora, M. Vatsa, R. Singh, and A. Jain. Iris recognition under alcohol influence: A preliminary study. In *2012 5th IAPR International Conference on Biometrics (ICB)*, pages 336–341. IEEE, mar 2012.

[174] J. E. Tapia, C. A. Perez, and K. W. Bowyer. Gender Classification From the Same Iris Code Used for Recognition. *IEEE Transactions on Information Forensics and Security*, 11(8):1760–1770, 2016.

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

[176] University of Notre Dame. Nd-crosssensor-iris-2013. https://cvrl.nd.edu/projects/data/#nd-crosssensor-iris-2013-data-set, 2013.

[177] D. Kim, Y. Jung, K. Toh, B. Son, and J. Kim. An empirical study on iris recognition in a mobile phone. *Expert Systems with Applications*, 54:328–339, jul 2016.

[178] R. Raghavendra, K. B. Raja, V. K. Vemuri, S. Kumari, P. Gacon, E. Krichen, and C. Busch. Influence of cataract surgery on iris recognition: A preliminary study. In *2016 International Conference on Biometrics (ICB)*, pages 1–8, 2016.

[179] M. Karakaya. A study of how gaze angle affects the performance of iris recognition. *Pattern Recognition Letters*, 82:132 – 143, 2016. An insight on eye biometrics.

[180] O. M. Kurtuncu, G. N. Cerme, and M. Karakaya. Comparison and evaluation of datasets for off-angle iris recognition. In Edward M. Carapezza, editor, *Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security, Defense, and Law Enforcement Applications XV*, volume 9825, pages 122 – 133. International Society for Optics and Photonics, SPIE, 2016.

[181] S. J. Garbin, Y. Shen, I. Schuetz, R. Cavin, G. Hughes, and S. S. Talathi. OpenEDS: Open Eye Dataset. *CoRR*, abs/1905.03702:1–11, 2019.

[182] IRISKING. Irisking. `http://www.irisking.com/`, 2017. Acessado em 2021-02-02.

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

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

[185] Int. Std. ISO/IEC 19794-6. Information technology-biometric data interchange formats-part 6: Iris image data. `https://www.iso.org/standard/50868.html`, 2011. Acessado em 2021-02-02.

[186] M. Karakaya, D. Barstow, H. Santos-Villalobos, and J. Thompson. Limbus impact on off-angle iris degradation. In *2013 International Conference on Biometrics (ICB)*, pages 1–6, 2013.

[187] M. Karakaya. Deep learning frameworks for off-angle iris recognition. In *2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pages 1–8, 2018.

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

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

[190] H. Proenca and L. A. Alexandre. UBIRIS: A noisy iris image database. In *13th International Conference on Image Analysis and Processing - ICIAP 2005*, volume 3617, pages 970–977. Springer Berlin Heidelberg, 2005.

[191] M. S. Hosseini, B. N. Araabi, and H. Soltanian-Zadeh. Pigment melanin: Pattern for iris recognition. *IEEE Transactions on Instrumentation and Measurement*, 59(4):792–804, 2010.

[192] A. Sharma, S. Verma, M. Vatsa, and R. Singh. On cross spectral periocular recognition. In *2014 IEEE International Conference on Image Processing (ICIP)*, pages 5007–5011, Paris, France, Oct 2014. IEEE.

[193] F. M. Algashaam, K. Nguyen, M. Alkanhal, V. Chandran, W. Boles, and J. Banks. Multispectral periocular classification with multimodal compact multi-linear pooling. *IEEE Access*, 5:14572–14578, 2017.

[194] K. B. Raja, R. Raghavendra, V. K. Vemuri, and C. Busch. Smartphone based visible iris recognition using deep sparse filtering. *Pattern Recognition Letters*, 57:33–42, may 2015.

[195] M. Dobeš, L. Machala, P. Tichavský, and J. Pospíšil. Human eye iris recognition using the mutual information. *Optik - International Journal for Light and Electron Optics*, 115(9):399–404, jan 2004.

[196] S. Siena, V. N. Boddeti, and B. V. K. Vijaya Kumar. Coupled marginal fisher analysis for low-resolution face recognition. In *European Conference on Computer Vision (ECCV)*, pages 240–249, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

[197] A. F. Sequeira, J. Murari, and J. S. Cardoso. Iris Liveness Detection Methods in Mobile Applications. In *2014 International Conference on Computer Vision Theory and Applications (VISAPP)*, volume 3, pages 22–33, Jan 2014.

[198] G. Santos, E. Grancho, M. V. Bernardo, and P. T. Fiadeiro. Fusing iris and periocular information for cross-sensor recognition. *Pattern Recognition Letters*, 57:52–59, may 2015.

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

[200] R. Donida Labati, A. Genovese, V. Piuri, F. Scotti, and S. Vishwakarma. I-social-db: A labeled database of images collected from websites and social media for iris recognition. *Image and Vision Computing*, page 104058, 2020.

[201] J. M. Smereka, V. N. Boddeti, and B. V. K. Vijaya Kumar. Probabilistic deformation models for challenging periocular image verification. *IEEE Transactions on Information Forensics and Security*, 10(9):1875–1890, sep 2015.

[202] J. Fierrez et al. BiosecurID: a multimodal biometric database. *Pattern Analysis and Applications*, 13(2):235–246, may 2010.

[203] J. Ortega-Garcia, J. Fierrez, F. Alonso-Fernandez, J. Galbally, M. R. Freire, J. Gonzalez-Rodriguez, C. Garcia-Mateo, J. Alba-Castro, E. Gonzalez-Agulla, E. Otero-Muras, S. Garcia-Salicetti, L. Allano, B. Ly-Van, B. Dorizzi, J. Kittler, T. Bourlai, N. Poh, F. Deravi, M. Ng, M. Fairhurst, J. Hennebert, A. Humm, M. Tistarelli, L. Brodo, J. Richiardi, A. Drygajlo, H. Ganster, F. M. Sukno, S. Pavani, A. Frangi, L. Akarun, and A. Savran. The Multiscenario Multienvironment BioSecure Multimodal Database (BMDB). *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(6):1097–1111, jun 2010.

[204] NIST. Multiple Biometric Grand Challenge (MBGC). https://www.nist.gov/programs-projects/multiple-biometric-grand-challenge-mbgc, 2010.

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

[206] NIST. Face and Ocular Challenge Series (FOCS). https://www.nist.gov/programs-projects/face-and-ocular-challenge-series-focs, 2010.

[207] B. Ríos-Sánchez, M. F. Arriaga-Gómez, J. Guerra-Casanova, D. de Santos-Sierra, I. de Mendizábal-Vázquez, G. Bailador, and C. Sánchez-Ávila. gb2sμMOD: A MUlti-MODal biometric video database using visible and IR light. *Information Fusion*, 32:64–79, nov 2016.

[208] K. Hollingsworth, T. Peters, K. W. Bowyer, and P. J. Flynn. Iris recognition using signal-level fusion of frames from video. *IEEE Transactions on Information Forensics and Security*, 4(4):837–848, dec 2009.

[209] J.R. Matey, O. Naroditsky, K. Hanna, R. Kolczynski, D.J. LoIacono, S. Mangru, M. Tinker, T.M. Zappia, and W.Y. Zhao. Iris on the move: Acquisition of images for iris recognition in less constrained environments. *Proceedings of the IEEE*, 94(11):1936–1947, nov 2006.

[210] P.J. Phillips, P.J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the face recognition grand challenge. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 947–954, San Diego, CA, USA, 2005. IEEE.

[211] D. L. Woodard, S. J. Pundlik, J. R. Lyle, and P. E. Miller. Periocular region appearance cues for biometric identification. In *IEEE Conference on Computer Vision and Pattern Recognition - Workshops (CVPRW)*, pages 162–169, San Francisco, CA, USA, jun 2010. IEEE.

[212] Q. Wang, X. Zhang, M. Li, X. Dong, Q. Zhou, and Y. Yin. Adaboost and multi-orientation 2D Gabor-based noisy iris recognition. *Pattern Recognition Letters*, 33(8):978–983, 2012.

[213] K. B. Raja, R. Raghavendra, S. Venkatesh, and C. Busch. Multi-patch deep sparse histograms for iris recognition in visible spectrum using collaborative subspace for robust verification. *Pattern Recognition Letters*, 91:27–36, may 2017.

[214] M. Zhang, Q. Zhang, Z. Sun, S. Zhou, and N. U. Ahmed. The BTAS*Competition on Mobile Iris Recognition. In *IEEE International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pages 1–7, Nova York (USA), sep 2016. IEEE.

[215] W. Sankowski, K. Grabowski, M. Napieralska, M. Zubert, and A. Napieralski. Reliable algorithm for iris segmentation in eye image. *Image and Vision Computing*, 28(2):231–237, feb 2010.

[216] P. De Almeida. A knowledge-based approach to the iris segmentation problem. *Image and Vision Computing*, 28(2):238–245, feb 2010.

[217] P. Li, X. Liu, L. Xiao, and Q. Song. Robust and accurate iris segmentation in very noisy iris images. *Image and Vision Computing*, 28(2):246–253, feb 2010.

[218] D. S. Jeong, J. W. Hwang, B. J. Kang, K. R. Park, C. S. Won, D. Park, and J. Kim. A new iris segmentation method for non-ideal iris images. *Image and Vision Computing*, 28(2):254–260, feb 2010.

[219] Y. Chen, M. Adjouadi, C. Han, J. Wang, A. Barreto, N. Rishe, and J. Andrian. A highly accurate and computationally efficient approach for unconstrained iris segmentation. *Image and Vision Computing*, 28(2):261–269, feb 2010.

[220] R. Donida Labati and F. Scotti. Noisy iris segmentation with boundary regularization and reflections removal. *Image and Vision Computing*, 28(2):270–277, feb 2010.

[221] M. A. Luengo-Oroz, E. Faure, and J. Angulo. Robust iris segmentation on uncalibrated noisy images using mathematical morphology. *Image and Vision Computing*, 28(2):278–284, feb 2010.

[222] G. Santos and E. Hoyle. A fusion approach to unconstrained iris recognition. *Pattern Recognition Letters*, 33(8):984–990, jun 2012.

[223] K. Y. Shin, G. P. Nam, D. S. Jeong, D. H. Cho, B. J. Kang, K. R. Park, and J. Kim. New iris recognition method for noisy iris images. *Pattern Recognition Letters*, 33(8):991–999, jun 2012.

[224] P. Li, X. Liu, and N. Zhao. Weighted co-occurrence phase histogram for iris recognition. *Pattern Recognition Letters*, 33(8):1000–1005, jun 2012.

[225] M. De Marsico, M. Nappi, and D. Riccio. Noisy iris recognition integrated scheme. *Pattern Recognition Letters*, 33(8):1006–1011, jun 2012.

[226] P. Li and H. Ma. Iris recognition in non-ideal imaging conditions. *Pattern Recognition Letters*, 33(8):1012–1018, jun 2012.

[227] R. Szewczyk, K. Grabowski, M. Napieralska, W. Sankowski, M. Zubert, and A. Napieralski. A reliable iris recognition algorithm based on reverse biorthogonal wavelet transform. *Pattern Recognition Letters*, 33(8):1019–1026, jun 2012.

[228] M. Haindl and M. Krupicka. Unsupervised detection of non-iris occlusions. *Pattern Recognition Letters*, 57:60–65, 2015.

[229] K. Ahuja, R. Islam, F. A. Barbhuiya, and K. Dey. Convolutional neural networks for ocular smartphone-based biometrics. *Pattern Recognition Letters*, 91(2):17–26, may 2017.

[230] A. Abate, S. Barra, L. Gallo, and F. Narducci. Skipsom: Skewness & kurtosis of iris pixels in self organizing maps for iris recognition on mobile devices. In *23rd ICPR*, pages 155–159, Cancun, Mexico, Dec 2016. IEEE.

[231] A. F. Abate, S. Barra, L. Gallo, and F. Narducci. Kurtosis and skewness at pixel level as input for SOM networks to iris recognition on mobile devices. *Pattern Recognition Letters*, 91:37–43, may 2017.

[232] C. Galdi and J. Dugelay. Fusing iris colour and texture information for fast iris recognition on mobile devices. In *International Conference on Pattern Recognition (ICPR)*, pages 160–164, Cancun, Mexico, dec 2016. IEEE.

[233] C. Galdi and J. Dugelay. FIRE: Fast Iris REcognition on mobile phones by combining colour and texture features. *Pattern Recognition Letters*, 91:44–51, may 2017.

[234] N. Aginako, J. M. Martinez-Otzerta, B. Sierra, M. Castrillon-Santana, and J. Lorenzo-Navarro. Local descriptors fusion for mobile iris verification. In *ICPR*, pages 165–169, Cancun, Mexico, Dec 2016. IEEE.

[235] N. Aginako, M. Castrillón-Santana, J. Lorenzo-Navarro, J. M. Martínez-Otzeta, and B. Sierra. Periocular and iris local descriptors for identity verification in mobile applications. *Pattern Recognition Letters*, 91:52–59, may 2017.

[236] N. Aginako, J. M. Martinez-Otzeta, I. Rodriguez, E. Lazkano, and B. Sierra. Machine learning approach to dissimilarity computation: Iris matching. In *ICPR*, pages 170–175, Cancun, Mexico, dec 2016. IEEE.

[237] N. Aginako, G. Echegaray, J. M. Martínez-Otzeta, I. Rodríguez, E. Lazkano, and B. Sierra. Iris matching by means of Machine Learning paradigms: A new approach to dissimilarity computation. *Pattern Recognition Letters*, 91:60–64, may 2017.

[238] P. Viola and M. J. Jones. Robust Real-Time Face Detection. *International Journal of Computer Vision*, 57(2):137–154, 2004.

[239] A. Das, U. Pal, M. Blumenstein, C. Wang, Y. He, Y. Zhu, and Z. Sun. Sclera segmentation benchmarking competition in cross-resolution environment. In *2019 International Conference on Biometrics (ICB)*, pages 1–7, 2019.

[240] R. A. Naqvi and W. Loh. Sclera-net: Accurate sclera segmentation in various sensor images based on residual encoder and decoder network. *IEEE Access*, 7:98208–98227, 2019.

[241] C. Wang, Y. He, Y. Liu, Z. He, R. He, and Z. Sun. Scleraseg net: an improved u-net model with attention for accurate sclera segmentation. In *2019 International Conference on Biometrics (ICB)*, pages 1–8, 2019.

[242] M. Vitek, P. Rot, V. Štruc, and P. Peer. A comprehensive investigation into sclera biometrics: a novel dataset and performance study. *Neural Computing and Applications*, 32:17941–17955, 2020.

[243] P. Rot, M. Vitek, K. Grm, Z. Emeršič, P. Peer, and V. Štruc. *Deep Sclera Segmentation and Recognition*, pages 395–432. Springer International Publishing, Cham, 2020.

[244] M. S. Maheshan, B. S. Harish, and N. Nagadarshan. A convolution neural network engine for sclera recognition. *International Journal of Interactive Multimedia and Artificial Intelligence*, 6(1):78–83, 2020.

[245] M. Vitek et al. Ssbc 2020: Sclera segmentation benchmarking competition in the mobile environment. In *2020 International Joint Conference on Biometrics (IJCB)*, pages 1–10, 2020.

[246] A. Das, U. Pal, M. A. Ferrer, and M. Blumenstein. Ssrbc 2016: Sclera segmentation and recognition benchmarking competition. In *2016 International Conference on Biometrics (ICB)*, pages 1–6, 2016.

[247] J. L. G. Rodríguez and Y. D. Rubio. A new method for iris pupil contour delimitation and its application in iris texture parameter estimation. In Alberto Sanfeliu and Manuel Lazo Cortés, editors, *Progress in Pattern Recognition, Image Analysis and Applications*, pages 631–641, Berlin, Heidelberg, 2005. Springer Berlin Heidelberg.

[248] Y. Alvarez-Betancourt and M. Garcia-Silvente. A fast iris location based on aggregating gradient approximation using qma-owa operator. In *International Conference on Fuzzy Systems*, pages 1–8, July 2010.

[249] Z. Yu and W. Cui. A rapid iris location algorithm based on embedded. In *2012 International Conference on Computer Science and Information Processing (CSIP)*, pages 233–236, Aug 2012.

[250] L. Zhou, Y. Ma, J. Lian, and Z. Wang. A new effective algorithm for iris location. In *2013 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, pages 1790–1795, Dec 2013.

[251] W. Zhang and Y. D. Ma. A new approach for iris localization based on an improved level set method. In *2014 11th International Computer Conference on Wavelet Actiev Media Technology and Information Processing(ICCWAMTIP)*, pages 309–312, Dec 2014.

[252] L. Su, J. Wu, Q. Li, and Z. Liu. Iris location based on regional property and iterative searching. In *2017 IEEE International Conference on Mechatronics and Automation (ICMA)*, pages 1064–1068, Aug 2017.

[253] F. Jan. Segmentation and localization schemes for non-ideal iris biometric systems. *Signal Processing*, 133(November 2016):192–212, apr 2017.

[254] A. Gangwar, A. Joshi, A. Singh, F. Alonso-Fernandez, and J. Bigun. Irisseg: A fast and robust iris segmentation framework for non-ideal iris images. In *2016 International Conference on Biometrics (ICB)*, pages 1–8, June 2016.

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

[256] S. Van Dongen and A. J. Enright. Metric distances derived from cosine similarity and Pearson and Spearman correlations. *CoRR*, 2012.

[257] J. Daugman. The importance of being random: statistical principles of iris recognition. *Pattern Recognition*, 36(2):279–291, 2003.

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

[259] F. Alonso-Fernandez K. Hernandez-Diaz and J. Bigun. Cross spectral periocular matching using resnet features. In *International Conference on Biometrics(ICB)*, pages 1–6, 2019. In Press.

[260] T. Ojala, M. Pietikainen, and D. Harwood. Performance evaluation of texture measures with classification based on kullback discrimination of distributions. In *ICPR*, volume 1, pages 582–585. IEEE, 1994.

[261] T. Ojala, M. Pietikäinen, and D. Harwood. A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition*, 29(1):51–59, 1996.

[262] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *CVPR*, volume 1, pages 886–893, June 2005.

[263] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60:91–110, 2004.

[264] V. Ojansivu and J. Heikkilä. Blur insensitive texture classification using local phase quantization. In *International conference on image and signal processing*, pages 236–243. Springer, 2008.

[265] N. Reddy, A. Rattani, and R. Derakhshani. Comparison of deep learning models for biometric-based mobile user authentication. In *IEEE International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pages 1–6, 2018.

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