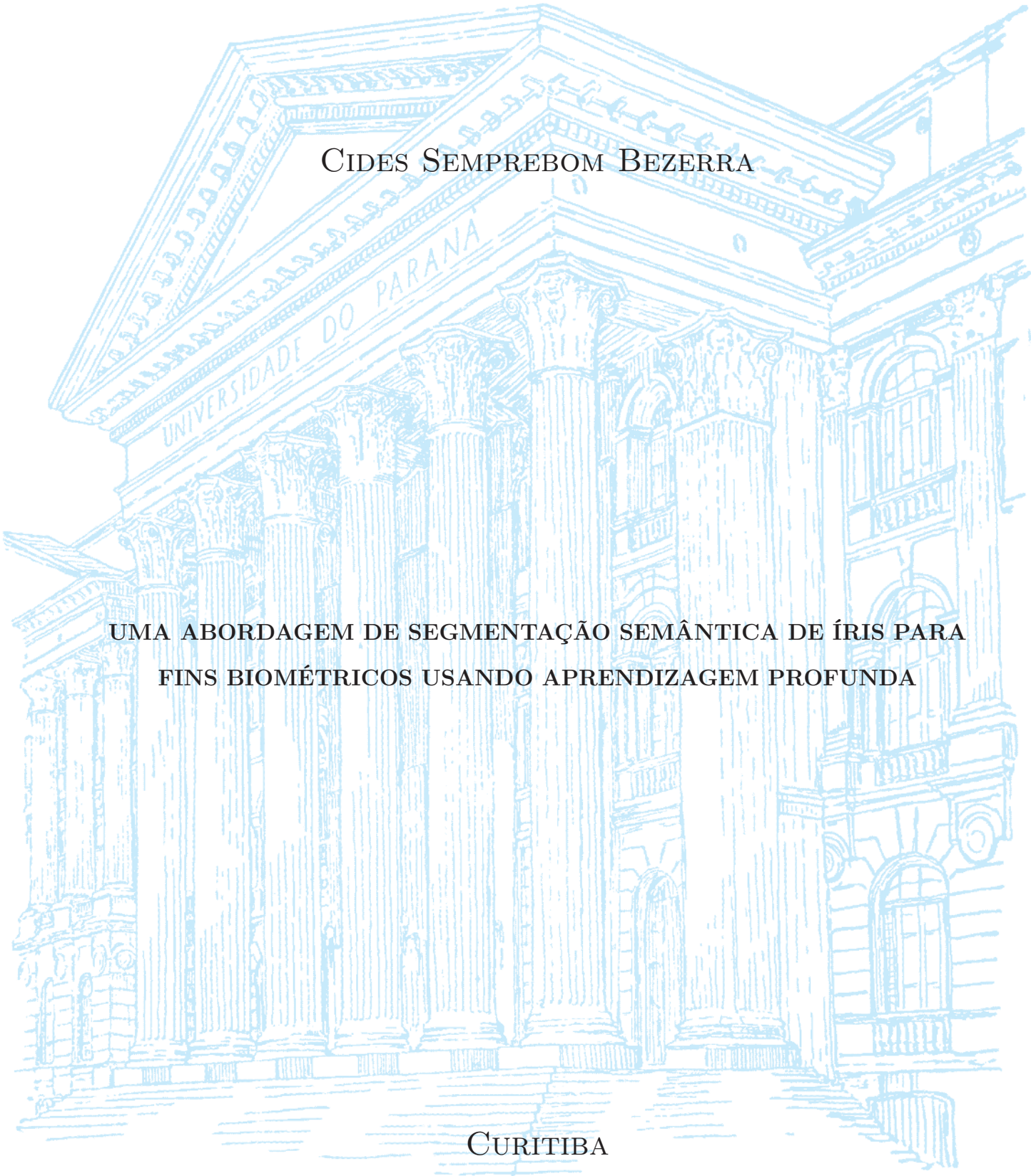
Universidade Federal do Paraná

Cides Semprebom Bezerra



uma abordagem de segmentação semântica de íris para fins biométricos usando aprendizagem profunda

Curitiba
2018

Cides Semprebom Bezerra

UMA ABORDAGEM DE SEGMENTAÇÃO SEMÂNTICA DE ÍRIS PARA FINS BIOMÉTRICOS USANDO APRENDIZAGEM PROFUNDA

Dissertação apresentada como requisito parcial à obtenção do grau de Mestre em Informática no Programa de Pós-Graduação em Informática, Setor de Ciências Exatas, da Universidade Federal do Paraná.

Área de concentração: *Ciência da Computação.*

Orientador: David Menotti Gomes.

Curitiba
2018

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

B574a Bezerra, Cides Semprebom
Uma abordagem de segmentação semântica de íris para fins biométricos usando aprendizagem profunda / Cides Semprebom Bezerra. – Curitiba, 2018.

Dissertação - Universidade Federal do Paraná, Setor de Ciências Exatas, Programa de Pós-Graduação em Informática, 2018.

Orientador: David Menotti Gomes.

1. Transferência de aprendizagem. 2. Aprendizagem. 3. Identificação biométrica. I. Universidade Federal do Paraná. II. Gomes, David Menotti. III. Título.

CDD: 570.15195

Bibliotecária: Vanusa Maciel CRB- 9/1928

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 Dissertação de Mestrado de **CIDES SEMPREBOM BEZERRA** intitulada: **Uma abordagem de segmentação semântica de íris para fins biométricos usando aprendizagem profunda**, após terem inquirido o aluno e realizado a avaliação do trabalho, são de parecer pela sua APROVADO no rito de defesa.

A outorga do título de mestre 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, 26 de Setembro de 2018.

DAVID MENOTTI GOMES
Presidente da Banca Examinadora (UFPR)

LUCAS FERRARI DE OLIVEIRA
Avaliador Interno (UFPR)

ALCEU DE SOUZA BRITTO JR
Avaliador Externo (PUCPR)

EDUARDO TODT
Avaliador Interno (UFPR)

MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DO PARANÁ
Setor de Ciências Exatas
Coordenação do Programa de Pós-Graduação em Informática

Agradecimentos

Agradeço à minha família, em especial meus pais Gildenilton Bezerra e Maria Cristina Semprebom Bezerra pela educação, pelos ensinamentos para a vida e principalmente pela força e apoio durante a construção deste trabalho.

Agradeço também, ao meu orientador David Menotti Gomes pela grande confiança e paciência ao longo desta trajetória. A todos os amigos do laboratório de Visão, Robótica e Imagem (VRI), em especial Rayson Laroca, Valber Lemes Zacarkim, Caroline Quadros Cordeiro, Felipe Bombardeli, Jeovane Honório Alves, André Gustavo Hochuli e Luiz Zanlorensi pelos momentos de descontração, incentivo e apoio agradeço imensamente.

Sou grato também aos amigos dos outros laboratórios de pesquisa, pela troca de experiências que de certa forma contribuíram direta e indiretamente para o desenvolvimento desta dissertação.

A todos os demais que aqui não foram citados, mas que contribuíram para que eu alcançasse este objetivo, tenham a certeza que sou imensamente grato.

Muito obrigado a todos vocês.

RESUMO

As características presentes na íris tornam-na uma das mais representativas modalidades biométricas com alto grau de unicidade. Entretanto, diversos fatores presentes na íris podem interferir no desempenho do sistema biométrico, sendo necessário removê-los de forma mais precisa possível, deixando apenas a região da íris. *Convolutional Neural Network*s (CNNs) empregadas em diversos problemas de segmentação, tais como segmentação de próstata, de esclera e de estradas por exemplo, alcançaram resultados estado-da-arte. Portanto, inspirado nesses bons resultados, o presente trabalho apresenta 5 abordagens que envolvem aprendizagem profunda para segmentar imagens da íris de 7 diferentes *datasets* de ambos os domínios *Near Infra-red* (NIR) e *Visible* (VIS) e cenários controlados e não controlados. Duas dessas abordagens (VggFCN-$fc7$ e VggFCN-*pool*5) são inspiradas no modelo *Visual Geometry Group* (VGG)16 e as outras três (ResNet-50, ResNet-101 e ResNet-152) baseiam-se na *Residual Network* (ResNet), conforme uma arquitetura codificador-decodificador. Todas as 5 abordagens são combinadas com a arquitetura *Fully Convolutional Network* (FCN) no decodificador, e utilizam a técnica de transferência de aprendizagem, com pesos pré-treinados a partir do *dataset ImageNet Large-Scale Visual Recognition Challenge* (ILSVRC). As abordagens propostas apresentaram os melhores resultados em comparação com 4 *baselines* utilizados. VggFCN-$fc7$ e VggFCN-*pool*5 obtiveram respectivamente 01,05% e 00,96% de erro E de segmentação no *dataset* NICE.I. Nas ResNet-50, ResNet-101 e ResNet-152 o E foi respectivamente de 01,13%, 00,98% e 00,97% também no *dataset* NICE.I. Além disso, após uma análise estatística, por meio dos testes de Friedman e Nemenyi, constatou-se que a abordagem ResNet-101 foi melhor nos *datasets* NIR, enquanto a VggFCN-*pool*5 foi melhor nos VIS. Ao mesclar os *datasets* NIR e VIS três abordagens (VggFCN-$fc7$, ResNet-50 e VggFCN-*pool*5) ficaram empatadas.

Palavras-chave: Aprendizagem Profunda. Segmentação de íris. Modalidade biométrica. Transferência de aprendizagem.

ABSTRACT

The characteristics present in the iris make it one of the most representative biometric modalities with a high degree of uniqueness. However, several factors present in the iris can interfere with the performance of the biometric system, and it is necessary to remove them as accurately as possible, leaving only the region of the iris. *Convolutional Neural Network*s (CNNs) employed in various segmentation problems, such as prostate, sclera and road segmentation, for example, have achieved state-of-the-art results. So, inspired by these good results, the present paper presents 5 approaches that involve deep learning to segment different 7 iris images *datasets* of both *Near Infra-red* (NIR) and *Visible* (VIS) domains and controlled and uncontrolled scenarios. Two of these approaches (VggFCN-$fc7$ and VggFCN-*pool*5) are inspired by the *Visual Geometry Group* (VGG)16 model and the other three (ResNet-50, ResNet-101 and ResNet-152) are based on *Residual Network* (ResNet), according to an encoder-decoder architecture. All 5 approaches are combined with the *Fully Convolutional Network* (FCN) architecture in the decoder, and use the transfer learning technique, with pre-trained weights from *dataset ImageNet Large-Scale Visual Recognition Challenge* (ILSVRC). The proposed approaches presented the best results compared to 4 *baselines* used. VggFCN-$fc7$ and VggFCN-*pool*5 obtained respectively 01.05% and 00.96% of segmentation error E in NICE.I *dataset*. In ResNet-50, ResNet-101 and ResNet-152 the E were respectively 01.13%, 00.98% and 00.97% also in NICE.I. In addition, after a statistical analysis, through the tests of Friedman and Nemenyi, it was found that the ResNet-101 approach was better in *datasets* NIR, while VggFCN-*pool*5 was better in VIS . By merging *datasets* NIR and VIS three approaches (VggFCN-$fc7$, ResNet-50, and VggFCN-*pool*5) are tied.

Keywords: Deep Learning. Iris segmentation. Biometric modality. Transfer learning.

Lista de Figuras

Lista de Tabelas

Lista de acrônimos

BioSec	*Biometrics and Security*
CASIA	*Chinese Academy of Sciences e Institute of Automation*
CasiaI3	*Casia-Iris-v3-Interval*
CasiaT4	*Casia-Iris-v4-Thousand*
CasiaV1	*Casia-Iris-v1*
CasiaV2	*Casia-Iris-v2*
CasiaL3	*CASIA-Iris-LampV3*
CasiaTW3	*CASIA-Iris-TwinsV3*
CasiaD4	*CASIA-Iris-DistanceV4*
CasiaS4	*CASIA-Iris-SynV4*
CD	*critical difference*
CNN	*Convolutional Neural Network*
CrEye-Iris	*Cross-Spectral Iris/Periocular*
CSIP	*Cross-Sensor Iris and Periocular*
CuIris	*Dark-Skinned African Iris Dataset*
DCNN	*Deep Convolutional Neural Network*
F1	*F-Measure*
FCN	*Fully Convolutional Network*
FN	*False Negative*
FP	*False Positive*
FPR	*False Positive Rate*
GAC	Contornos Ativos Geodésicos
GS4	*Galaxy Samsung IV*
GT2	*Galaxy Tablet II*
GT	*Ground-Truth*
HOG	*Histogram of Oriented Gradients*
HCNN	*Hierarchical Convolutional Neural Network*
HT	*Hough Transform*
IoU	*Intersection over Union*
ICE	*Iris Challenge Evaluation*
ILSVRC	*ImageNet Large-Scale Visual Recognition Challenge*
IITD-1	*IITD Iris Image Database 1.0*
IP5	*iPhone 5*

IrisDenseNet	*Densely Connected Fully Convolutional Network*
IRISSEG	*Iris Segmentation Framework*
KNN	*K-Nearest Neighbor*
LBP	*Local Binary Pattern*
LBD	*Learned Boundary Detectors*
Mean Acc.	*Mean Accuracy*
MICHE-I	*Mobile Iris CHallenge Evaluation I*
MICHE-II	*Mobile Iris CHallenge Evaluation II*
MFCN	*Multi-scale Fully Convolutional Network*
MRS	*Maximum Radial Suppression*
NIR	*Near Infra-red*
NICE.I	*Noisy Iris Challenge Evaluation - Part I*
NICE.II	*Noisy Iris Challenge Evaluation - Part II*
NTDame	*Notredame 0405 Iris*
OSIRISv4.1	*Open Source Iris Recognition System*
OCAC	*Optical Correlation based Active Contours*
Prec	*Precision*
PR	*Precision-Recall*
Rec	*Recall*
ResNet	*Residual Network*
RGB	*Red, Green and Blue*
RNA	Redes Neurais Artificiais
ROC	*Receiver Operating Characteristics*
SegNet	*Segmentation Network*
SLIC	*Simple Linear Iterative Clustering*
SVM	*Support Vector Machines*
SPDNN	*Semi Parallel Deep Neural Network*
TN	*True Negative*
TP	*True Positive*
TPR	*True Positive Rate*
TVM	*Total Variation Model*
UBIRIS.v1	*University of Beira Interior Iris - I*
UBIRIS.v2	*University of Beira Interior Iris - II*
UTIRIS	*University of Tehran IRIS*
VIS	*Visible*
VOC	*Pascal Visual Object Classes*
VGG	*Visual Geometry Group*

SUMÁRIO

1 Introdução

A identificação de pessoas por meio de suas características físicas e comportamentais surgiu na área criminal, com a finalidade de identificar criminosos. As características físicas proporcionam grau de confiabilidade superior em comparação com outros meios de identificação, como senhas por exemplo (Crihalmeanu et al., 2007). Diversas destas características podem ser utilizadas para este fim. Dentre essas características, estão a face, voz, anatomia das mãos, impressão digital, forma de andar, formato das orelhas, íris, esclera, retina, arcada dentária, nariz, dentre outras. Entretanto, as impressões digitais, face e íris são as modalidades mais comuns a serem utilizadas nos sistemas biométricos (Jain et al., 1997, 2004, 2016).

Dentre estas comumente utilizadas, as características presentes na íris a tornam uma das mais representativas e segura modalidade biométrica devido ao seu alto grau de unicidade (Turk e Pentland, 1991; Daugman, 1993; Wildes, 1997). Atualmente esse tipo de biometria é utilizada em diversas áreas como controle de fronteiras internacionais, transações financeiras e segurança computacional. Além disso, a aquisição das imagens pode ser feita de forma não invasiva e discreta (Daugman, 1993). Daugman (1993) relata que diversos oftalmologistas observaram alto nível de detalhes de informações e textura presentes na íris. Essas informações não sofrem alterações no decorrer do tempo de vida das pessoas, sem levar em consideração fatores externos como (doenças e/ou acidentes), reforçando ainda mais a importância do seu uso.

No olho humano, a íris é a região texturizada, geralmente colorida, onde seus limites interno e externo, correspondem respectivamente à pupila e à esclera, conforme ilustrado na Figura 1.1. Observe que as partes superiores e inferiores da íris sofrem oclusão pela presença das pálpebras e dos cílios respectivamente, considerados certos tipos de ruídos. Além desses ruídos, conforme será apresentado na Seção 1.2, existem reflexos, causados pela superfície refletiva do globo ocular, cabelos, óculos também são tipos de ruídos, que atrapalham a identificação/reconhecimento dos indivíduos.

1.1 Sistema Biométrico de Íris

Um sistema biométrico automatizado para reconhecimento de íris, conforme descrito em Jillela e Ross (2015), é composto de alguns passos fundamentais para seu correto funcionamento. Esse sistema geralmente é dividido em cinco passos: *Aquisição das imagens*, *segmentação da íris*, *normalização*, *extração de características* e o *matching*.

- *Aquisição das imagens:* corresponde à entrada da informação no sistema, que é feita por meio de um sensor para a captura da imagem. Normalmente as imagens são capturadas no espectro *Near Infra-red* (NIR), que representa melhor as características de textura, principalmente das íris muito escuras (Zhou et al., 2010; Chen et al., 2016; Jain et al., 2016).

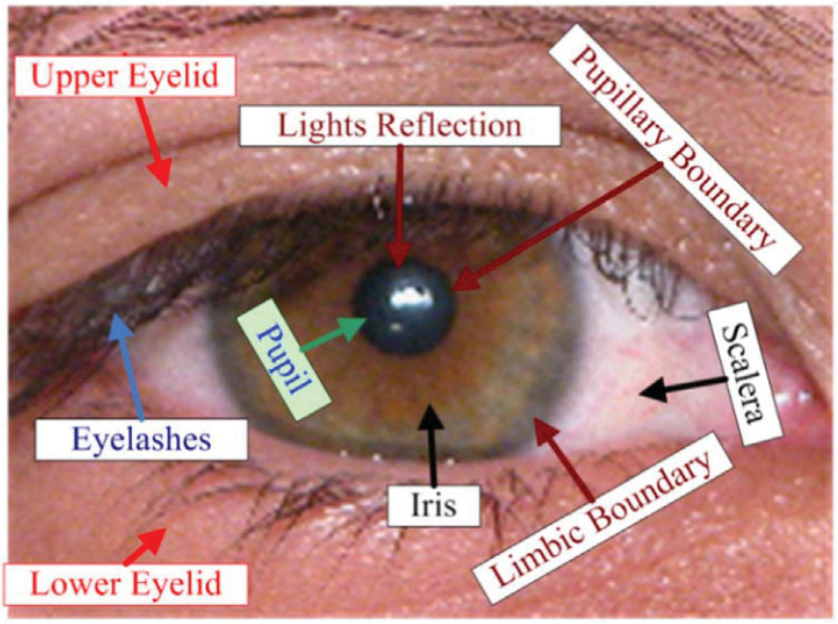


Figura 1.1: Componentes da região periocular em uma imagem colorida. Fonte: (Jan, 2017)

- *Segmentação da íris:* é o processo de localizar e isolar a íris do restante da imagem. É o passo mais desafiador e crucial em um sistema biométrico. Caso a segmentação falhar ou ocorrerem erros, os passos posteriores são prejudicados e o desempenho do sistema biométrico é afetado (Jillela e Ross, 2016; Pundlik et al., 2008; Jillela e Ross, 2015).

- *Normalização:* corresponde na transformação da região circular da íris, que possui coordenadas polares, para uma imagem retangular (coordenadas cartesianas). Essa transformação serve para deixar todas as imagens com o mesmo tamanho (Daugman, 1993; Jillela e Ross, 2015).

- *Extração de características:* é o processo que identifica padrões mais discriminantes e representativos da íris. Esses padrões são as características, similares para o mesmo indivíduo, e totalmente distintas para indivíduos diferentes (Daugman, 1993, 2001, 2003; Jillela e Ross, 2015).

- *Matching:* é a comparação das características de uma imagem, com todas as outras imagens da base. Essa comparação gera valores, dos quais é possível identificar o indivíduo, caso suas características estejam cadastradas na base (Daugman, 2007; Jillela e Ross, 2015; Bashir et al., 2008).

A Figura 1.2 ilustra a estrutura de um sistema biométrico comum para reconhecimento de íris, de acordo com os cinco passos citados anteriormente. Observe que esse sistema apresenta a aquisição da imagem de uma maneira um tanto quanto controlada, isso influencia no sucesso do reconhecimento, pelo fato de que o cenário de aquisição da imagem seja restrito.

Com essa restrição, as chances de ocorrerem ruídos na imagem são reduzidas em relação à uma tomada de imagem em ambiente livre, “mascarando” e diferenciando o sistema biométrico de uma situação totalmente real. Em contrapartida, a presença dos diversos ruídos presentes em um cenário mais próximo do real, como um sistema de

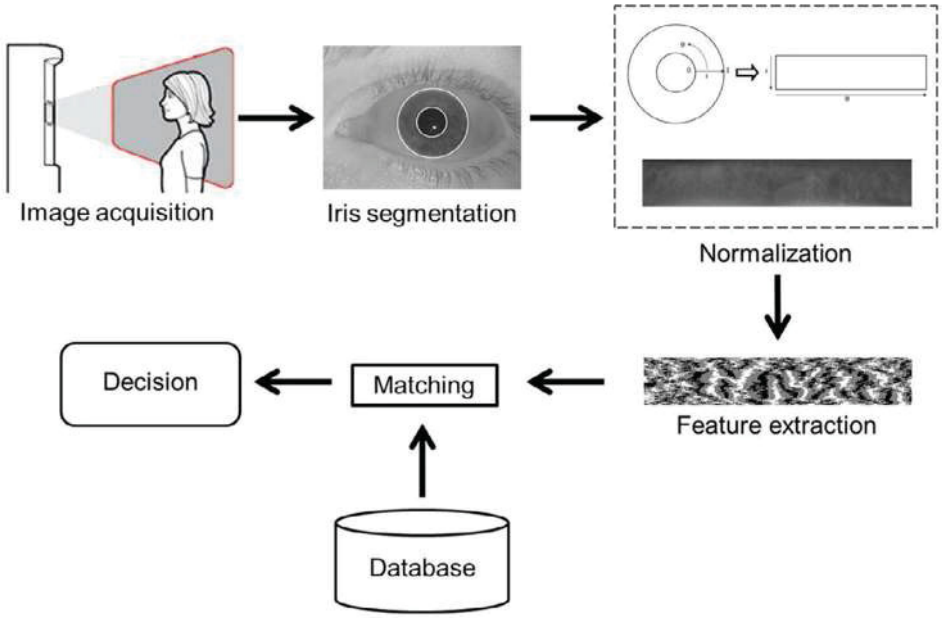
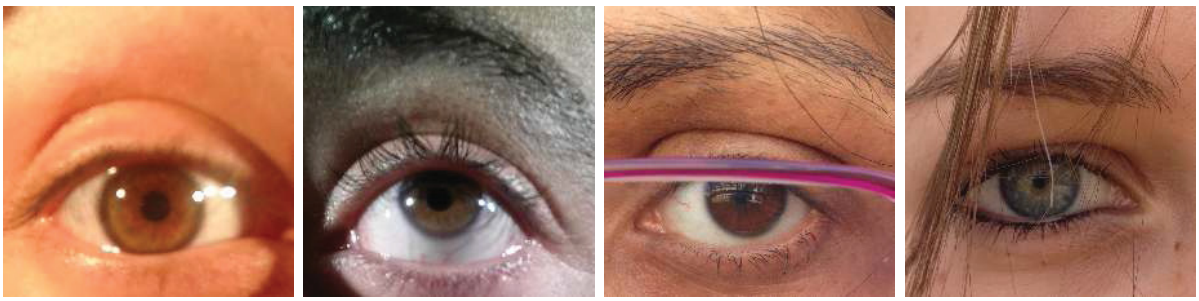


Figura 1.2: Diagrama de blocos de um típico sistema de reconhecimento de íris. Fonte: (Jillela e Ross, 2015).

câmeras de vigilância de um aeroporto por exemplo, é o grande fator desafiador para o sistema biométrico.

1.2 Definição do Problema

As imagens utilizadas como entrada para um sistema de reconhecimento de íris compreendem a região em torno dos olhos, conhecida como região periocular. A região periocular possui componentes como: íris, pupila, esclera, reflexos, pálpebras, cílios, cabelo e sobrancelhas (Jillela et al., 2013). Alguns desses componentes não são usados no sistema de reconhecimento de íris e interferem diretamente no desempenho do mesmo, sendo necessário removê-los, de forma mais precisa possível. Na Figura 1.3, é possível observar os diversos tipos de ruídos presentes nas imagens da íris.



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

Figura 1.3: Exemplos dos ruídos que atrapalham a correta segmentação da íris. (a): reflexo e pouca oclusão pelas pálpebras; (b): reflexos, oclusão pela pálpebra superior e sombra; (c): reflexos e oclusão pela armação do óculos; (d): oclusão pela pálpebra superior e cabelo. Fonte: (De Marsico et al., 2015).

Diversas técnicas reportadas na literatura propõem um sistema biométrico completo (segmentação e reconhecimento), tendo como pioneira a proposta de Daugman (1993).

Entretanto, algumas concentram seus esforços apenas na tarefa de segmentação, com o intuito de melhorar o desempenho dos sistemas biométricos já existentes no âmbito da identificação/reconhecimento de indivíduos.

A segmentação visa localizar e separar corretamente a íris dos ruídos, das oclusões e do restante da imagem (ver Figura 1.3), por isso, a correta segmentação é a tarefa mais desafiadora de um sistema biométrico (Jan, 2017; Santos et al., 2015). Neste caso, o sistema ao invés de gerar as características referentes ao indivíduo e tentar reconhecê-lo, tenta localizar e/ou delimitar as regiões da imagem que pertencem ou não à região da íris. Normalmente esses sistemas geram, para uma imagem de entrada, uma imagem de saída (máscara de segmentação) com essas regiões da íris mapeadas, conforme ilustrado na Figura 1.4b.

(a) (b)

Figura 1.4: Resultado de segmentação esperado. (a): imagem de entrada para o sistema; (b): mapa de segmentação representando a saída esperada. Fonte: (De Marsico et al., 2015).

A Figura 1.4a mostra como a presença de reflexos pode atrapalhar a segmentação. Por isso esta etapa deve ser capaz de isolar esses reflexos do restante da íris.

Na literatura existem diversas técnicas que se propõem a resolver este problema usando as informações geométricas da íris. Essas técnicas utilizam detecção de bordas, binarização, crescimento de regiões, segmentos de contornos e detecção de círculos por exemplo.

No entanto, os resultados apresentados por essas técnicas funcionam apenas em domínios específicos (NIR ou *Visible* (VIS)) separadamente. Além disso, essas técnicas também dependem da qualidade e da quantidade de ruídos presentes nas imagens. Portanto, ainda existem problemas a serem resolvidos em relação à segmentação de íris, principalmente ao mesclar os domínios NIR e VIS.

É importante ressaltar que, muito recentemente, Zanlorensi et al. (2018) alcançaram resultados promissores no reconhecimento de íris sem o estágio de segmentação, isto é, utilizando apenas o bounding box da íris como entrada do sistema. No entanto, os experimentos foram realizados em uma única base de dados e uma análise mais detalhada é necessária.

1.3 Motivação

*Convolutional Neural Network*s (CNNs) são modelos computacionais com várias camadas, que têm a capacidade de aprender representações a partir de um conjunto de amostras de treinamento (Bengio et al., 2013). Após esse aprendizado, as CNNs podem fazer previsões em outras amostras a fim de reconhecê-las. O uso de CNNs tem permitido a obtenção de resultados estado-da-arte em diversos problemas de visão computacional como segmentação, detecção, reconhecimento e classificação em aplicações de biometria, imagens médicas, sistemas de segurança, monitoramento de trânsito, etc. (Severo et al., 2018; LeCun et al., 2015; Ahuja et al., 2017; Dumoulin e Visin, 2016; Krizhevsky et al., 2012; Laroca et al., 2018).

Especificamente na segmentação, as CNNs realizam uma classificação a nível de *pixel* na imagem. Essa classificação atribui uma classe a cada *pixel*, agrupando-os em regiões pertencentes ao mesmo objeto (Thoma, 2016). Em Shelhamer et al. (2015), os autores fazem uma adaptação de algumas CNNs, como por exemplo *Visual Geometry Group* (VGG) (Simonyan e Zisserman, 2014), onde usam uma técnica de transferência de aprendizagem deslocando as representações aprendidas por um modelo de CNN para afinar a segmentação, isto é, aprimorando-a.

As CNNs mostraram bom desempenho, ao serem empregadas em outras abordagens de segmentação, como em segmentação de estradas, segmentação de imagens da próstata, segmentação de nódulos de câncer e segmentação de esclera (Teichmann et al., 2016; Tian et al., 2017; Brandao et al., 2017; Lucio et al., 2018). Sendo assim, inspirado por esse bom desempenho de aprendizagem das representações e do reconhecimento das CNNs, este trabalho apresenta 5 abordagens de CNN, empregadas ao problema de segmentação de íris, utilizando 7 *datasets* disponíveis na literatura.

Esses *datasets* são distribuídos, de maneira que 4 possuem imagens obtidas no espectro NIR e os outros 3, no espectro VIS. As imagens NIR possuem uma maior representação das informações de textura da íris, porém, são mais próximas de um cenário/ambiente controlado. Já as imagens VIS possuem a presença de ruídos bem mais intensa, e próximas a um ambiente totalmente não controlado, que é muito mais desafiador, do que um cenário controlado.

Os modelos VGG (VggFCN-$fc7$ e VggFCN-*pool*5) e *Residual Network* (ResNet) (ResNet-50, ResNet-101 e ResNet-152) foram empregados, na forma de codificador, com a técnica de *Fully Convolutional Network* (FCN) (Shelhamer et al., 2015) combinada no decodificador, para refinar a segmentação. Essas 5 abordagens propostas utilizaram a técnica de transferência de aprendizagem e *fine-tunning* no processo de treinamento, devido à pouca quantidade de imagens presentes nos *datasets* utilizados. Além disso, essas abordagens propostas foram comparadas com 4 métodos *baseline* disponíveis na literatura.

1.4 Objetivos

Conforme a discussão do problema de segmentação de íris, e dos bons resultados alcançados pelas CNNs no âmbito da segmentação, este trabalho tem por objetivo geral a avaliação e proposta de aperfeiçoamento de redes convolucionais e redes residuais para segmentação de imagens de íris. Tais imagens são oriundas de ambientes cooperativos e não cooperativos em ambos os domínios (NIR e VIS), comparando esses modelos de redes,

com métodos que utilizam abordagens clássicas de segmentação, avaliando o desempenho das redes.

A partir disso, almeja-se os seguintes objetivos específicos:

- Testar 5 abordagens de segmentação de íris propostas nesta dissertação, duas a partir do modelo VGG e três oriundas do modelo ResNet.
- Comparar essas 5 abordagens com 4 métodos *baseline*.
- Empregar a estratégia de transferência de aprendizagem, no contexto de segmentação de íris.
- Analisar a capacidade de aprendizagem/discriminação das 5 abordagens propostas.
- Definir protocolos de treinamento/teste para realização dos experimentos.
- Mesclar os *datasets* de maneira a possibilitar a avaliação de convergência/generalização dessas abordagens.
- Rotular manualmente 2.431 imagens para permitir uma avaliação automática, quantitativa e objetiva.

1.5 Contribuições

De acordo com a definição apresentada anteriormente aos objetivos, este trabalho apresenta as seguintes contribuições:

- Proposta de 5 abordagens de CNN para a segmentação de íris.
- Avaliação destas abordagens em 7 *datasets* distintos, mostrando os desafios em cada domínio (NIR e VIS).
- Definição de protocolos de treinamento/teste dos modelos, com base na distribuição das imagens.
- Comparação entre os modelos e os métodos da literatura, usando 5 métricas diferentes.
- Disponibilização dos modelos gerados (pesos aprendidos dos 5 modelos de rede e código fonte desenvolvido `http://web.inf.ufpr.br/vri/databases/iris-segmentation-annotations/`.
- Disponibilização também das imagens rotuladas manualmente para fins de pesquisa acadêmica no site `http://web.inf.ufpr.br/vri/databases/iris-segmentation-annotations/`.

1.5.1 Artigos Publicados

1. (Bezerra e Menotti, 2017) Bezerra, C. S., Menotti, D. (2017) Fully Convolutional Neural Network for Occular Iris Semantic Segmentation. In XIII Workshop de Visão Computacional (WVC 2017), Qualis CC: B5;

2. (Severo et al., 2018) Severo, E., Laroca, R., Bezerra, C. S., Zanlorensi, L. A., Weingaertner, D., Moreira, G., Menotti, D. (2018). A Benchmark for Iris Location and a Deep Learning Detector Evaluation. In International Joint Conference on Neural Networks (IJCNN 2018), Qualis CC: A1

3. (Bezerra et al., 2018) Bezerra, C. S., Laroca, R., Lucio, D. R., Severo, E., Oliveira, L. F., Britto Jr, A. S., Menotti, D. (2018). Robust Iris Segmentation Based on Fully Convolutional Networks and Generative Adversarial Networks. In 31st Conference on Graphics, Patterns and Images (SIBGRAPI 2018), Qualis CC: B1

1.6 Estrutura do Documento

Esta dissertação está estruturada em 6 capítulos. No Capítulo 2 são apresentadas as principais técnicas utilizadas para segmentação da íris (estado-da-arte), incluindo as competições e os *datasets* existentes. Em seguida, o Capítulo 3, apresenta a descrição das abordagens propostas, como configurações de arquitetura, e os tipos de redes. No Capítulo 4 são descritos protocolos e métricas utilizados para comparação dessas abordagens com os métodos da literatura, as bases de dados de íris utilizadas nos experimentos, define-se também a especificação dos *datasets* analisados e os testes estatísticos para validação estatística das abordagens propostas. No Capítulo 5 os resultados de todos os experimentos são apresentados e discutidos, conforme distribuição dos *datasets*, validação estatística e análise qualitativa dos erros de segmentação. Por fim, no Capítulo 6 são apresentadas as conclusões obtidas com base no desenvolvimento deste trabalho, sobre os pontos relevantes identificados durante a revisão do estado-da-arte e de acordo com os experimentos realizados, incluindo uma breve apresentação de possíveis objetos de pesquisa na forma de trabalhos futuros.

2 Estado da Arte

Este Capítulo apresenta a descrição dos métodos de segmentação de iris presentes na literatura, de acordo com abordagens geométricas (Seção 2.1), abordagens em aprendizagem (Seção 2.2), algumas competições de segmentação (Seção 2.3). Por fim, *datasets* existentes são apresentados na Seção 2.4.

Diversas técnicas para segmentação de íris são reportadas na literatura. Muitas dessas técnicas são empregadas a um sistema de reconhecimento completo (segmentação e reconhecimento). Entretanto, resumiremos a seguir apenas as etapas de segmentação de íris de algumas dessas técnicas, de acordo com o escopo deste trabalho.

2.1 Abordagens Geométricas

O uso de uma abordagem geométrica para segmentar a íris foi inicialmente proposta por Daugman (1993), onde um operador integro-diferencial foi usado para aproximar os limites internos e externos da íris. Esse operador gera a saída x, y, r, onde x, y são as coordenadas do centro da pupila e da íris, e r o raio correspondente a ambas.

A *Hough Transform* (HT) modificada, com uma ordem de detecção reversa, foi utilizada para segmentar a íris em Liu et al. (2005). Primeiro são detectados os limites externos da íris (íris e esclera), e em seguida os limites internos (íris e pupila). Dentro da região da íris, os pontos de bordas ruidosos são reduzidos por meio da eliminação de *pixels* com nível de intensidade extremamente alto/baixo, de acordo com um *threshold* (limite) definido.

A segmentação de imagens da íris em ambientes não controlados proposta por Proença e Alexandre (2006) consiste em classificar cada *pixel* da imagem como pertencente a um grupo, com base nas coordenadas e distribuição de intensidade desses *pixels*. O agrupamento é feito por um algoritmo chamado *Fuzzy K-means*. Em seguida, o detector de bordas de Canny (1986) é aplicado na imagem com os *pixels* agrupados, gerando um mapa de bordas. A HT é usada para detectar as bordas correspondentes aos contornos da íris.

Contornos Ativos Geodésicos (GAC) foram usados em Shah e Ross (2009) para segmentar a íris a partir de estruturas circulares. Esses GAC combinam minimização de energia com contornos ativos baseados em evolução de curvas e são comumente utilizados para segmentação de imagens médicas. No caso da íris, a imagem primeiro passa por uma suavização, com um filtro de média em duas dimensões, seguida de uma binarização, que resulta em uma imagem com a pupila. A partir do ponto central e do raio da pupila, os contornos internos e externos da íris são aproximados de acordo com o uso dos coeficientes das séries de *Fourier*. A textura da íris, contribui para determinar seus limites, sendo influenciada pelas propriedades globais e locais da imagem.

Rankin et al. (2010) propuseram uma análise sobre o quanto a dilatação/contração da pupila interfere no reconhecimento de imagens da íris, pois isso altera visualmente os tecidos da íris. Uma substância é colocada no olho para dilatar a pupila, em seguida as imagens do mesmo olho (com a pupila em estado normal) são comparadas com essas imagens (com a pupila dilatada). A segmentação da pupila e da íris foi feita com os algoritmos propostos por Daugman (2004), e constataram que o grau de dilatação afeta significativamente o reconhecimento, com base na dissimilaridade intra-classe.

Conforme Podder et al. (2015) descrevem, os cílios e as pálpebras são ruídos que atrapalham o desempenho dos sistemas de reconhecimento de íris. Neste caso, esses ruídos são removidos ao aplicar a técnica de supressão radial máxima. O detector de bordas de Canny (1986) é aplicado na imagem para facilitar a detecção dos limites da íris e da pupila pela HT.

Vera et al. (2015) propuseram o uso do sistema embarcado *BeagleBone Black Rev C*, objetivando a redução de custos de implementação. A imagem da íris é binarizada com base em seu histograma para separar a pupila. As bordas da pupila e da íris são detectadas com o uso do operador Canny (1986) e da HT.

Em Pundlik et al. (2008) diferentes níveis de intensidade da imagem são usados para segmentar as regiões da íris. Primeiro, as reflexões são removidas com uma etapa de pré-processamento e informações de textura são calculadas conforme a variação de intensidade em uma vizinhança do *pixel*. Um mapa de probabilidades é gerado para atribuir o *pixel* a uma região de textura. Esse mapa de probabilidades é usado para binarizar a imagem, por meio de cortes gráficos, separando os *pixels* pertencentes aos cílios. As regiões da íris são estimadas usando a intensidade dos tons de cinza combinadas com operações morfológicas e refinadas por elipses.

Em Alkassar et al. (2016) a ideia é segmentar a esclera com base em suas formas de contorno. O operador integro-diferencial de Daugman (2004) e um método de contornos ativos sem bordas Chan e Vese (2001) são usados para segmentar as regiões da íris em dois arcos (esquerdo e direito). As pálpebras são detectadas por meio da intensidade de cores *Red, Green and Blue* (RGB) de cada *pixel*, dentro dos arcos pertencentes à íris, de acordo com uma heurística que usa um mapa de distância de cores. Duas sementes, uma à esquerda e outra à direita de cada arco da íris são usadas para representar a região da esclera. As coordenadas dessas sementes são definidas pelas coordenadas x, y centrais e o raio da íris.

A Tabela 2.1 apresenta as abordagens de segmentação que baseiam-se nas formas geométricas da íris, conforme detalhado anteriormente.

Tabela 2.1: Técnicas de segmentação que utilizam abordagens geométricas. Fonte: Autoria própria.

Artigos	**Técnica de Segmentação**
Liu et al. (2005)	*Hough Transform*
Proença e Alexandre (2006)	*Fuzzy K-means* e *Hough Transform*
Pundlik et al. (2008)	Cortes Gráficos e operações morfológicas
Shah e Ross (2009)	Contornos Ativos Geodésicos e *Fourier*
Rankin et al. (2010)	Operador integro-diferencial
Podder et al. (2015)	Supressão Radial Máxima e *Hough Transform*
Vera et al. (2015)	*Hough Transform* e Canny (1986)
Alkassar et al. (2016)	Operador integro-diferencial e Contornos Ativos

Além das abordagens que utilizam informações referentes a geometria da íris, da esclera e das pálpebras, comentadas acima, outras abordagens que baseiam-se em técnicas de aprendizagem também são utilizadas para segmentar a íris. Essas abordagens serão apresentadas na próxima seção.

2.2 Abordagens Baseadas em Aprendizagem

Uma abordagem que envolve a técnica de aprendizagem para detecção de face, em tempo real, foi proposta por Viola e Jones (2001). No trabalho de He et al. (2009), o algoritmo de detecção de face foi treinado com imagens de íris, com o intuito de localizar a posição do centro da íris. Um pré-processamento é feito na imagem da íris para encontrar pontos de reflexo, que são removidos por interpolação bilinear. Um modelo elástico chamado *Pulling and Pushing* definem os limites da íris. Em seguida as pálpebras são detectadas por meio de filtragem de histograma e a íris é então segmentada por meio da distribuição de intensidade dos seus *pixels*, agrupando-os de acordo com essa distribuição.

O algoritmo proposto por Li et al. (2012), faz uma busca de segmentos de contorno, e utiliza um classificador (regressão logística em *AdaBoost* cascata, chamado *Learned Boundary Detectors* (LBD)) treinado com os limites da pupila e da íris para eliminar bordas ruidosas. O detector de bordas de Canny (1986) é aplicado na imagem para destacar bordas candidatas. Essas bordas são agrupadas com base em informações de forma (circular). As bordas que não pertencem a essa formação circular, são consideradas ruídos e ignoradas. Após localizar a pupila, os limites esquerdos e direitos da íris são encontrados pelo classificador.

Outro algoritmo proposto por Badejo et al. (2016) para segmentar imagens da íris com baixo contraste (íris muito escuras) também utiliza uma abordagem de aprendizagem. A técnica de segmentação consiste em treinar o classificador *K-Nearest Neighbor* (KNN) (Cover e Hart, 1967) com características extraídas por um descritor de textura *Local Binary Pattern* (LBP) invariante à rotação, baseado em co-ocorrência proposto por Nosaka et al. (2013). São geradas amostras positivas (que contém íris/pupila) e negativas (não contém íris/pupila), que são utilizadas para treinar o KNN. Em seguida, o KNN classifica as regiões das imagens de teste em positivas e negativas. As regiões positivas (resultado da classificação) são agrupadas e marcadas, com um retângulo, como região pertencente à íris.

2.2.1 *Convolutional Neural Network*s (CNNs) para Segmentação de Íris

Considerando a segmentação de íris usando CNNs, Liu et al. (2016a) propuseram duas abordagens chamadas *Hierarchical Convolutional Neural Network*s (HCNNs) e *Multi-scale Fully Convolutional Network*s (MFCNs) para executar uma predição densa dos *pixels* usando janela deslizante combinando camadas rasas e profundas.

Em Jalilian e Uhl (2017) os autores propuseram três abordagens de utilizam FCN e codificador-decodificador para segmentar imagens da íris, baseando-se no modelo *Segmentation Network* (SegNet) (Badrinarayanan et al., 2015). A primeira abordagem é o modelo original da SegNet. Na segunda abordagem os autores utilizam uma variação abstrata da SegNet, enquanto na terceira abordagem a forma de predição (a nível de *pixel*) é feita com base em decisão bayesiana (Kendall et al., 2015).

Em Arsalan et al. (2018), os autores propõem uma abordagem chamada *Densely Connected Fully Convolutional Network* (IrisDenseNet). Essa abordagem determina os

limites da íris utilizando informações de gradiente entre os blocos densos da arquitetura de rede utilizada. Essa arquitetura baseia-se em codificador densamente conectado (*densely connected*) e decodificador SegNet (Badrinarayanan et al., 2015). As camadas da rede estão distribuídas entre blocos, onde cada camada possui conexão direta com as camadas subsequentes dentro do mesmo bloco. Os autores utilizaram técnicas de *data augmentation* para treinar essa abordagem, sem a necessidade de transferência de aprendizagem.

Em Bazrafkan et al. (2018), os autores propõem uma abordagem chamada *Semi Parallel Deep Neural Network* (SPDNN) para gerar mapas de íris a partir de imagens de baixa qualidade. Essa abordagem consiste em combinar vários modelos de *Deep Convolutional Neural Network*s (DCNNs) com intuito de aproveitar as vantagens de cada modelo específico pode proporcionar. Essa combinação dos modelos é realizada por meio de teoria dos grafos. Além disso, os autores também utilizaram *data augmentation* para o trenamento da abordagem.

Uma visão geral das técnicas de segmentação referentes à abordagens de aprendizagem, descritas nas Seções 2.2 e 2.2.1 é apresentada conforme a Tabela 2.2.

Tabela 2.2: Técnicas de segmentação que utilizam abordagens de aprendizagem. Fonte: Autoria própria.

Artigos	**Técnica de Segmentação**
He et al. (2009)	AdaBoost Cascata (Viola e Jones, 2001)
Li et al. (2012)	AdaBoost Cascata (regressão logística)
Badejo et al. (2016)	*Local Binary Pattern* (Nosaka et al., 2013)
	K-Nearest Neighbor
Liu et al. (2016a)	*Hierarchical Convolutional Neural Network*
	Multi-scale Fully Convolutional Network
Jalilian e Uhl (2017)	SegNet, SegNet com Variação Abstrata
	SegNet bayesiana
Arsalan et al. (2018)	*Densely Connected Fully Convolutional Network*
Bazrafkan et al. (2018)	*Semi Parallel Deep Neural Network*

2.3 Competições Sobre Segmentação de Íris

2.3.1 *Noisy Iris Challenge Evaluation - Part I*

Uma competição chamada *Noisy Iris Challenge Evaluation - Part I* (NICE.I) avalia os algoritmos voltados para segmentação de imagens da íris (da base de dados UBIRIS II, ver Seção 2.4.2), bem como a detecção de ruídos como reflexos, pálpebras e cílios por exemplo (Proença e Alexandre, 2007). Os competidores recebem 500 imagens para utilizar como entrada do sistema, a fim de treinar seus algoritmos de segmentação. Os algoritmos devem gerar, em sua saída para cada imagem de entrada, imagens binárias (máscaras) correspondentes apenas à região da íris, com os ruídos eliminados. O teste é feito com outras 500 imagens e suas máscaras (disponibilizadas na competição). A avaliação dos resultados de segmentação é feita conforme a Equação 4.1 descrita na Seção 4.2.

A NICE.I contou com 97 participantes de 22 continentes em 2008. Os 8 melhores resultados (menor taxa de erro) foram convidados a submeter um artigo descrevendo seu algoritmo. Esses melhores resultados serão descritos a seguir e estão apresentados na Tabela 2.3.

Tabela 2.3: Classificação da competição NICE.I. Fonte: Adaptado de (Proença e Alexandre, 2010).

Rank	**Autores**	**Técnica de Segmentação**	**Erro** (E %)
1	Tan et al. (2010)	Operador Integro-diferencial, Modelo de curvatura aprendido;	01,31
2	Sankowski et al. (2010)	Operador integro-diferencial, Modelagem paramétrica;	01,62
3	Almeida (2010)	Máscara probabilística, segmentos de arcos;	01,80
4	Li et al. (2010)	AdaBoost em cascata de Viola e Jones (2004), k-means com histograma de co-ocorrência, RANSAC;	02,24
5	Jeong et al. (2010)	Circular Edge Detection de ho Cho et al. (2006), AdaBoost do OpenCV, Janela local e kernel de convolução de Kang e Park (2007);	02,82
6	Chen et al. (2010)	*Hough Transform*, Detector de bordas Sobel;	02,97
7	Labati e Scotti (2010)	Operador integro-diferencial de Daugman (2004), Interpolação linear,	03,01
8	Luengo-Oroz et al. (2010)	Métodos de Morfologia matemática Serra (1983)	03,05

Algoritmo vencedor da NICE.I foi proposto por Tan et al. (2010) e consiste primeiramente em remover os pontos de reflexão por meio de limiarização adaptativa e interpolação bi-linear proposta pelos mesmos autores (no artigo (He et al., 2009)) (ver Seção 2.2). A localização grosseira da íris é dada pelo agrupamento e rotulação de regiões como pele, íris e sobrancelha, pois possuem diferença em sua estrutura. O agrupamento ocorre por meio de 8 vizinhos conectados, similar a um algoritmo de crescimento de regiões. Então, a região agrupada da íris é identificada, com base em sua forma, por descritores de *Fourier* (forma e momento). Em seguida, os limites da pupila e da iris são modelados pelo operador integro-diferencial (Daugman, 2004) modificado (chamado de *integro-differential constellation*). Essa modificação visa encontrar o caminho mais curto para maximizar a equação do algoritmo original, melhorando a convergência global. Os limites da pupila e da iris são refinados por meio de estatísticas de intensidade (proposto por Daugman (2004)). A oclusão da íris por parte dos cílios e das pálpebras são removidas respectivamente pela aplicação de um filtro horizontal 1-D e um modelo de curvatura aprendido. Esse modelo de curvatura encontra estatisticamente o melhor *threshold*, de acordo com a intensidade entre diferentes regiões da íris, para limiarizar a mesma, separando a região sobreposta pela sombra da pálpebra (oclusão) e deixando apenas a íris "limpa". Os erros de segmentação nos conjuntos de treino e de teste foram respectivamente 1,29% e 1,31% (ver Tabela 2.3). Segundo os autores Tan et al. (2010), apesar dos resultados alcançados, a segmentação de imagens de íris ruidosas ainda é um problema em aberto.

O segundo lugar da NICE.I pertence a Sankowski et al. (2010) onde a localização dos reflexos ocorre na imagem convertida para escala de cinza com base em *luminance in-phase quadrature* (YIQ), onde um *threshold* é calculado com base no histograma da imagem, gerando um mapa de reflexos. Cada *pixel* pertencente a essa região de reflexo é preenchido pela interpolação dos valores dos *pixels* em RGB baseando-se em 4 *pixels* vizinhos (esquerda, direita, superior e inferior). Conforme os autores Sankowski et al. (2010), esse preenchimento visa melhorar a tarefa de segmentação e não é utilizado para o reconhecimento. Primeiro encontra-se os limites externos da íris, em seguida, os limites internos, no componente de cor vermelho do espaço RGB, pois a pupila é melhor representada nesse componente. A localização da pálpebra inferior é feita pela aplicação

do detector de bordas de *Sobel* usando o kernel para bordas horizontais. A pálpebra superior é localizada com base na análise de pequenas regiões (esclera e pele) à esquerda e à direita da íris. A média dos valores RGB dos *pixels* de cada região é calculada, com isso, é possível identificar a transição entre o fim da esclera e o início da pálpebra. As coordenadas das duas transições são usadas para gerar uma linha, eliminando da íris a parte ocluída pela pálpebra. O erro de segmentação da abordagem proposta foi de 1,62%.

No terceiro lugar da NICE.I, a abordagem proposta por Almeida (2010) consiste em localizar os pontos de reflexão com base no valor de intensidade dos *pixels* acima de 250 nas imagens convertidas para escala de cinza. Em seguida, três canais de cores RGB são separados e utilizados para gerar uma imagem em escala de cinza com um contraste elevado. Esse aumento do contraste destaca os pontos pretos da pupila e clareia o restante da imagem, facilitando a detecção da pupila. Uma máscara probabilística foi aplicada, no canal vermelho, para destacar a pupila e a íris da região periocular. Essa máscara consiste em clarear regiões menos possíveis de ocorrer a íris, com base na análise de todas as imagens do conjunto de treinamento. Depois da filtragem pela máscara, a imagem é "percorrida" por alguns retângulos, de tamanho pré-definido, em busca da pupila. Quando ela é encontrada (pertencendo dentro do retângulo), algumas sementes são colocadas nessa região com o intuito de estimar o centro e o raio da pupila. A íris foi localizada da mesma forma. Os componentes de cor verde e azul foram usados para detectar as pálpebras, com base em segmentos de arcos. Duas sementes (a partir da posição e tamanho da pupila e da íris) determinam o centro dos arcos, que encontram as pálpebras superior e inferior com base na mudança de contraste entre a esclera e a pele. O erro de segmentação alcançado pelo algoritmo proposto foi de 1,8%.

Outro algoritmo competidor da NICE.I, proposto por Li et al. (2010) localiza a região dos olhos com o *framework* de (Viola e Jones, 2004) que utiliza *AdaBoost* em cascata para detecção de face. Em seguida a imagem do olho é recortada da região periocular. A técnica de agrupamento *k-means*, com base no histograma de co-ocorrência em níveis de cinza é aplicada na imagem recortada (do olho) para agrupar as regiões (íris, pálpebra, esclera, cílios). Após esse agrupamento, a aplicação do detector de bordas de Canny (1986) cria um mapa de bordas, na qual a transformada de *Hough* elíptica afina a região do olho. Na sequência, a transformada circular de *Hough* localiza a região da íris. O operador integro-diferencial de Daugman (1993) é aplicado dentro do círculo da íris para localizar a pupila. Uma técnica similar a RANSAC Fischler e Bolles (1981) é utilizada para selecionar possíveis parábolas candidatas à pálpebra. As pálpebras (superior e inferior) verdadeiras são localizadas por uma modificação do operador integro-diferencial de Daugman (1993) buscando detectar a melhor parábola dada pelo RANSAC. Por fim, o histograma da região segmentada da íris é calculado e filtrado com uma gaussiana a fim de obter um valor limiar equivalente aos reflexos. O erro na segmentação foi de 2,24% e resultou no quarto lugar no *rank* da competição.

O algoritmo proposto na competição por Jeong et al. (2010) converte a imagem em RGB para escala de cinza e detecta os limites interno e externo da íris por meio do uso da detecção de duas bordas circulares (*circular edge detection* - CED) proposto por ho Cho et al. (2006). Os pontos de reflexão são determinados por valores de intensidade dos *pixels* maiores que 250. Se a quantidade desses pontos de reflexão for maior que 1 a imagem é "classificada" como boa detecção, caso contrário má detecção. Quando ocorre a má detecção, de acordo com a classificação, o método de detecção baseado em AdaBoost da biblioteca OpenCV [1], é utilizado para forçar uma nova detecção. Verifica-se então essa

[1] http://opencvlibrary.sourceforge.net

nova detecção seguindo os mesmo critérios de classificação, e caso persista a má detecção, significa que o olho está fechado. Uma área de busca da pálpebra superior é definida com base na descontinuidade dos pontos na borda externa da íris. Uma máscara de detecção de pálpebra é aplicada nessa área de busca, com o intuito de identificar pálpebras candidatas. A transformada parabólica de *Hough* encontra a verdadeira pálpebra, entre as candidatas. Essa pálpebra é utilizada como referência para o método de detecção dos cílios. Esse método foi proposto por Kang e Park (2007) e consiste em destacar os cílios por meio de uma janela local e um *kernel* de convolução, com base nos valores de média e desvio padrão da intensidade dos *pixels* dentro da janela. O quinto lugar no *rank* da competição NICE.I foi obtido, com erro de 2,8%.

O sexto lugar da NICE.I pertence a Chen et al. (2010), onde a esclera é extraída com base no modelo de cores HSI, no qual a saturação é utilizada para encontrar um *threshold* que separa a esclera do restante da imagem. Em seguida um limiar adaptativo é utilizado para binarizar essa imagem, deixando apenas a área da esclera. As informações de coordenadas extremas (esquerda, direita, superior e inferior) da mesma são usadas para determinar uma área retangular com a localização do olho. O detector de bordas de *Sobel*, apenas na direção horizontal, é aplicado dentro dessa área retangular, gerando um mapa de bordas. Então, a transformada circular de *Hough* é utilizada para localizar o contorno externo da íris dentro do retângulo. Em seguida, o detector de bordas de *Sobel*, na direção vertical, é aplicado para gerar o mapa de bordas das pálpebras. No mapa de borda das pálpebras é aplicado a transformada linear de *Hough*, que define as pálpebras por meio de segmentos de linhas. Devido ao fato de que pode ocorrer íris com formatos não circulares, um método que corrige essas formas consiste em verificar o verdadeiro centro da íris (pupila) por meio de uma grade, dentro do círculo, onde a menor intensidade média em cada célula dessa grade, representa esse centro. Em seguida, os centros são ajustados e novos círculos são procurados, atribuindo votos a esses círculos. A intersecção de dois círculos e as linhas das pálpebras representam a verdadeira região da íris. Um *threshold* é aplicado nessa íris para remover os reflexos, e em seguida uma equalização de histograma eleva o contraste da íris. O detector de bordas de *Sobel* gera um mapa de bordas da íris e por fim, a transformada circular de *Hough* localiza o contorno da pupila. A abordagem proposta obteve um erro de segmentação de 2,9%.

As imagens coloridas são convertidas em escala de cinza e, em seguida, o operador integro-diferencial de Daugman (2004) é utilizado para localizar os pontos centrais dos limites internos e externos da íris. O gradiente radial da imagem é calculado em torno dos pontos centrais dos contornos. As informações desse gradiente são utilizadas para extrair duas partes da imagem (em forma de tiras), que correspondem a região dos contornos interno e externos da íris. Essas tiras são então linearizadas, com base em um algoritmo de interpolação linear proposto por Park e Chirikjian (2007), com o intuito de refinar os contornos da íris.

2.3.2 *Mobile Iris CHallenge Evaluation I*

A *Mobile Iris CHallenge Evaluation I* (MICHE-I) é uma base de dados que contém imagens da iris VIS (ver Seção 2.4.2), que foi criada para ser utilizada na primeira parte de uma competição de segmentação de íris.

Uma abordagem proposta por Haindl e Krupicka (2015) para segmentar as imagens da iris, consiste na utilização da Transformada de *Hough* generalizada de Ballard (1981) para localizar a região grosseira dos olhos. Em seguida, a modificação do operador integro-

diferencial de Daugman (1993) é aplicada nessa região para refinar os contornos externos da íris. Os contornos da pupila são localizados no canal vermelho da imagem, pelo operador integro-diferencial original. A íris é então normalizada, para posterior remoção dos ruídos (pálpebras e reflexos) nos canais vermelho e azul, respectivamente. Um polinômio de terceira ordem define os contornos das pálpebras superiores, enquanto que as inferiores são definidas pela estimativa de média e desvio padrão. Por fim, os reflexos são detectados, por meio da combinação entre limiarização adaptativa e modelos de textura Markovianos de Haindl (2012). A abordagem proposta foi testada nas bases de dados UBIRIS II e MICHE. Entretanto, pelo fato de que apenas a UBIRIS II possui as máscaras segmentadas, o método foi comparado com os oito melhores resultados da competição NICE.I (ver Tabela 2.3), seguindo os mesmos critérios de avaliação (dado pela Equação 4.1). O erro de segmentação na UBIRIS II foi de 1,24%, superando o algoritmo vencedor da NICE.I. Na MICHE o erro de segmentação foi analisado visualmente. Essa abordagem é utilizada por todos os participantes para segmentar as imagens da competição *Mobile Iris CHallenge Evaluation II* (MICHE-II), pelo fato da mesma ser voltada apenas para o reconhecimento.

Os autores Hu et al. (2015) propuseram o uso de histograma de correlação de *superpixels* por meio de *Simple Linear Iterative Clustering* (SLIC) para localizar a região grosseira da iris. Em seguida, essa região é binarizada com o método de *otsu*. Um quadrado, com tamanho pré-definido, é colocado nessa região binarizada (que pertence a íris), onde o centro desse quadrado representa o centro aproximado da íris. Vários possíveis raios são ajustados nesse quadrado, expandindo-se do centro para as bordas da imagem, em busca dos possíveis contornos externos da íris e da pupila. Cada contorno "candidato" recebe um valor de score dado por uma norma de regressão conforme o operador integro-diferencial de Daugman (1993). Esses contornos são então ajustados seguindo três modelos diferentes: um círculo e duas elipses. Características são extraídas com *Histogram of Oriented Gradients* (HOG) para treinar um *Support Vector Machines* (SVM), que define o melhor contorno. A localização das pálpebras superior e inferior ocorreu com base no algoritmo proposto por Li et al. (2010) (ver Tabela 2.3) que obteve o 4º lugar na competição NICE.I (ver Seção 2.3.1). Os *pixels* com reflexo são removidos por meio de um ajuste de contraste, no canal vermelho da imagem, quando o valor de intensidade desse *pixel* for maior que um *threshold*. O algoritmo proposto foi testado em quatro *datasets* (ver Seção 2.4): MICHE, UBIRIS I e II e FRGC, alcançando respectivamente os erros de segmentação, (dado pela Equação 4.1), de 1,93%, 1,43%, 1,39% e 1,37%.

Outra abordagem proposta por Raja et al. (2015), o olho é encontrado na imagem por meio de um *framework* para detecção de objetos, disponível em *Matlab*, com base em *Haar cascade*. Em seguida, um mapa de saliências, proposto por Cheng et al. (2015), explora a mudança de contraste presente nas bordas da imagem do olho, na qual a íris possui um alto contraste, sendo possível aproximar o diâmetro da mesma. A técnica de difusão anisotrópica invariante à rotação é utilizada para minimizar baixas mudanças de contraste, suavizando-as, a fim de refinar o mapa de saliências. Esse mapa de saliências refinado é então usado como entrada para o *framework Open Source Iris Recognition System* (OSIRISv4.1) (Sutra et al., 2012) que localiza precisamente os limites interno e externo da íris. Por fim, a íris é normalizada, conforme o algoritmo proposto por Daugman (2004). Os testes de acurácia na segmentação foram realizados por meio da comparação manual com o OSIRISv4.1 na base de dados MICHE e na base criada pelos autores chamada *Visible Spectrum Smartphone Iris Database* (VSSIRIS). O erro médio de segmentação alcançado pelo método proposto na MICHE foi de 7,64% e na VSSIRIS 1,96%. No OSIRISv4.1 o erro foi de 12,83% e 8,07% respectivamente.

O artigo de Abate et al. (2015) utiliza a transformada *watershed* para localizar a fronteira externa da iris e combina os scores da iris e da região periocular, por meio da soma, para melhorar a etapa de reconhecimento. Uma correção de cor e iluminação é feita na imagem em escala de cinza, por meio de um filtro gaussiano e interpolação linear no pré-processamento. O detector de bordas de *Sobel* e aplicado nessa imagem para realçar as bordas. Então, a transformada *watershed* agrupa a imagem em pequenas regiões, com base na uniformidade dos *pixels* de cada região, similar a um algoritmo de crescimento de regiões. Uma binarização é feita na imagem resultado da transformada *watershed* para separar as regiões de pele (mais claras) do restante da imagem. Os limites da iris e da pupila são refinados por meio de uma técnica de ajuste de círculos, no qual, cada círculo candidato recebe um voto, e o melhor votado é dado como limite verdadeiro. Os experimentos foram realizados na base de dados MICHE. Os resultados foram comparados com dois algoritmos. Um foi implementado pelos autores (o vencedor da NICE.I (He et al., 2009)) e o outro chamado de ISIS proposto por De Marsico et al. (2010). O centro e o raio da íris foram estimados na máscara gerada pela abordagem proposta e no *Ground-Truth* (GT) obtido manualmente. A divergência entre essas informações (centro e raio) de ambas imagens, resultam no erro de segmentação.

A abordagem proposta por Santos et al. (2015) utiliza a região periocular e a íris para o reconhecimento biométrico. As imagens da íris foram segmentadas por meio do uso das máscaras obtidas pelo algoritmo vencedor da NICE.I de Tan et al. (2010). Em seguida, as imagens segmentadas são normalizadas, pela transformação de suas coordenadas polares para cartesianas, por meio do tradicional método de Daugman (2004). Os experimentos foram realizados em um *dataset* proprietário, porém, disponível para a comunidade acadêmica, chamada de *Cross-Sensor Iris and Periocular* (CSIP) [2] (ver Seção 2.4.2 para maiores informações).

O artigo de Barra et al. (2015) propõe uma metodologia para autenticação de indivíduos em dispositivos moveis, por meio de histogramas espaciais. A segmentação das imagens da iris é feita com o algoritmo chamado ISIS proposto por De Marsico et al. (2010). O artigo aborda principalmente a etapa de reconhecimento, logo, ela não será tratada pois foge ao escopo deste trabalho. Os experimentos foram realizados nos *datasets* UBIRIS, UPOL [3] e MICHE.

Uma visão geral dos resultados e das técnicas de segmentação referentes aos artigos da competição MICHE-I, descritos na Seção 2.3.2, é apresentada na Tabela 2.4.

2.4 *Datasets*

As imagens utilizadas na análise de desempenho dos sistemas de reconhecimento de íris estão distribuídas em vários *datasets*. Esses *datasets* possuem imagens no espectro NIR e VIS. As imagens NIR são mais comuns, nestes sistemas biométricos, pois revelam informações relacionadas à textura da íris, ignorando as cores. Entretanto, as imagens VIS preservam informações de cor e são usadas em sistemas onde a colaboração do indivíduo não é necessária e/ou as imagens são capturadas a alguns metros de distância do sensor de aquisição (Jillela e Ross, 2016).

Imagens NIR fornecem um melhor desempenho no reconhecimento, em comparação com imagens VIS. Contudo, a maioria dos algoritmos existentes no estado-da-arte superam

[2] http://csip.di.ubi.pt

[3] Os links não estão mais disponíveis.

Tabela 2.4: Resultados e as técnicas de segmentação utilizadas pelos autores para a competição MICHE-I em 2015.

Autores	**Base Utilizada**	**Técnica de Segmentação**	*E* %
Haindl e Krupicka (2015)	MICHE, UBIRIS-II	*Hough Transform* (Ballard, 1981), Operador integro-diferencial (Daugman, 1993); Modelo de Textura Markoviano (Haindl, 2012)	01,24
Hu et al. (2015)	MICHE-I, UBIRIS I e II, FRGC;	Seleção de modelos (círculo e elipses), HOG, SVM;	01,93, 01,43, 01,39, 01,37;
Raja et al. (2015)	MICHE-I e VSSIRIS;	*Haar Cascade*, difusão anisotrópica, OSIRISv4.1;	07,64, 01,96;
Abate et al. (2015)	MICHE-I	Transformada *Watershed*, *Sobel*, Ajuste de círculos;	nd
Santos et al. (2015)	CSIP	Algoritmo vencedor da NICE.I (Tan et al., 2010)	nd
Barra et al. (2015)	UBIRIS, UPOL, MICHE-I	Algoritmo ISIS (De Marsico et al., 2010)	nd

as dificuldades dos *datasets* que usam imagens NIR. Imagens VIS são bem mais desafiadoras, principalmente por simular situações encontradas em ambientes do mundo real e possuir maior presença de ruídos (qualidade ruim, oclusões de cílios, pálpebras, óculos, olhos fechados) entre outros (Jan, 2017; Jillela e Ross, 2016). Os *datasets*, com imagem em ambos os espectros NIR e VIS, serão apresentados detalhadamente nas Seções 2.4.1 e 2.4.2 respectivamente.

2.4.1 *Datasets* no Espectro *Near Infra-red*

Um *dataset* NIR chamado *Chinese Academy of Sciences e Institute of Automation* (CASIA) (Tan e Sun, 2002) foi disponibilizado buscando promover pesquisas acadêmicas na área. Esse *dataset* possui várias versões com diferentes características:

- *Casia-Iris-v1* (CasiaV1): Imagens obtidas em ambiente controlado e com a cooperação dos indivíduos. Possui 108 olhos, 7 imagens de cada olho, totalizando 756 imagens, com extensão *.bmp* e resolução 320 × 280 *pixels*.
- *Casia-Iris-v2* (CasiaV2): Imagens obtidas em ambiente controlado, com dois dispositivos diferentes e em duas seções. Possui 60 indivíduos (classes) com 1200 imagens, totalizando 2400 com resolução de 640 × 480 *pixels*.
- *Casia-Iris-v3-Interval* (CasiaI3): Imagens obtidas em ambiente interno e em duas seções. Todas elas estão em escala de cinza 8 *bits*, com extensão *.jpeg*. Possui 249 indivíduos, com 395 classes, totalizando 2639 imagens com resolução de 320 × 280 *pixels*.
- *CASIA-Iris-LampV3* (CasiaL3): Imagem obtida em ambiente interno com variação de luz no momento da captura da imagem. Essa variação de luz provoca dilatações

e contrações na pupila, aumentando as variações intra-classe. Possui 411 indivíduos, 819 classes, totalizando 16.212 imagens com resolução de 640 × 480 *pixels*.

- *CASIA-Iris-TwinsV3* (CasiaTW3): imagens obtidas em ambiente aberto, 200 indivíduos (100 pares de gêmeos). Possui 400 classes, totalizando 3.183 imagens com resolução de 640 × 480 *pixels*.

- *CASIA-Iris-DistanceV4* (CasiaD4): Imagens obtidas em ambiente interno e capturadas a distância (3 metros) com a região da face. Pode ser utilizada em sistemas de biometria da região periocular ou multimodal (íris e face). Possui 142 indivíduos, 284 classes, totalizando 2.567 imagens com resolução de 2352 × 1728 *pixels*.

- *Casia-Iris-v4-Thousand* (CasiaT4): Imagens obtidas em ambiente interno, contendo a presença de óculos e reflexos. Possui 1.000 indivíduos e 2.000 classes, totalizando 20.000 imagens com resolução de 640 × 480 *pixels*.

- *CASIA-Iris-SynV4* (CasiaS4): Imagens da íris criadas sinteticamente a partir da CasiaV1. Essa sintetização inclui deformações, borrões e rotações. Possui 1.000 classes e indivíduos, totalizando 10.000 imagens com resolução 640 × 480 *pixels*.

Outra base de dados que possui imagens NIR é a *Iris Challenge Evaluation* (ICE) (Phillips et al., 2008), criada em 2005 para o primeiro desafio de reconhecimento de íris. A base contém imagens de 132 indivíduos, divididas entre olhos direito e esquerdo, com resolução de 640 × 480 *pixels* totalizando 2.953 imagens. Essas imagens foram obtidas em ambiente controlado e o sensor de aquisição manipulado manualmente. Entretanto, o *dataset* possui imagens de alta e baixa qualidade, inclusive com a presença de ruídos.

Com uma característica um pouco diferente dos dois *datasets* citadas à pouco, a *Dark-Skinned African Iris Dataset* (CuIris) (Badejo et al., 2011) foi criada com o intuito de testar algoritmos de segmentação e reconhecimento em imagens de íris com baixo contraste entre a íris e a pupila. Ou seja, imagens com a íris muito escuras (quase pretas). As imagens foram obtidas de pessoas do continente africano (devido a peculiaridade das íris bastante escuras), em um ambiente controlado (sensor de aquisição próprio), no espectro NIR, em escala de cinza e com resolução de 640 × 480 *pixels*. Entretanto, as imagens possuem presença de ruído, principalmente das pálpebras. No total são 2.058 imagens de 8 *bits*, distribuídas em 296 sujeitos e 432 classes.

2.4.2 *Datasets* no Espectro *Visible*

Um *dataset* que possui imagens VIS é a *University of Beira Interior Iris - I* (UBIRIS.v1) que foi criado em 2004 para simular diversas situações reais, como por exemplo, imagens desfocadas, oclusões, reflexos, sombras, que um sistema de reconhecimento de íris pode encontrar. A primeira seção UBIRIS.v1 é composta por 241 indivíduos, formando um total de 1877 imagens. Dessas imagens, dois *subsets* estão em RGB com resoluções de 800 × 600 e 200 × 150 *pixels* (ambos com 24 *bits*) e outro *subsets* em escala de cinza e resolução 200 × 150 *pixels*. Na segunda seção *University of Beira Interior Iris - II* (UBIRIS.v2) a captura das 11102 imagens com resolução de 800 × 600 *pixels* ocorreu de forma não controlada, ocasionando variações de luminosidade, foco, contraste e presença de reflexos (Proença e Alexandre, 2005; Proença et al., 2010).

O *University of Tehran IRIS* (UTIRIS) foi criado em 2007 e compreende imagens NIR e VIS com a finalidade de fundir as características da íris obtidas em ambos esses espectros, visando melhorar o desempenho no reconhecimento (Hosseini et al., 2010). UTIRIS é composto por 770 imagens NIR e 770 VIS, totalizando 1540 imagens dos olhos esquerdo e direito de 79 indivíduos, sendo distribuídas em 158 classes. As imagens NIR possuem dimensões de 1000×776 *pixels* e extensão *bmp*. A dimensão das imagens VIS é de 2048×1360 *pixels*, com extensão *.jpeg* e espaço de cores RGB.

A MICHE-I é uma base de dados que contém imagens da iris obtida VIS, que foi criada para ser utilizada na primeira parte de uma competição de segmentação de íris. Os sensores para a captura das imagens são dos dispositivos móveis: *iPhone 5* (IP5), *Galaxy Samsung IV* (GS4) e *Galaxy Tablet II* (GT2). As imagens da base de dados dos 92 sujeitos estão dispostas entre dispositivo e resolução, conforme (i) IP5: 1262 imagens de 1536×2048 *pixels*; (ii) GS4: 1297 imagens de 2322×4128 *pixels* e (iii) GT2: 632 imagens de 640×480 *pixels*. A MICHE-I possui diversos tipos de ruídos como: reflexos (naturais e artificiais), variação de foco, pálpebras, cílios, óculos, movimentos durante a captura das imagens, sombras, entre outros (De Marsico et al., 2015).

Uma visão geral das bases de dados descritas nas Seções 2.4.1 e 2.4.2 é apresentada conforme a Tabela 2.5.

Tabela 2.5: Características das bases de dados para biometria da iris.

Base	**Ambiente control.**	**Número Sujeitos**	**Número Imagens**	**Resolução (*pixels*)**	**Espec.**	**Ano**
CASIA-IrisV1	Sim	108*	756	320×280	NIR	2002
CASIA-IrisV2	Sim	60	2400	640×480	NIR	2004
CASIA-Iris-IntervalV3	Sim	249	2639	320×280	NIR	2010
CASIA-Iris-LampV3	Sim	411	16212	640×480	NIR	2010
CASIA-Iris-TwinsV3	Sim	200	3183	640×480	NIR	2010
CASIA-Iris-DistanceV4	Sim	142	2567	2352×1728	NIR	2010
CASIA-Iris-ThousandV4	Sim	1000	20000	640×480	NIR	2010
CASIA-Iris-SynV4	Sim	1000	10000	640×480	NIR	2010
ICE2005 (Phillips et al., 2008)	Sim	132	2953	640×480	NIR	2005
CUiris (Badejo et al., 2011)	Sim	296	2058	640×480	NIR	2011
UBIRIS I (Proença e Alexandre, 2005)	Não	241	1877	800×600 200×150	VIS	2005
UBIRIS II (Proença et al., 2010)	Não	261	11102	800×600	VIS	2010
UTIRIS (Hosseini et al., 2010)	Sim	79	1540	1000×776 2048×1360	NIR VIS	2007
MICHE-I (De Marsico et al., 2015)	Não	92	3191	1536×2048 2322×4128 640×480	VIS	2015

* Olhos.

2.5 Considerações Finais

Este capítulo apresentou uma breve descrição das abordagens clássicas de segmentação de íris presentes na literatura. Essas abordagens clássicas foram estruturadas em categorias, como abordagens que utilizam informações da geometria da íris e abordagens que utilizam técnicas de aprendizagem/treinamento. Competições de segmentação de íris (NICE.I e MICHE-I) também foram apresentadas, inclusive com as abordagens melhor colocadas nestas competições. Por fim, os *datasets* (NIR e VIS) utilizados em segmentação de íris foram descritos detalhadamente.

3 Abordagem Proposta

Conforme apresentado no Capítulo 2, as técnicas de segmentação de íris existentes na literatura possuem diferentes abordagens. Cada abordagem foi desenvolvida para uma finalidade específica e assim foi direcionada para um *dataset* específico, seja NIR ou VIS, controlado ou não controlado. Além disso, essas abordagens realizam em algum momento, um pré- ou pós-processamento nas imagens (Pundlik et al., 2008; Shah e Ross, 2009; He et al., 2009; Podder et al., 2015; Vera et al., 2015).

Abordagens que utilizam CNN para segmentação semântica, como em (Liu et al., 2016b) por exemplo, não necessitam de pré- ou pós-processamento. Neste capítulo, especifica-se detalhadamente as abordagens propostas para segmentar a íris utilizando modelos de CNNs.

Exploramos a arquitetura VGG (Simonyan e Zisserman, 2014), mais especificamente o modelo com 16 camadas (VGG16), distribuídas entre convolução e *pooling*, chamado de codificador, combinada com a arquitetura da FCN (Shelhamer et al., 2015), chamada de decodificador (VggFCN-$fc7$ e VggFCN-*pool*5). De forma similar, exploramos também 3 arquiteturas do modelo ResNet (He et al., 2016), mais especificamente os modelos com 50, 101 e 152 camadas, chamados respectivamente ResNet-50, ResNet-101 e ResNet-152.

Em resumo, propomos utilizar 5 abordagens codificador-decodificador que possuem uma arquitetura mista que combina cada um dos modelos VggFCN-$fc7$, VggFCN-*pool*5, ResNet-50, ResNet-101 e ResNet-152, com o modelo FCN para segmentar a íris em ambientes cooperativos e não-cooperativos. Essas arquiteturas codificador e decodificador são descritas, respectivamente, nas Seções 3.1 e 3.2.

3.1 Codificador *Visual Geometry Group*

O codificador da DCNN possui filtros de convolução com resolução (*receptive fields*) muito pequenos (i.e., 3×3 e 1×1), com o intuito de reduzir o número de parâmetros da rede. Segundo (Simonyan e Zisserman, 2014), 2 camadas de convolução com filtros 3×3 empilhadas possuem os *receptive fields* equivalentes a uma convolução com filtros 5×5. Da mesma forma, 3 camadas empilhadas, possuem os *receptive fields* equivalentes a uma convolução com filtro 7×7. A ideia base para a redução do número de parâmetros da rede é exemplificada da seguinte forma: 3 camadas de convolução, com C canais (filtros) 3×3, empilhadas possuem 3 $(3^2C^2) = 27C^2$ parâmetros. Já uma única camada, com filtro 7×7, possui $7^2C^2 = 49C^2$ parâmetros. Além disso, ao aumentar a profundidade (quantidade de camadas de convolução) da rede, aumenta-se também a representatividade dos mapas de características, devido ao tamanho dos *receptive fields* dos filtros (Simonyan e Zisserman, 2014).

A arquitetura do codificador da DCNN é apresentada na Figura 3.1, conforme a disposição das camadas de convolução, *max-pooling*, totalmente conectadas e *soft-max* ilustradas respectivamente em azul, vermelho, verde e amarelo. As 2 primeiras camadas de convolução (blocos ilustrados em azul) possuem 64 filtros com resolução 3×3 que geram uma saída (mapa de características resultantes) com a mesma resolução espacial (i.e., altura h e largura w) da entrada e 64 (C) canais. Cada camada de *max-pooling* realiza um *downsampling* por um fator de 2, enquanto que a quantidade de filtros das camadas de convolução após cada *max-pooling* são aumentadas também por um fator de 2 até o terceiro *max-pooling*.

Essa forma de construção faz com que a resolução espacial do mapa de características final seja $\frac{h \times w}{2^n}$ e a quantidade de canais seja $C_n = [64, 128, 256, 512, 512]$ onde n corresponde à n-ésima camada de *max-pooling* da rede VGG16. Por exemplo, uma imagem de entrada para o codificador com resolução espacial de 300×400 e 3 canais gera, de acordo com cada camada de *max-pooling* (n), as seguintes resoluções dos mapas de características: $n_i = 150 \times 200 \times 64$; $n_{ii} = 75 \times 100 \times 128$; $n_{iii} = 37 \times 50 \times 256$; $n_{iv} = 18 \times 25 \times 512$ e $n_v = 9 \times 12 \times 512$.

Em aplicações de reconhecimento de imagens em larga escala, finalidade do modelo VGG16, sua configuração original possui, após a quinta camada de *max-pooling*, 2 camadas totalmente conectadas (*fully connected*). Essas camadas totalmente conectadas (ilustradas em verde na Figura 3.1) representam o vetor de características da imagem de entrada. Esse vetor de característica é gerado a partir da aplicação de 4096 filtros, que possuem a mesma resolução do mapa de características resultante dessa quinta camada de *max-pooling*. Portanto, esse vetor de características possui a resolução de $1 \times 1 \times 4096$. Por fim, esse vetor de característica é classificado, por uma função de decisão *soft-max*, ilustrada em amarelo na Figura 3.1, de acordo com as 1000 classes presentes no *dataset ImageNet Large-Scale Visual Recognition Challenge* (ILSVRC) (Russakovsky et al., 2015), que é utilizado para treinamento do modelo VGG16.

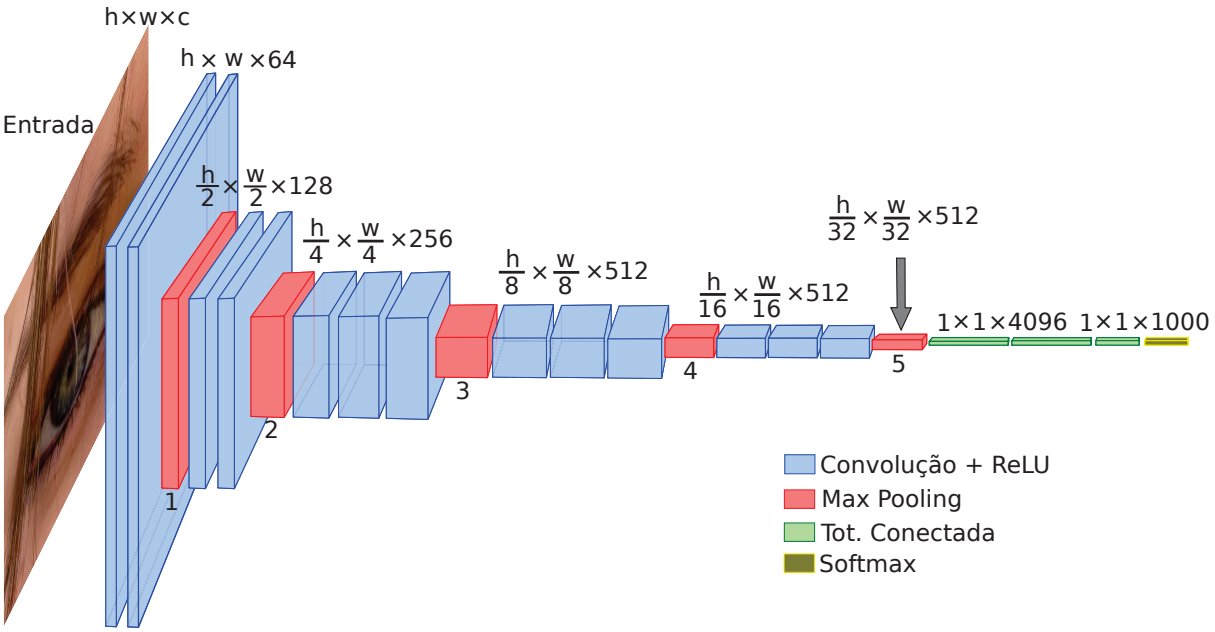


Figura 3.1: Arquitetura do modelo utilizado como base para o codificador das abordagens VggFCN-$fc7$ e VggFCN-*pool*5 para segmentação de íris. Fonte: adaptado de Simonyan e Zisserman (2014).

Entretanto, a configuração original do modelo VGG16, é utilizada pela comunidade acadêmica, especificamente para aplicações que envolvem reconhecimento de imagens, por meio de transferência de aprendizagem e *fine-tuning*.

As possíveis combinações da arquitetura do modelo VGG16 para geração dos mapas de características necessários para a realização do *fine-tuning* podem ser observadas na Figura 3.2. Essas combinações sugerem o uso do mapa de características após a saída de cada camada de *max-pooling* por causa da redução da resolução espacial desse mapa. As Figuras 3.2a e 3.2b por exemplo, ilustram (em vermelho) o mapa de características resultante da primeira e segunda camadas de *max-pooling*. Esses mapa de características possuem respectivamente, $\frac{1}{2}$ e $\frac{1}{4}$ da resolução da imagem de entrada e apenas 64 e 128 canais, quando comparada com outras redes como a ResNet, por exemplo.

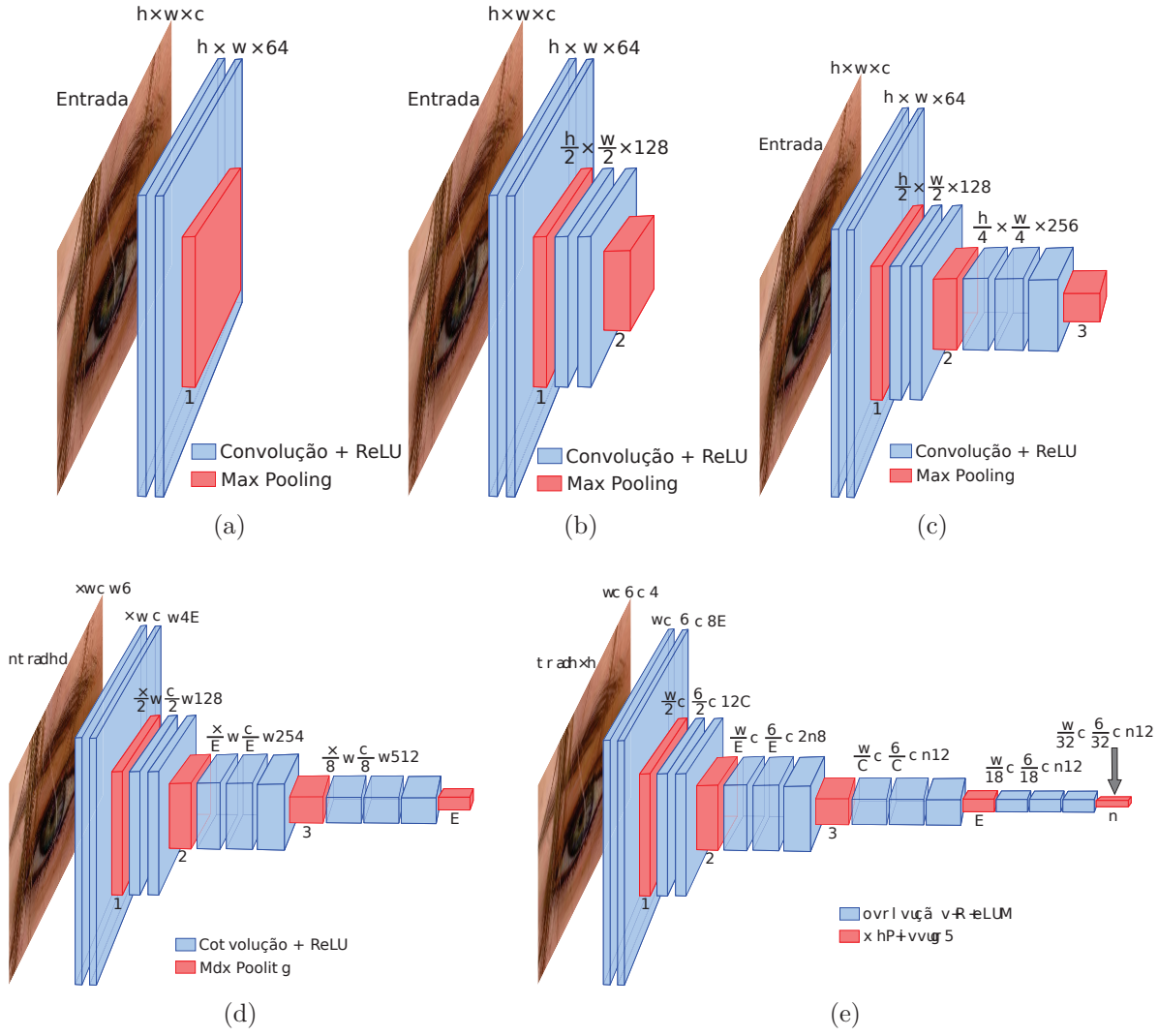


Figura 3.2: Possíveis combinações para realização do *fine-tuning* a partir do modelo VGG16 para o reconhecimento de imagens. Fonte: adaptado de (Simonyan e Zisserman, 2014).

As características presentes nesses mapas não são tão representativas, pois são oriundas apenas de convoluções aplicadas às camadas superficiais de uma CNN, que neste caso, os filtros aprendem apenas informações comuns a qualquer tipo de imagem, como bordas, contornos e formas (Simonyan e Zisserman, 2014). Entretanto, nos casos onde os mapas de características são resultantes da terceira, quarta e quinta camadas

de *max-pooling*, ilustrados em vermelho nas Figuras 3.2c, 3.2d e 3.2e, a quantidade de camadas de convolução e de filtros é bem maior. Isso faz com que essas características representem melhor as informações referentes a íris, ao contrário das informações comuns de qualquer imagem, comentado anteriormente.

Os pesos extraídos da quinta camada de *max-pooling* são disponibilizados[1] por Simonyan e Zisserman (2014) para realização do *fine-tuning* em diversas abordagens que utilizam CNNs voltadas para reconhecimento de imagens e extração de características (Minaee et al., 2016; Tarawneh et al., 2018). Para utilizar o modelo da VGG16 no contexto de segmentação de imagens, são necessárias algumas adaptações da arquitetura original. Essas adaptações são descritas na Seção 3.2.

3.2 Adaptação do Modelo *Visual Geometry Group* para *Fully Convolutional Network*

Conforme Shelhamer et al. (2015), a maioria das CNNs voltadas para o reconhecimento/detecção de objetos necessita de imagens de entrada com tamanho fixo, e produzem, por meio das camadas totalmente conectadas, uma saída sem informação espacial (vetor de características). Essas camadas totalmente conectadas também possuem tamanho fixo, descartando as coordenadas espaciais. Shelhamer et al. (2015) propuseram transformar essas camadas totalmente conectadas em totalmente convolucionais, e chamaram de *Fully Convolutional Network*.

Essa transformação consiste em realizar duas convoluções (*conv*6 e *conv*7), pois são duas camadas totalmente conectadas, com 4096 *kernels* (k) de tamanho 1×1 na quinta camada de *max-pooling* (*pool*5). Isso gera um mapa de características com resolução de $\frac{H}{32} \times \frac{W}{32} \times 4096$, independente da resolução ($H \times W$) da imagem de entrada. Em seguida, uma convolução é realizada nesse mapa de caraterísticas com 2 *max-pooling* 1×1. Isso gera um mapa de predição com resolução $\frac{H}{32} \times \frac{W}{32} \times 2$, onde 2 equivale a classe íris e *background*.

No entanto, devido ao processo de *downsampling* gerado pelas camadas de *max-pooling* no codificador, é necessário ampliar (*upsampling*) o mapa de predição até obter a mesma resolução espacial da imagem de entrada da rede. O *upsampling* consiste na aplicação de convoluções transpostas empilhadas, que possuem filtros aprendidos durante o treinamento, e que funcionam de forma similar a convolução realizada no codificador (Dumoulin e Visin, 2016). Essa capacidade de aprendizagem das camadas de convolução transposta estende-se também a forma de interpolação bilinear dos *pixels*.

Como o mapa de predição possui resolução $\frac{H}{32} \times \frac{W}{32}$ em relação a imagem de entrada ($H \times W$), a convolução transposta é realizada com *stride* de 32 *pixels*, para "compensar" a perda de informações ocasionadas pelo *downsampling*, mantendo o padrão de conectividade.

O nível de detalhe ao usar $stride = 32$ (FCN32) é baixo, gerando uma imagem resultante com uma segmentação grosseira, devido a interpolação. A Figura 3.3 representa o modelo VGG16, com as camadas totalmente conectadas, modificadas para totalmente convolucionais (*conv*6 e *conv*7), tornando a arquitetura do modelo FCN. Também é possível observar, nesta mesma figura, as imagens resultantes do *upsampling* dos mapas de predição em 3 escalas de refinamento (FCN32, FCN16 e FCN8), de acordo com cada valor de *stride*.

[1]Disponível em : `http://www.robots.ox.ac.uk/~vgg/research/very_deep/`

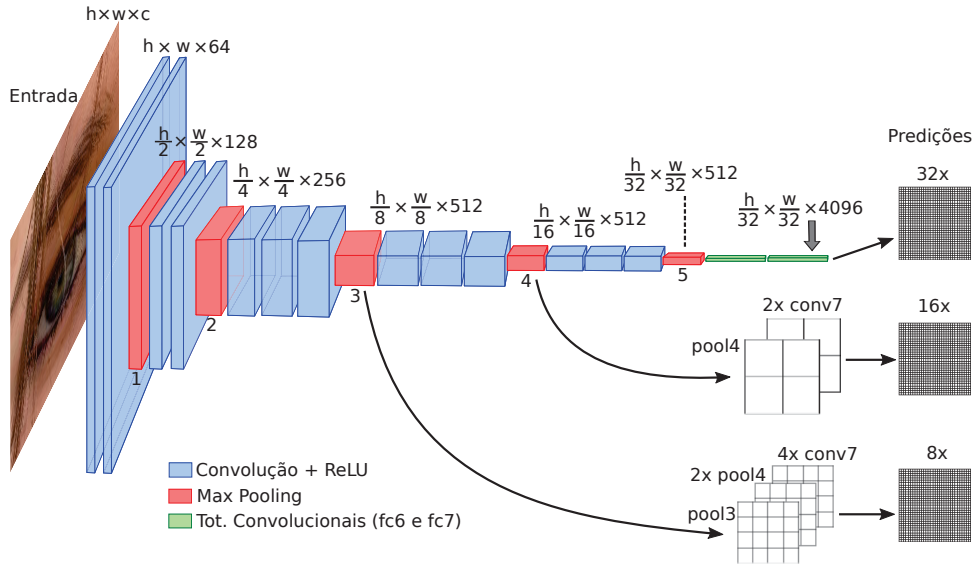


Figura 3.3: Arquitetura do codificador da VggFCN-$fc7$ adaptado do modelo VGG16 para FCN para segmentação de íris. Fonte: (Simonyan e Zisserman, 2014; Shelhamer et al., 2015).

A FCN utiliza informações das camadas de *max-pooling* do codificador, a fim de refinar o mapa de predição e melhorar a segmentação resultante. Essa combinação ocorre por meio de saltos (*skips*), apresentados na próxima seção (Seção 3.2.1).

3.2.1 Refinamento Por Meio de Saltos

O refinamento é possível pois a resolução dos mapas de características, em relação a resolução da imagem de entrada, é proporcional ao *max-pooling*, como discutido na Seção 3.1. Consequentemente, o valor do *stride* também é menor. A combinação permite a realização de predições que respeitem a estrutura global da imagem de entrada, pois as camadas são mais superficiais.

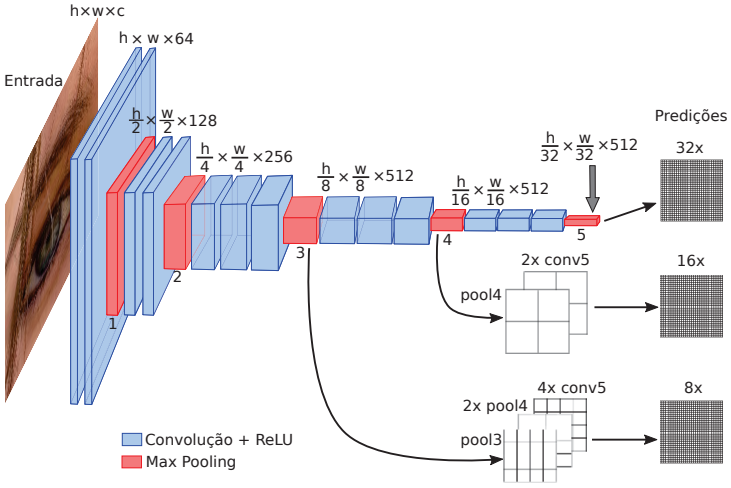
Na FCN16, o mapa de características da camada *pool*4 ($\frac{H}{16} \times \frac{W}{16} \times 512$) recebe uma convolução 1×1, resultando em $\frac{H}{16} \times \frac{W}{16}$. Em seguida o mapa de predição da camada *conv*7, (o mesmo com *stride*$= 32$) recebe 2 vezes o *upsampling* e é combinado, por meio da soma, com a predição da camada *pool*4, gerando a imagem de segmentação resultante, com *stride* igual a 16. Para a FCN8 a combinação ocorreu de forma similar. O mapa de características da camada *pool*3 é combinado com 4 vezes o *upsampling* da camada *conv*7 e 2 vezes o mapa de predição da camada *pool*4. Essa combinação também é representada na Figura 3.3.

Ao combinar as camadas *pool*2 e *pool*1, segundo Shelhamer et al. (2015), os resultados de segmentação foram piores, portanto não utilizaram essa combinação. Na FCN8 a presença de detalhes foram mais refinados, em comparação com a FCN16 e FCN32, portanto, utilizamos a FCN8 na abordagem proposta VggFCN-$fc7$ e VggFCN-*pool*5 para segmentar a íris.

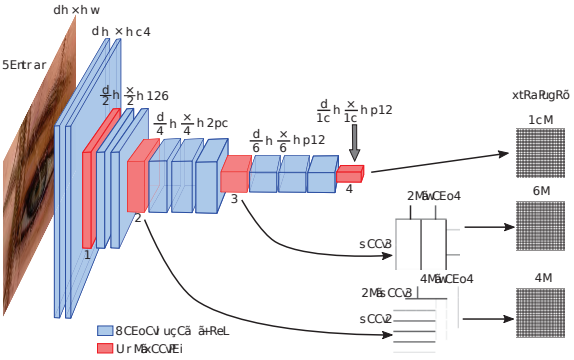
3.3 Possíveis Arquiteturas

Possíveis combinações de arquitetura para o codificador inspirado no modelo VGG16 são ilustradas conforme a Figura 3.4. Na Figura 3.4a a arquitetura de uma das

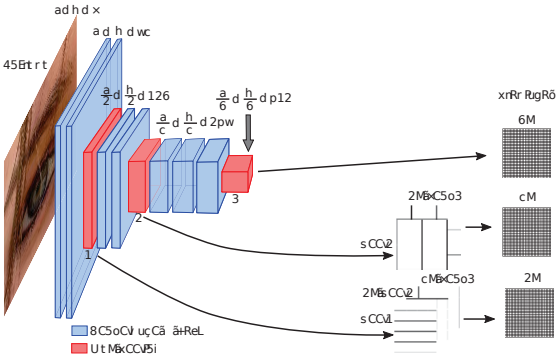
abordagens propostas (VggFCN-*pool*5) possui praticamente a mesma configuração da VggFCN-$fc7$ (Figura 3.3), diferindo apenas na remoção das duas camadas totalmente convolucionais.



(a) Arquitetura VggFCN-*pool*5.



(b) Combinação de arquitetura, realizando a predição a partir da camada *pool*4.



(c) Combinação de arquitetura, realizando a predição a partir da camada *pool*3.

Figura 3.4: Possíveis combinações de arquitetura do codificador das abordagens inspiradas no modelo VGG16 para segmentação de íris. (a) representa a arquitetura da VggFCN-*pool*5, na qual as camadas totalmente convolucionais foram removidas. (b) e (c) representam duas possíveis arquiteturas que podem ser utilizadas para segmentação de íris. Fonte: adaptado de Simonyan e Zisserman (2014); Shelhamer et al. (2015).

Observe que a Figura 3.4b apresenta, no mapa de predição, a proporção de 16×, ou seja 2^4. Essa proporção é dada de acordo com a resolução da imagem de entrada, e o *upsampling* é realizado nesse mapa de predição, e refinado com base nos saltos, usando informações das duas camadas de *pooling* anteriores, da mesma maneira que na Figura 3.3. Essa configuração segue de forma similar para a arquitetura apresentada na Figura 3.4c. Para o problema de segmentação, usando o modelo VGG16 e FCN, não é possível a obtenção de outras combinações dessa forma, pois o processo de *upsampling* que utiliza duas camadas de *pooling* fica comprometido.

3.4 Codificador *Residual Network*

O modelo ResNet utilizado neste trabalho tem por característica principal o uso de conexões de atalho entre as camadas de convolução. Essas conexões de atalho pulam uma ou mais camadas e estão dispostas em forma de blocos, chamados de bloco residual. Esses blocos possuem sempre 3 camadas de convolução com k filtros 1×1, 3×3 e 1×1 respectivamente para cada camada, conforme representado pela Figura 3.5. A estrutura desses blocos residuais é a mesma para os modelos ResNet-50, ResNet-101 e ResNet-152 utilizados neste trabalho. Dada uma entrada $\mathbf{x}$, o mapeamento residual $\mathcal{F}(\mathbf{x})$

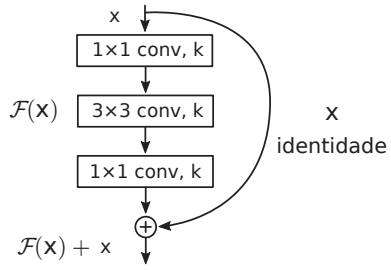


Figura 3.5: Estrutura do bloco residual dos modelos ResNet-50, ResNet-101 e ResNet-152. Fonte: Adaptado de He et al. (2016).

é denotado por $\mathcal{F}(\mathbf{x}) + \mathbf{x}$, ou seja, esse mapeamento soma a entrada de um bloco residual ($\mathbf{x}$ identidade) a saída resultante desse mesmo bloco, conforme ilustrado na Figura 3.5. Essa soma é realizada enquanto os mapas de características resultantes das camadas de convolução possuem a mesma resolução. Entretanto, quando as resoluções dos mapas de características são diferentes, essa soma ocorre com preenchimento extra de zero para aumentar a resolução dos mapas menores (He et al., 2016). Essas conexões de atalho não adicionam nenhum parâmetro extra na complexidade computacional do modelo. Além disso, elas possibilitam um aprendizado residual da rede, a fim de resolver o problema de treinamento presente em redes muito profundas. Esse problema refere-se a variações no gradiente de aprendizagem, fazendo com que haja apenas atualização dos pesos nas camadas mais profundas da rede, interferindo na convergência da mesma (He et al., 2016; Glorot e Bengio, 2010; Goodfellow et al., 2016).

Além disso, a ResNet possui uma quantidade de camadas de convolução empilhadas bem maior, em comparação ao modelo VGG16, como pode ser visto na Figura 3.6. Segundo os autores da ResNet (He et al., 2016), as características podem ser melhor representadas de acordo com o aumento do número de camadas empilhadas, por isso, essa maior profundidade. Outra característica da ResNet é possuir apenas uma camada de *max-pooling* logo após a execução da primeira convolução, e uma camada de *average-pooling* na antepenúltima camada, mais precisamente antes da camada totalmente conectada. O *downsampling* é realizado principalmente pelas camadas de convolução, por meio de *stride* $= 2$, ao contrário do modelo VGG16, que utiliza *max-pooling* (ver Seção 3.1 para maiores detalhes).

O modelo ResNet-50 é representado na Figura 3.6a, onde a imagem de entrada recebe 64 filtros de convolução (*conv*1), com *kernel* 7×7 e *stride* igual a 2. Em seguida, o mapa de características resultante dessa convolução, recebe um *downsampling*, por meio de *max-pooling* (*pool*1) com *kernel* 3×3 e também *stride* de 2. A saída de *pool*1 é utilizada como entrada para o primeiro bloco residual do modelo (ver Figura 3.5), que possui 3 camadas de convolução com *kernel* de tamanhos $1 \times 1 \times 64$, $3 \times 3 \times 64$ e $1 \times 1 \times 256$. Os

blocos residuais são conectados e agrupados formando blocos maiores chamados de *macro blocos*. A primeira camada de convolução de cada *macro bloco* utiliza *stride* = 2, para realizar o *downsampling* do mapa de características resultante do *macro bloco* anterior.

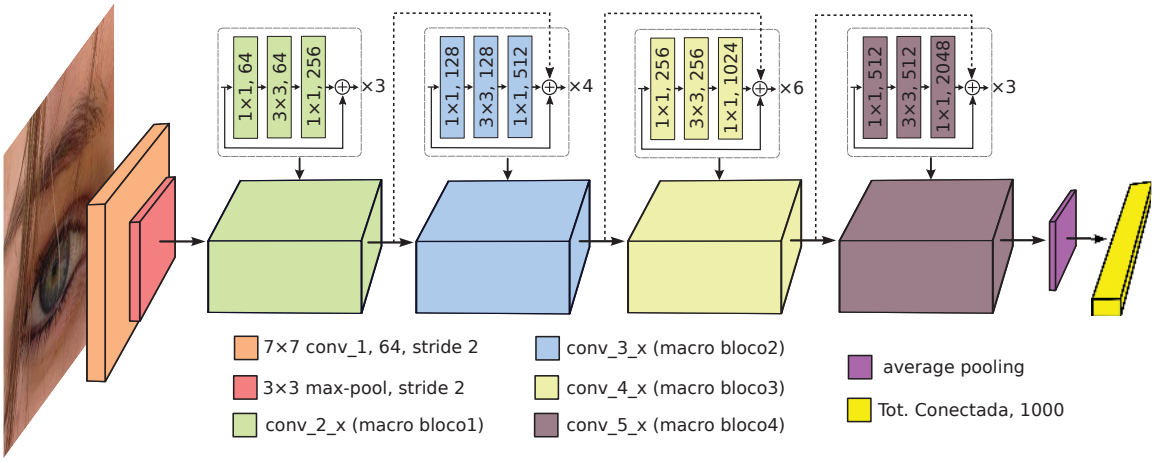
O primeiro macro bloco (*macro bloco*1) contém 3 blocos residuais, totalizando 9 camadas de convolução com a mesma resolução espacial e 3 conexões de atalho (uma em cada bloco residual). A saída do último bloco residual do *macro bloco*1 é conectada, por meio da conexão de atalho (ilustrada pelas linhas pontilhadas da Figura 3.6), à saída do primeiro bloco residual do *macro bloco*2. Esse padrão de conexão entre os *macro blocos* permanece o mesmo para todos eles. O *macro bloco*2 possui 4 blocos residuais, com *kernels* $1 \times 1 \times 128$, $3 \times 3 \times 128$ e $1 \times 1 \times 512$, totalizando 12 camadas de convolução.

O *macro bloco*3, por sua vez, possui 6 blocos residuais, com *kernels* $1 \times 1 \times 256$, $3 \times 3 \times 256$ e $1 \times 1 \times 1024$, com o total de 18 camadas de convolução, conectados como comentado anteriormente. O *macro bloco*4 possui a mesma quantidade de blocos residuais (3) que o *macro bloco*1, entretanto a quantidade de filtros é diferente, sendo ela 512, 512 e 2048. Após o último *macro bloco*, uma camada de *average pooling* realiza um *downsampling* no mapa de características. Por fim, a última camada do modelo ResNet é uma camada totalmente conectada, com um vetor de características $1 \times 1 \times 1000$, onde 1000 refere-se as classes presentes no *dataset* ILSVRC (Russakovsky et al., 2015). A função de classificação *softmax* é aplicada nessa camada totalmente conectada, gerando as probabilidades referentes a cada uma das 1000 classes.

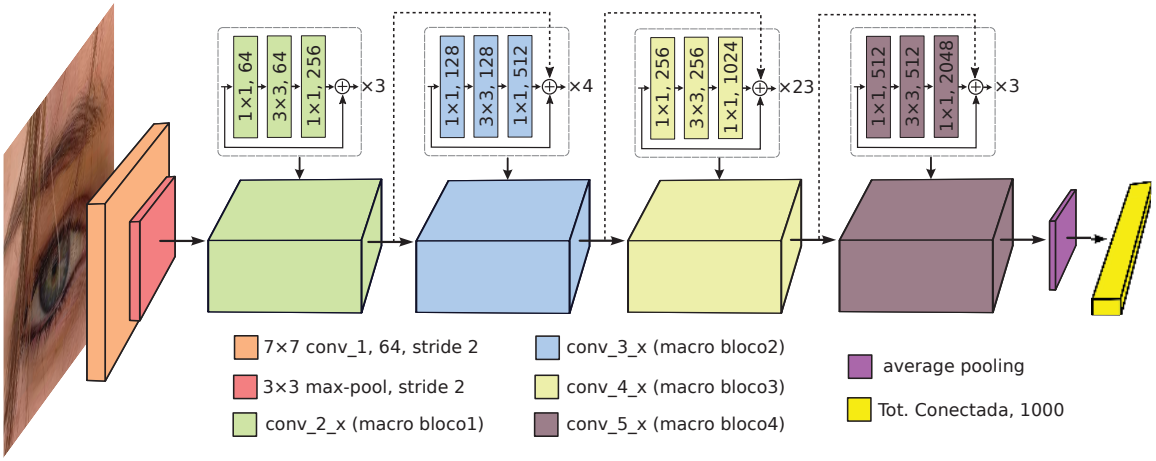
O que difere cada um dos 3 modelos da ResNet é a quantidade de blocos residuais agrupados nos macro blocos 2 e 3. Todos os blocos residuais são conectados e agrupados formando 4 blocos maiores chamados de macro blocos.

3.5 Considerações Finais

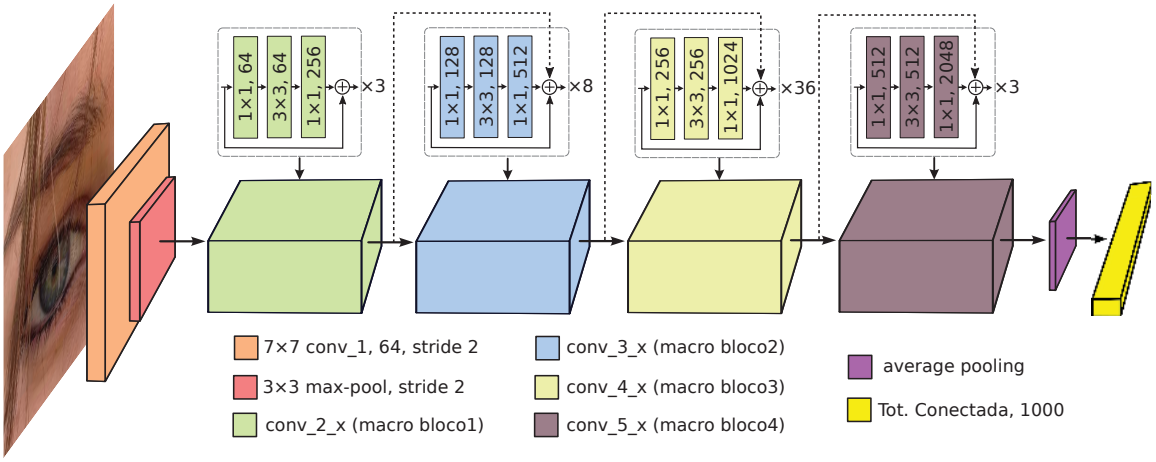
Este capítulo apresentou as arquiteturas e configurações das 5 abordagens propostas para segmentar imagens de íris. Inicialmente apresentamos as especificações dos modelos "base" (VGG16, ResNet e FCN) e em seguida descrevemos cada arquitetura utilizada, e suas características. Duas das abordagens propostas (VggFCN-$fc7$ e VggFCN-*pool*5) são inspiradas no modelo VGG16 e as demais (ResNet-50, ResNet-101 e ResNet-152) baseiam-se no modelo de redes residuais ResNet. Todas as abordagens propostas utilizam a arquitetura codificador-decodificador, onde o codificador baseia-se nos modelos VGG16 e ResNet, enquanto o decodificador utiliza FCN.



(a) Arquitetura ResNet-50.



(b) Arquitetura ResNet-101. Alteração para 23 blocos residuais no *macro bloco* 3, gerando o modelo com 101 camadas.



(c) Arquitetura ResNet-152. Alteração para 8 e 36 blocos residuais nos *macro blocos* 2 e 3 respectivamente, gerando o modelo com 152 camadas.

Figura 3.6: Arquitetura dos 3 modelos ResNet. As linhas de ligação em cada bloco residual (dentro do retângulo tracejado) representam as conexões de atalho referentes ao conjunto de 3 camadas de convolução. Os blocos maiores representam um conjunto de blocos residuais, chamado de macro bloco. Linhas de ligação pontilhadas representam as conexões de atalho entre as camadas de convolução com diferentes resoluções, de acordo com o processo de *downsampling*.

4 Experimentos

Neste capítulo, especifica-se detalhadamente os experimentos realizados para validar as abordagens de segmentação de íris propostas. A Seção 4.1 apresenta os detalhes do *software* e do *hardware* utilizados neste trabalho. As métricas utilizadas são descritas na Seção 4.2. Os *baselines* para comparação são apresentados na Seção 4.4. Os *datasets* utilizados para a execução dos experimentos são apresentados na Seção 4.3. As especificações referentes ao treinamento, como por exemplo, a taxa de aprendizagem, o algoritmo de otimização, a função de custo e a escolha da quantidade de iterações utilizadas são descritas na Seção 4.5. Por fim, os métodos para validação estatística são apresentados na Seção 4.6.

4.1 Especificação de *Hardware* e *Software*

Todos os experimentos foram realizados em um servidor com sistema operacional *Ubuntu* 16.04, processador *Quad-Core AMD Opteron*^{TM} *CPU* 8387, 64GB RAM, placa de vídeo *Nvidia TITAN Xp* 12GB GDDR5X de memória e 3840 CUDA®cores. A implementação ocorreu em *TensorFlow* 1.4.0, usando a linguagem *Python* 2.7.12 com a biblioteca de aprendizagem de máquina *Scikit-learn* 0.19.1.

4.2 Métricas

A avaliação dos resultados e de desempenho da abordagem proposta, contribui para a escolha do melhor modelo de classificação. Essa avaliação se dá por meio de algumas métricas, que serão descritas a seguir. Problemas de classificação que possuem apenas duas classes (positivas e negativas) são chamados de problemas binários. Cabe ao classificador inferir qual classe melhor representa a amostra testada. Essa inferência é definida como predição, e comumente é dada na forma de classe ou de probabilidade (Powers, 2011).

As predições do classificador são comparadas com as classes originais e os resultados dessa comparação compõem a matriz de confusão, ou também conhecida como tabela de contingência.

A matriz de confusão possui dimensões (x, y) igual ao número de classes. Nesse caso, por tratar-se de um problema binário, apenas quatro resultados são possíveis durante a predição (Fawcett, 2006):

- *True Positive* (TP): Uma amostra positiva classificada corretamente;
- *True Negative* (TN): Uma amostra negativa classificada corretamente;
- *False Positive* (FP): Uma amostra negativa classificada incorretamente;

- *False Negative* (FN): Uma amostra positiva classificada incorretamente.

A matriz de confusão é representada conforme a Figura 4.1, onde TN e TP equivalem às predições corretas, enquanto FN e FP equivalem aos erros de predição, com base na classe original. As Classes negativas e as positivas são representadas por 0 e 1 respectivamente. Neste trabalho, assume-se como classe negativa os *pixels* não pertencentes a íris (reflexos, pálpebras, cílios, pupila, pele, etc.), e como classe positiva, apenas os *pixels* pertencentes a íris.

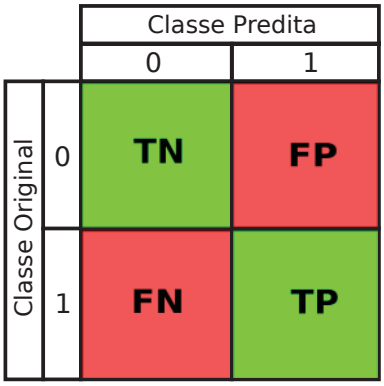


Figura 4.1: Matriz confusão de um problema de classificação binária. Fonte: Adaptado de (Powers, 2011).

Outra métrica utilizada segue os protocolos de avaliação da competição de segmentação de íris NICE.I (Proença e Alexandre, 2007). A avaliação é feita por meio da comparação *pixel*-a-*pixel* entre as máscaras binárias originais (GT produzidas manualmente pelos organizadores) e as máscaras geradas pelos algoritmos dos competidores. O erro médio de segmentação E é calculado pela divergência dos *pixels*, dada pelo operador $\otimes$ (XOR), da seguinte forma

$$E = \frac{1}{k \times h \times w} \sum_{i} \sum_{j} M(i,j) \otimes GT(i,j) \tag{4.1}$$

onde i e j são as coordenadas da máscara binária M gerada pelo algoritmo competidor, GT é o *Ground-Truth*, h e w são as linhas e colunas da imagem respectivamente, k é a quantidade de imagens. Observe que valores de E mais próximos de 0 representam melhores resultados (menor erro).

Outras métricas de avaliação comumente utilizadas em problemas binários: Precision (Prec) (Eq. 4.2), Recall (Rec) (Eq. 4.3), *F-Measure* (F1) (Eq. 4.4), e *Intersection over Union* (IoU) (Eq. 4.5) (Fritsch et al., 2013; Ge et al., 2006).

$$prec = \frac{TP}{TP + FP} \tag{4.2}$$

$$rec = \frac{TP}{TP + FN} \tag{4.3}$$

$$F1 = (1 + \beta^2) \frac{prec \cdot rec}{\beta^2 prec + rec} \tag{4.4}$$

Neste trabalho, optou-se por $\beta = 1$, isto é, o mesmo peso foi dado para $prec$ e rec.

A métrica utilizada para avaliar a segmentação, conforme a competição *Pascal Visual Object Classes* (VOC) (Everingham et al., 2015), é definida como IoU, dada pela Equação 4.5.

$$IoU = \frac{TP}{TP + FP + FN} \tag{4.5}$$

onde I é a intersecção entre os *pixels* preditos e o GT (máscara) da imagem e U é a união entre TP, FP e FN.

4.3 *Datasets* Utilizados

Na literatura existem diversos tipos de *datasets* que foram criados para os mais diversos fins (ver Seção 2.4). Neste trabalho, foram selecionados 7 (sete) *datasets* (*Biometrics and Security* (BioSec), CasiaI3, CasiaT4, *IITD Iris Image Database 1.0* (IITD-1), NICE.I, *Cross-Spectral Iris/Periocular* (CrEye-Iris) e MICHE-I), levando em consideração fatores como:

- *dataset* público;
- Imagens adquiridas em ambientes controlados e não controlados;
- Presença e ausência de ruídos (oclusão, reflexos, cílios, pálpebras);
- Imagens adquiridas no espectro NIR e VIS;
- Disponibilidade das máscaras (GT);

BioSec: *dataset* multimodal que contém impressão digital, face, íris e voz. Possui 3,200 imagens de íris, obtidas NIR de 25 sujeitos, com resolução de 640×480 *pixels* (Fierrez et al., 2007). Devido a problemas[1] com as máscaras de segmentação disponibilizadas, foi possível utilizar apenas as primeiras 400 imagens (ver Figura 4.3a).

CasiaI3: possui 2.639 imagens obtidas em ambiente interno/controlado, de 249 sujeitos NIR, com detalhes de textura da íris com poucos ruídos e resolução de 320×280 *pixels* (Tan e Sun, 2005a), (Figura 4.3b).

CasiaT4: contém 20.000 imagens NIR de 1.000 sujeitos, com resolução de 640×480 *pixels*, obtidas em ambiente interno, sob variações de iluminação (Tan e Sun, 2005b). (Figura 4.3c) As primeiras 1.000 imagens de 50 sujeitos foram rotuladas manualmente pelo autor.

IITD-1: possui 2.240 imagens NIR de 224 sujeitos com idade entre 14-55 anos, correspondendo a um total de 176 homens e 48 mulheres. Todas as imagens possuem resolução de 320×240 *pixels* obtidas em ambiente interno (Kumar e Passi, 2010), (Figura 4.3d).

NICE.I: é um *subset* da UBIRIS.v2 (Seção 2.4.2), criada para avaliar os métodos de segmentação de íris em imagens ruidosas. Possui 500 imagens para treinamento e 445 para teste, conforme o protocolo original. A resolução das imagens é de 400×300 *pixels*, obtidas em um ambiente não controlado, apresentando variação de iluminação, foco e presença de reflexos e óculos, (Figura 4.3e). Esse *dataset* é um dos mais desafiadores, no âmbito de segmentação de íris (Proença e Alexandre, 2012).

[1] Foi observado que a partir da 400 imagem, as máscaras não correspondiam com o restante do *dataset*.

CrEye-Iris: é composto por 3.840 imagens de 120 sujeitos. As imagens foram obtidas com um sensor adaptado para ambos NIR e VIS espectro, e divididas em três *subsets*: íris, máscara periocular (região em torno dos olhos) e ocular (apenas do olho) (Sequeira et al., 2016). As primeiras 1.000 imagens VIS com resolução de 400×300 *pixels* foram manualmente rotuladas, apenas do *subset* de íris, (Figura 4.3f).

MICHE-I: possui 3.191 VIS imagens de 92 sujeitos, sob configurações não controladas, usando três dispositivos móveis: IP5, GS4 e GT2, com (1.262, 1.297 e 632 imagens, respectivamente). Essas imagens possuem resolução de 1536×2048, 2320×4128 e 640×480 *pixels*, respectivamente (De Marsico et al., 2015). Foram utilizados 569 os GT fornecidos por Hu et al. (2015) e outros 431 manualmente criados pelo autor, para completar 1.000 imagens, (ver Figura 4.3g).

Uma visão geral dos *datasets* utilizados neste trabalho é apresentada na Tabela 4.1. Os GTs dos *datasets* BioSec, CasiaI3 e IITD-1 foram fornecidos por Hofbauer et al. (2014). Já os *datasets* CasiaT4, CrEye-Iris e MICHE-I foram rotulados manualmente pelo autor. Além disso, devido a segmentação de íris ser um problema binário, i,e., classe positiva

Tabela 4.1: Visão geral dos *datasets* utilizados neste trabalho. (*) significa que apenas parte do *dataset* foi utilizado. w e h representam respectivamente largura e altura das imagens.

Dataset	**Imagens**	**Sujeitos**	**Resolução** $(w \times h)$	**Espectro**
BioSec (Fierrez et al., 2007) (*)	400	25	640×480	NIR
CasiaI3 (Tan e Sun, 2005a)	2.639	249	320×280	NIR
CasiaT4 (Tan e Sun, 2005b) (*)	1.000	50	640×480	NIR
IITD-1 (Kumar e Passi, 2010)	2.240	224	320×240	NIR
NICE.I (Proença e Alexandre, 2012)	945	- - -	400×300	VIS
CrEye-Iris (Sequeira et al., 2016) (*)	1.000	120	400×300	VIS
MICHE-I (De Marsico et al., 2015) (*)	1.000	75	1536×2048 2320×4128 640×480	VIS

(íris) e classe negativa (não-íris), ilustramos a distribuição das classes na Figura 4.2 para todos os *datasets* utilizados neste trabalho.

Observe que todos os *datasets* possuem desbalanceamento das classes, principalmente pela grande quantidade de *pixels* negativos (barras ilustradas em azul na Figura 4.2). Entre os *datasets* NIR, a CasiaT4 possui apenas 5,84% dos *pixels* positivos, ou seja, de íris. Já entre os VIS, a MICHE-I possui apenas 1,19% dos *pixels* positivos. Uma justificativa para isso é que as imagens da MICHE-I possuem alta resolução espacial, como mostrado na Tabela 4.1.

A Figura 4.3 ilustra duas imagens de exemplo de cada um dos 7 *datasets* utilizados neste trabalho. É possível observar as diferenças entre as imagens NIR e VIS, como por exemplo, a qualidade superior da textura da íris na BioSec, em comparação com a MICHE-I.

Como pode ser observado na Tabela 4.1 e na Figura 4.3, as imagens dos *datasets* BioSec, CasiaI3, CasiaT4 e IITD-1 são similares pois foram adquiridas no espectro NIR. Devido a essa semelhança, decidiu-se considerar a mesclagem desses *datasets*, criando um novo *dataset* (NIR_{dt}) virtual. Da mesma forma, os *datasets* NICE.I, CrEye-Iris e MICHE-I são similares (VIS) portanto foram mesclados, criando virtualmente o VIS_{dt}. Além destas duas mesclagens citadas acima, uma nova junção entre ambos NIR e VIS resultou em outro *dataset*, chamado Todos_{dt}. A divisão dos *subsets* para treinamento e teste da abordagem proposta será explicada na próxima seção (Seção 4.3.1).

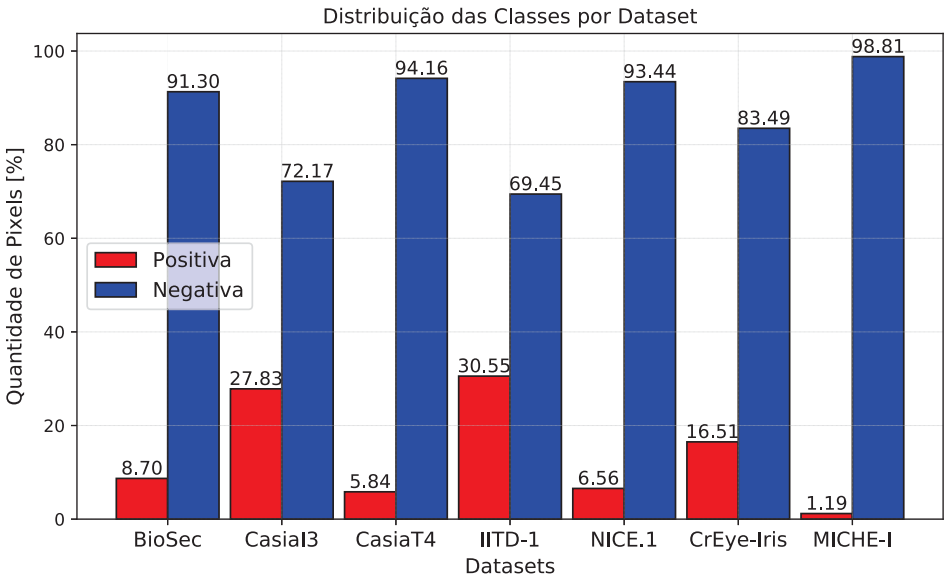


Figura 4.2: Distribuição das classes positivas e negativas para cada *dataset*. Barras em vermelho representam as classes positivas, enquanto as barras azuis representam as classes negativas.

4.3.1 Divisão dos *Datasets*

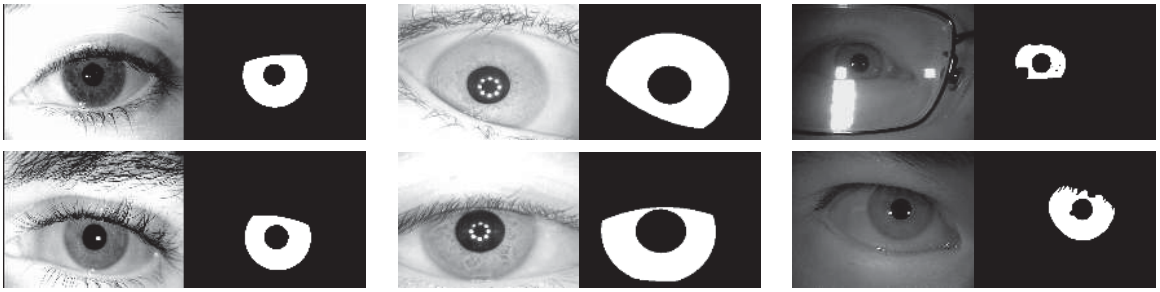
O único *dataset* que possui a divisão (treino e teste) pré-estabelecida é o NICE.I, conforme especificado na competição (Proença e Alexandre, 2007). Por isso, um experimento inicial foi realizado de acordo com essa divisão, com o intuito de estabelecer as configurações necessárias para o treinamento dos modelos de DCNN propostos (ver Seção 4.5).

Inicialmente as imagens do treinamento do *dataset* NICE.I (apenas 500 imagens) foram divididas aleatoriamente entre treinamento e validação, seguindo respectivamente a proporção de 80% e 20% conforme a distribuição de Pareto (Newman, 2005). Essa distribuição foi observada por Pareto, onde notou que normalmente 20% das amostras representam 80% da população em questão (Newman, 2005). Chen et al. (2018) propuseram uma abordagem para a segmentação do ventrículo esquerdo em imagens de ressonância magnética cardíaca, também utilizando CNNs e a mesma distribuição de Pareto, portanto adotamos essa distribuição como válida para essa divisão.

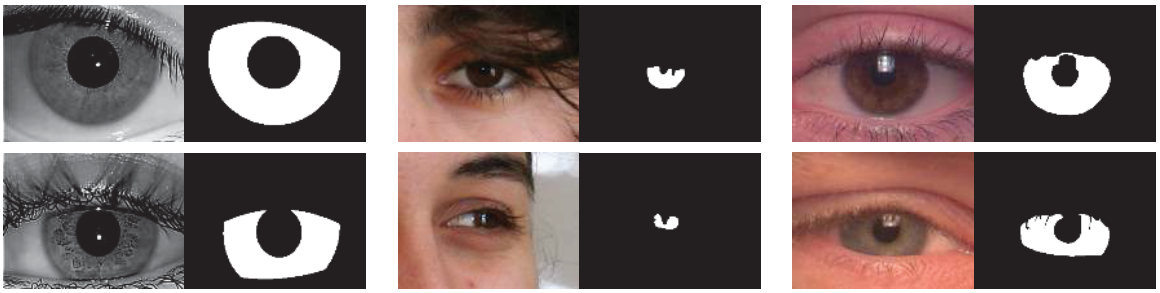
Subdividimos o treinamento entre *subsets* de treino e validação em 5-*folds* (Hastie et al., 2001), a fim de verificar se essas amostras garantem um treinamento satisfatório das abordagens propostas. Os resultados desses 5 *folds* para cada *dataset*, com base nos modelos VggFCN-$fc7$ e ResNet-50 são apresentados na Tabela 4.2.

Observe que o modelo VggFCN-$fc7$ obteve valores de F1 acima de 97% para os *datasets* BioSec, CasiaI3 e IITD-1, enquanto o CasiaT4 o resultado foi um pouco menor (95,83%). Já nos *datasets* NICE.I e CrEye-Iris, os resultados de F1 foram respectivamente 94,10%, 96,95%, enquanto que o MICHE-I, por possuir muito mais informações de ruídos e ser mais desafiador, esse resultado foi de 90,79%.

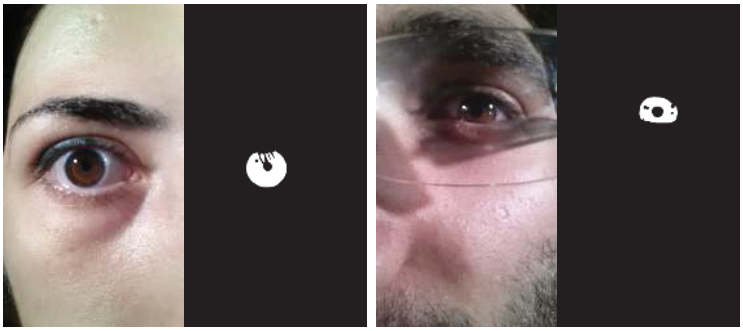
Já o modelo ResNet-50 apresentou F1 acima de 98% para os *datasets* BioSec e CasiaI3, $\approx$ 97% para os *datasets* CasiaT4, IITD-1 e CrEye-Iris. Para o *dataset* MICHE-I o F1 foi de 92,72%. De maneira geral, a ResNet-50 obteve valores de F1 pouco maiores do que VggFCN-$fc7$.



(a) BioSec (b) CasiaI3 (c) CasiaT4



(d) IITD-1 (e) NICE.I (f) CrEye-Iris



(g) MICHE-I

Figura 4.3: Exemplos dos sete *datasets* utilizados neste trabalho (à esquerda), com seus respectivos GT (à direita). (a), (b), (c) e (d) são os *datasets* NIR, enquanto que (e), (f) e (g) são VIS. Fonte: o autor.

Tabela 4.2: Resultados da validação cruzada (*cross-validation*) com 5 *folds* dos modelos VggFCN-$fc7$ (a) e ResNet-50 (b) com base na métrica F1 para os *datasets* BioSec, CasiaI3, CasiaT4, IITD-1, NICE.I, CrEye-Iris e MICHE-I.

(a) VggFCN-$fc7$

Datasets	**F1 %**
BioSec	$97{,}15 \pm 00{,}21$
CasiaI3	$97{,}96 \pm 00{,}08$
CasiaT4	$95{,}83 \pm 00{,}08$
IITD-1	$97{,}68 \pm 00{,}11$
NICE.I	$94{,}10 \pm 00{,}22$
CrEye-Iris	$96{,}95 \pm 00{,}15$
MICHE-I	$90{,}79 \pm 01{,}27$

(b) ResNet-50

Datasets	**F1 %**
BioSec	$98{,}47 \pm 00{,}32$
CasiaI3	$98{,}45 \pm 00{,}13$
CasiaT4	$97{,}44 \pm 00{,}10$
IITD-1	$97{,}78 \pm 00{,}25$
NICE.I	$95{,}54 \pm 00{,}18$
CrEye-Iris	$97{,}90 \pm 00{,}11$
MICHE-I	$92{,}72 \pm 00{,}98$

Com base nos resultados da Tabela 4.2, assume-se que as amostras escolhidas para o treinamento das abordagens propostas VggFCN-$fc7$ e ResNet-50 representam satisfatoriamente todos os *datasets* utilizados, devido ao baixo desvio padrão entre os 5 *folds*, mesmo no *dataset* MICHE-I. Portanto essas amostras foram utilizadas para todos os treinamentos. As demais abordagens propostas apresentaram valores semelhantes, por isso, utilizamos os modelos VggFCN-$fc7$ e ResNet-50 para representar esse treinamento. Na próxima seção (Seção 4.5), discutimos o treinamento das abordagens propostas, bem como os critérios para escolha da quantidade de iterações.

4.4 *Baselines* para Comparação

Foram selecionados 4 *frameworks*, utilizados como *baseline*, para comparar os resultados obtidos entre os modelos da abordagem proposta (VggFCN-$fc7$, VggFCN-*pool*5, ResNet-50, ResNet-101 e ResNet-152) e o estado da arte. Esses *frameworks* estão disponíveis na literatura em forma de código-fonte e arquivo executável. São eles: OSIRISv4.1, IrisSeg, Haindl e TVM.

O **OSIRISv**4.1 é composto de quatro módulos principais: segmentação, normalização, extração de características e *matching* (Othman et al., 2016). Entretanto, apenas o módulo de segmentação foi utilizado para comparação dos resultados. O OSIRISv4.1 possui parâmetros de entrada, como diâmetro mínimos e máximos da íris e da pupila em *pixels*. Esses parâmetros foram ajustados visualmente para cada *dataset*, a fim de obter os melhores resultados. A Tabela 4.3 apresenta os valores dos parâmetros ajustados de acordo com cada *dataset*. Embora o desempenho do OSIRISv4.1 tenha sido relatado apenas em *datasets* com imagens NIR, os experimentos foram executados em ambos NIR e VIS.

Tabela 4.3: Parâmetros do OSIRISv4.1 utilizados para segmentar a íris. **MinDp**, **MaxDp**, **MinDi** e **MaxDi** representam respectivamente os diâmetros mínimo e máximo da pupila e da íris.

Dataset	**MinDp**	**MaxDp**	**MinDi**	**MaxDi**
BioSec	50	140	150	200
CasiaI3	50	126	150	180
CasiaT4	50	126	150	180
IITD-1	50	126	150	180
NICE.I	50	90	100	180
CrEye-Iris	50	126	150	180
MICHE-I	50	126	100	900

O ***Iris Segmentation Framework*** **(IRISSEG)** (Gangwar et al., 2016) foi projetado especificamente para íris em condições não ideais (íris com fatores de má qualidade) e utiliza como base filtragem adaptativa, seguindo uma estratégia *coarse-to-fine*. A pupila e a íris são detectadas de forma grosseira, por meio de um método de busca iterativa, que analisa contornos candidatos a ambas. Em seguida, uma análise geométrica e de textura refina e ajusta os melhores contornos. Os autores enfatizam que o IRISSEG não requer ajuste de parâmetros para diferentes *datasets*. Assim como no OSIRISv4.1, os experimentos foram executados em ambos NIR e VIS *datasets*.

O **HAINDL** (Haindl e Krupicka, 2015) foi desenvolvido para imagens coloridas obtidas através de dispositivos móveis sem restrições, e é utilizado como algoritmo base de segmentação na competição de reconhecimento de íris chamada MICHE-II (Marsico et al.,

2017). Apenas os resultados obtidos pelo **HAINDL** nos *datasets* VIS foram comparados com as DCNNs, pois nos *datasets* NIR, o executável fornecido pelos autores não funcionou devido as imagens terem apenas uma banda.

Por fim, o ***Total Variation Model* (TVM)** (Zhao e Kumar, 2015) conforme os autores afirmam, esse método segmenta imagens de íris a partir de imagens faciais em ambos os espectros NIR e VIS. O método propõe melhorar um descritor de textura, regularizando a variação local dos *pixels*. Essa melhoria ocorre com base na regularização da norma l^1, que faz com que principalmente as regiões de ruído sejam suprimidas. Com isso, as partes do olho com regiões de texturas parecidas (i.e., limite das pálpebras, da esclera e da pupila) podem ser localizadas por meio da Transformada Circular de *Hough*. Além disso, o método também utiliza estratégias de pré- e pós-processamento, como correção da variação de iluminação por exemplo, para refinar os resultados da segmentação.

4.5 Treinamento das *Deep Convolutional Neural Network*s

O treinamento das abordagens propostas foi realizado com taxa de aprendizagem de 10^{-5} e o algoritmo de otimização *Adam optimizer*, visando minimizar o erro da função de custo, com base em descida de gradiente (Teichmann et al., 2016). A função de custo da segmentação é dada pela entropia cruzada (*cross-entropy*) conforme:

$$loss(p, q) = -\frac{1}{|I|} \sum_{i \in I} \sum_{c \in C} q_i(c) \log p_i(c) \tag{4.6}$$

onde p é a predição, q é o GT e C é a classe. O codificador de segmentação são inicializados de acordo com os pesos pré-treinados da ImageNet (Krizhevsky et al., 2012). O *weight decay* com fator de $5 \cdot 10^{-4}$ foi usado em todas as camadas.

4.5.1 Escolha das Iterações

A Figura 4.4 representa a curva de aprendizagem dos 5 modelos da abordagem proposta, para $128t$ iterações, com base na métrica F1 (Eq. 4.4) obtida nos *subsets* de validação. Inicialmente as DCNNs foram treinadas com $16t$ iterações, onde $t = 1000$, similarmente ao original (Teichmann et al., 2016). Entretanto, ao realizar uma análise visual da curva de aprendizagem, no conjunto de validação, das DCNNs apresentada na Figura 4.4, notou-se que essa curva ainda tende a convergir. Buscando encontrar a quantidade de iterações ideal para a convergência das DCNNs, o treinamento foi realizado dobrando a quantidade de iterações, a partir dos $16t$, para $32t$, $64t$ e $128t$.

Observe que todos os 5 modelos obtiveram comportamento semelhante a partir da iteração $28t \sim 32t$ até $128t$, portanto utilizamos a quantidade fixa de $32t$ iterações (linha pontilhada) para o treinamento de todas as abordagens.

4.6 Análise Estatística dos Resultados

O teste aqui apresentado busca avaliar o desempenho dos *baselines* e das abordagens propostas, com base em análise de significância estatística. Dada uma estimativa a, calcula-se um *p-value*, que é a probabilidade de se obter uma medida a ou superior, dada uma hipótese nula H_0. Então, compara-se esse *p-value* com um nível de confiança de 95%

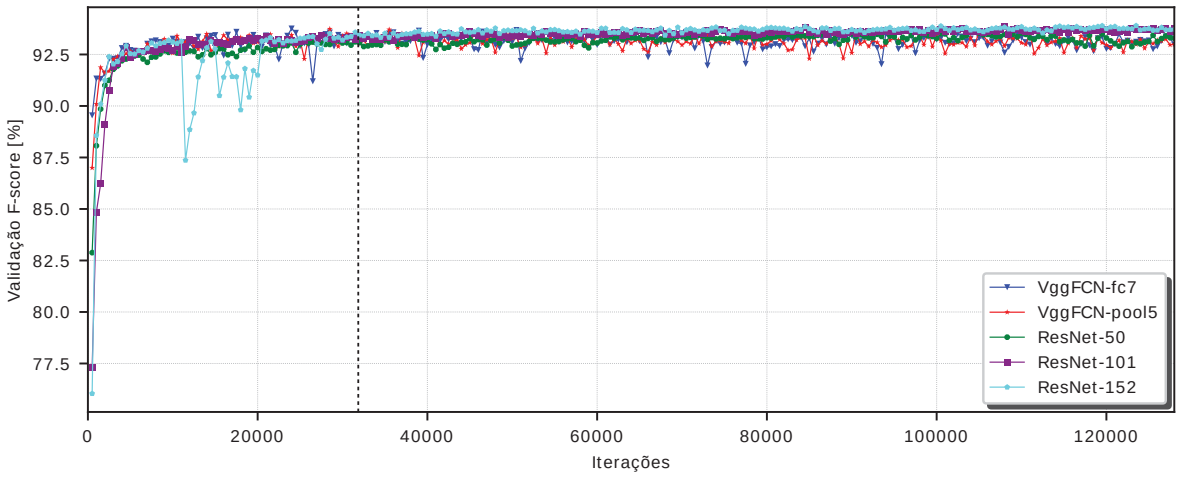


Figura 4.4: Comportamento da convergência da curva de aprendizagem com $128t$ iterações das 5 abordagens propostas.

($\alpha = 0{,}05$), aceitando a hipótese nula (quando não há diferença estatística) caso $p\text{-}value > \alpha$, e rejeitando essa hipótese (quando existe uma diferença), caso $p\text{-}value < \alpha$ (Demšar, 2006; Flach, 2012).

Para comparar k algoritmos de classificação, usando N conjuntos de dados, é necessário utilizar testes de significância específicos para esse cenário (múltiplos *datasets* e algoritmos). O método de Friedman (1937) é um método estatístico não paramétrico que classifica os k algoritmos de acordo com cada *dataset* N separadamente, gerando um *rank* conforme o desempenho desse algoritmo. O primeiro melhor tem *rank* 1, o segundo melhor *rank* 2, até *rank* k. Em caso de empate, um *rank* médio é atribuído (Demšar, 2006).

Seja r_i^j o *rank* j-ésimo do k algoritmo no i-ésimo N *dataset*. Em seguida, um *rank* médio é gerado, $R_j = \frac{1}{N}\sum_i r_i^j$, com base em todos os *ranks* de cada *dataset*. O teste de Friedman é dado conforme a Equação 4.7:

$$\chi_F^2 = \frac{12N}{k(k+1)}\left[\sum_j R_j^2 - \frac{k(k+1)^2}{4}\right] \tag{4.7}$$

onde χ_F^2 representa a estatística de Friedman, de acordo com $k-1$ graus de liberdade, N é a quantidade de *datasets*, k é a quantidade de algoritmos e R_j é o *rank* médio desses algoritmos (Demšar, 2006).

Como o teste de Friedman mostra os *ranks* médios de acordo com a classificação de cada algoritmo, é possível apenas identificar se existe ou não diferença estatística entre esses algoritmos, conforme rejeição/aceitação de H_0. Portanto, apenas com essa informação não é possível determinar, de fato, qual desses algoritmos é melhor/pior que o outro. Em casos onde H_0 é rejeitada, aplica-se o pós-teste de Nemenyi (1963), que serve para identificar o quão diferente, por meio da *critical difference* (CD), um algoritmo é do outro conforme:

$$CD = q_\alpha\sqrt{\frac{k(k+1)}{6N}} \tag{4.8}$$

onde q_α representa os valores críticos que se baseiam na estatística de intervalo *studentized* dividida por $\sqrt{2}$, k é a quantidade de algoritmos e N é a quantidade de *datasets* (Demšar, 2006).

A CD serve como um limiar para comparação par-a-par entre os algoritmos, agrupando os que estão dentro da faixa de CD e separando os que estão fora dessa faixa. Esse agrupamento é ilustrado por meio de um diagrama de diferença crítica, ilustrados na Seção 5.4. De acordo com Demšar (2006), as métricas utilizadas para a realização dessa avaliação não afetam o desempenho do teste estatístico. Portanto, o F1 foi escolhido por que leva em consideração tanto o erro positivo quanto negativo, por meio da média harmônica entre *prec* e *rec*.

4.7 Considerações Finais

Neste capítulo, protocolos e métricas para avaliação dos resultados das abordagens propostas (5 modelos de rede: VggFCN-$fc7$, VggFCN-*pool*5, ResNet-50, ResNet-101 e ResNet-152) foram definidos, bem como os *datasets* e as divisões dos mesmos, utilizados durante a execução dos experimentos foram descritos e ilustrados. Além disso, apresentamos os *baselines* de comparação e justificamos a escolha da quantidade de iterações para o treinamento dessa da abordagem proposta. Por fim, os métodos de Friedman e Nemenyi foram definidos, para avaliação estatística dos resultados entre os *baselines* e as abordagens propostas.

5 Resultados e Discussões

Neste capítulo apresentam-se os resultados obtidos durante a realização de todos os experimentos, conduzidos conforme a metodologia experimental discutida anteriormente no Capítulo 4. Apresentam-se os resultados obtidos para as métricas de Prec, Rec, F1, IoU e erro E por cada um dos 4 *frameworks* (ver Seção 4.4) usados como *baseline* (OSIRISv4.1, IrisSeg, Haindl e TVM) e dos 5 modelos de DCNN definidos como abordagem proposta (VggFCN-$fc7$, VggFCN-*pool*5, ResNet-50, ResNet-101 e ResNet-152). Esses resultados, para todas as métricas, são apresentados em tabelas, enquanto os resultados de F1 e E serão ilustrados em gráficos de barras. Por fim, com base em análise de significância estatística, define-se qual dos modelos das abordagens propostas obtiveram melhores resultados.

5.1 Resultados com o Protocolo da Competição *Noisy Iris Challenge Evaluation - Part I*

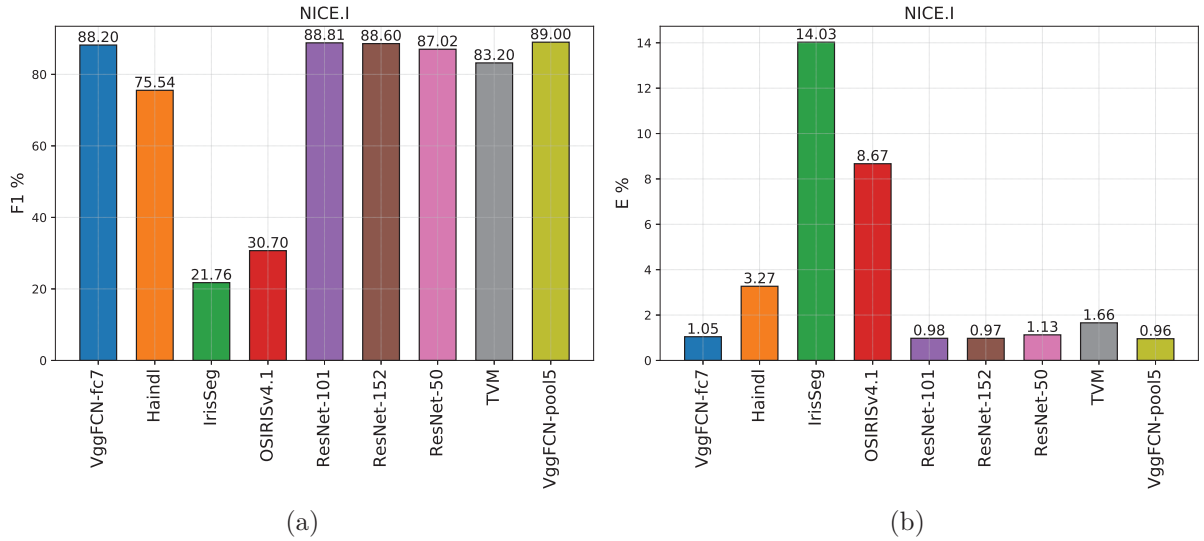
Conforme descrito na Seção 4.3.1, experimentos iniciais foram realizados conforme o protocolo da competição NICE.I, seguindo a mesma divisão deste *dataset*. Como pode ser observado na Tabela 5.1, tanto o OSIRISv4.1 quanto o IrisSeg apresentaram os menores resultados de F1, (por exemplo $30{,}70 \pm 32{,}00$ e $21{,}76 \pm 32{,}13$, respectivamente) para todas as métricas, exceto o E que foi os maiores, em comparação a todos os outros métodos.

Tabela 5.1: Resultados de segmentação da íris seguindo o protocolo da competição NICE.I.

Método	**Precision %**	**Recall %**	**F1 %**	**IoU %**	**E %**
OSIRISv4.1	$30{,}30 \pm 31{,}59$	$35{,}81 \pm 35{,}67$	$30{,}70 \pm 32{,}00$	$23{,}13 \pm 26{,}70$	$08{,}67 \pm 06{,}29$
IrisSeg	$20{,}67 \pm 31{,}17$	$25{,}88 \pm 34{,}93$	$21{,}76 \pm 32{,}13$	$17{,}15 \pm 27{,}67$	$14{,}03 \pm 12{,}33$
Haindl	$74{,}75 \pm 23{,}68$	$79{,}96 \pm 21{,}95$	$75{,}54 \pm 22{,}93$	$65{,}15 \pm 24{,}50$	$03{,}27 \pm 04{,}29$
TVM	$82{,}51 \pm 18{,}06$	$85{,}41 \pm 18{,}20$	$83{,}20 \pm 17{,}32$	$73{,}77 \pm 17{,}38$	$01{,}66 \pm 01{,}42$
VggFCN-$fc7$	$88{,}55 \pm 12{,}68$	$88{,}80 \pm 15{,}67$	$88{,}20 \pm 13{,}73$	$80{,}56 \pm 13{,}89$	$01{,}05 \pm 00{,}86$
VggFCN-*pool*5	$88{,}76 \pm 13{,}45$	$89{,}74 \pm 13{,}88$	$89{,}00 \pm 13{,}11$	$81{,}71 \pm 13{,}21$	$00{,}96 \pm 00{,}60$
ResNet-50	$87{,}62 \pm 14{,}78$	$87{,}28 \pm 15{,}08$	$87{,}02 \pm 14{,}39$	$78{,}85 \pm 14{,}63$	$01{,}13 \pm 00{,}94$
ResNet-101	$88{,}16 \pm 12{,}84$	$89{,}87 \pm 13{,}61$	$88{,}81 \pm 12{,}66$	$81{,}28 \pm 12{,}60$	$00{,}98 \pm 00{,}63$
ResNet-152	$88{,}14 \pm 13{,}49$	$89{,}45 \pm 13{,}74$	$88{,}60 \pm 13{,}05$	$81{,}03 \pm 12{,}97$	$00{,}97 \pm 00{,}54$

Esse menor valor de F1 pode ser justificado porque ambos foram desenvolvidos para imagens NIR. O Haindl e o TVM obtiveram resultados de F1 de $75{,}54 \pm 22{,}93$ e $83{,}20 \pm 17{,}32$, respectivamente. Acreditava-se que os resultados do Haindl seriam maiores que os demais *baselines*, pois ele foi desenvolvido especificamente para imagens VIS. Porém, o que alcançou maiores resultados entre os *baselines* foi o TVM.

Por outro lado dentre os modelos da abordagem proposta, o ResNet-50 foi o que obteve o menor valor para F1 ($87{,}02 \pm 14{,}39$) e maior para E= $01{,}13 \pm 00{,}94$. Já o modelo VggFCN-*pool*5 (linha verde da Tabela 5.1) obteve os melhores resultados em média para todas as métricas, com F1 e E de respectivamente $89{,}00 \pm 13{,}11$ e $00{,}96 \pm 00{,}60$. A Figura 5.1 ilustra os resultados de segmentação dos *baselines* e da abordagem proposta no *dataset* NICE.I. É possível observar o maior valor do modelo VggFCN-*pool*5 (barra em amarelo) para F1 e o menor valor para E.



(a) (b)

Figura 5.1: Resultados dos métodos no *dataset* NICE.I. (a) representa a métrica F1, enquanto (b) representa o E.

Também pode-se observar na Figura 5.1 valores similares para ResNet-50, ResNet-101 e ResNet-152, para ambos F1 e E, pois eles possuem uma configuração de arquitetura quase idêntica, apenas com variações na quantidade de camadas de convoluções (ver Capítulo 3 para mais detalhes).

5.2 Resultados em cada *Dataset*

Os experimentos realizados em cada um dos 4 *datasets* NIR serão apresentados na próxima seção. O Haindl não foi capaz de segmentar nenhuma imagem NIR, pois não foi desenvolvido para este domínio (Haindl e Krupicka, 2015).

5.2.1 Resultados nos *Datasets Near Infra-red*

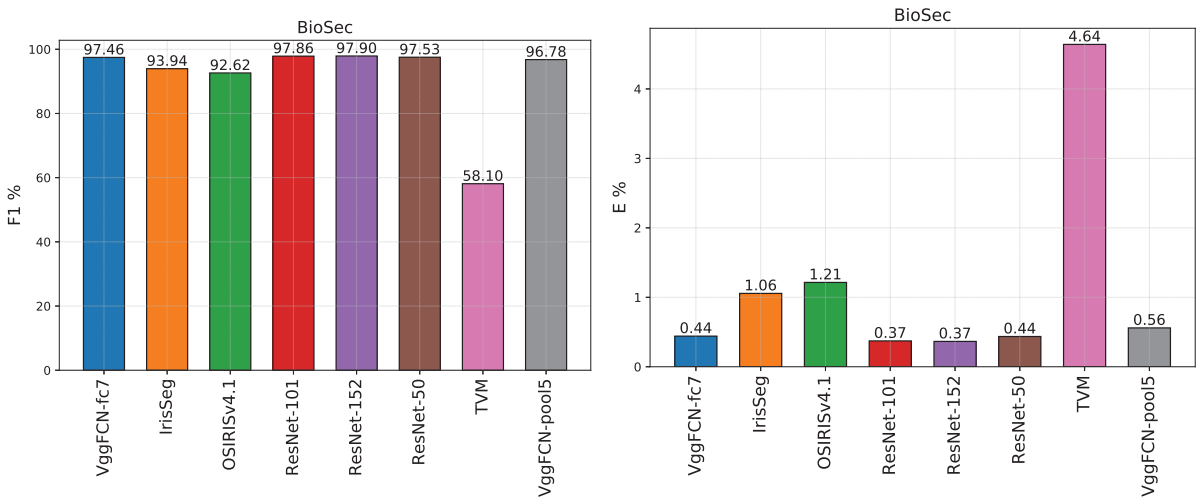
Os resultados de segmentação de íris no *dataset* BioSec são apresentados conforme a Tabela 5.2, onde é possível observar que o pior valor em média de F1 e E foi obtido por TVM ($58{,}10 \pm 38{,}40$ e $04{,}64 \pm 03{,}22$ respectivamente), mesmo este *dataset* não sendo tão desafiador, como os demais. Já o OSIRISv4.1 obteve $92{,}62 \pm 03{,}19$ e $01{,}21 \pm 00{,}47$, pouco pior em média do que o IrisSeg, que obteve $93{,}94 \pm 05{,}88$ e $01{,}06 \pm 01{,}20$ para F1 e E respectivamente.

O VggFCN-*pool*5 obteve o menor valor de F1 ($96{,}78 \pm 00{,}84$) em comparação com todas as abordagens propostas, que oscilaram entre $97{,}46 \pm 00{,}74$ (VggFCN-$fc7$) e $97{,}90 \pm 00{,}86$ (ResNet-152). A Figura 5.2 ilustra essa pequena oscilação do F1, principalmente

Tabela 5.2: Resultados de segmentação da íris no *dataset* **BioSec** pelos 3 *baselines* e pelas 5 abordagens propostas.

Método	Precision %	Recall %	F1 %	IoU %	E %
OSIRISv4.1	$96{,}29 \pm 03{,}12$	$89{,}42 \pm 05{,}22$	$92{,}62 \pm 03{,}19$	$86{,}42 \pm 05{,}46$	$01{,}21 \pm 00{,}47$
IrisSeg	$96{,}90 \pm 08{,}33$	$91{,}65 \pm 04{,}54$	$93{,}94 \pm 05{,}88$	$89{,}04 \pm 08{,}62$	$01{,}06 \pm 01{,}20$
Haindl	– – –	– – –	– – –	– – –	– – –
TVM	$70{,}55 \pm 38{,}30$	$54{,}09 \pm 36{,}73$	$58{,}10 \pm 38{,}40$	$50{,}17 \pm 34{,}25$	$04{,}64 \pm 03{,}22$
VggFCN-$fc7$	$97{,}18 \pm 01{,}54$	$97{,}77 \pm 01{,}21$	$97{,}46 \pm 00{,}74$	$95{,}05 \pm 01{,}40$	$00{,}44 \pm 00{,}12$
VggFCN-*pool5*	$96{,}46 \pm 01{,}54$	$97{,}13 \pm 01{,}29$	$96{,}78 \pm 00{,}84$	$93{,}77 \pm 01{,}57$	$00{,}56 \pm 00{,}14$
ResNet-50	$97{,}18 \pm 03{,}34$	$97{,}96 \pm 01{,}29$	$97{,}53 \pm 01{,}85$	$95{,}23 \pm 03{,}17$	$00{,}44 \pm 00{,}37$
ResNet-101	$97{,}43 \pm 01{,}67$	$98{,}32 \pm 01{,}10$	$97{,}86 \pm 00{,}77$	$95{,}82 \pm 01{,}46$	$00{,}37 \pm 00{,}13$
ResNet-152	$97{,}54 \pm 01{,}80$	$98{,}31 \pm 01{,}14$	$97{,}90 \pm 00{,}86$	$95{,}91 \pm 01{,}61$	$00{,}37 \pm 00{,}15$

pelos modelos ResNet-50, ResNet-101 e ResNet-152. Na Figura 5.2b, o valor de E também



(a) (b)

Figura 5.2: Resultados dos métodos no *dataset* BioSec. (a) representa a métrica F1, enquanto (b) representa o E.

obteve pouca variação para as abordagens propostas, ficando entre 00,37% e 00,56%.

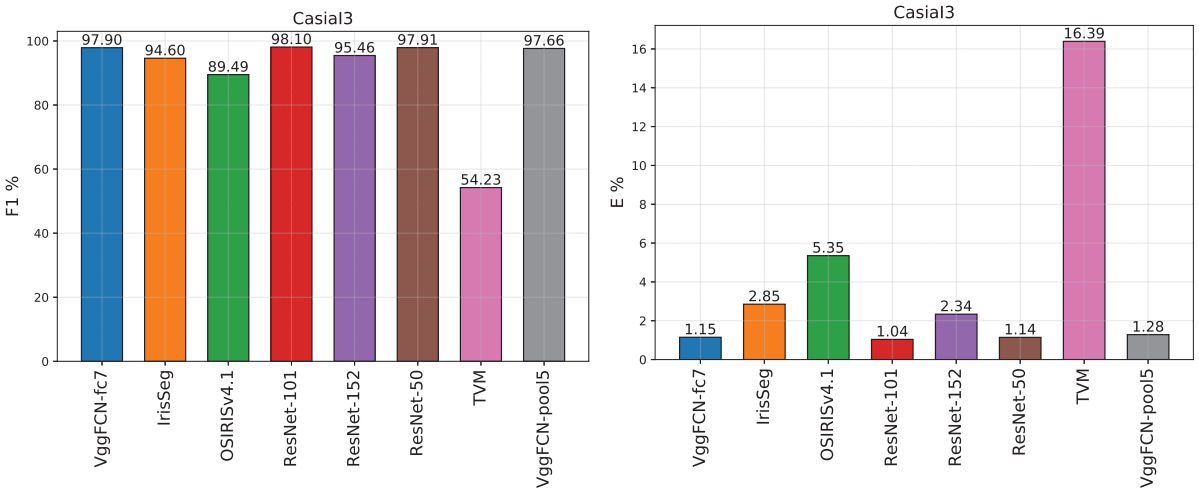
No *dataset* CasiaI3 o ResNet-101 apresentou (Tabela 5.3), maior valor de F1 ($98{,}10 \pm 00{,}59\%$) enquanto o ResNet-152 obteve o menor valor ($95{,}46 \pm 01{,}95\%$) para as abordagens propostas.

OSIRISv4.1 e IrisSeg obtiveram $89{,}49 \pm 05{,}78\%$ e $94{,}61 \pm 03{,}28\%$ de F1 respectivamente, e o Haindl $54{,}23 \pm 33{,}12\%$ obteve o menor resultado entre todos. Observe na Figura 5.3 o E mais alto de 16,39% para o TVM e 05,35 para o OSIRISv4.1, enquanto o ResNet-101 alcançou apenas 01,04%.

No *dataset* CasiaT4 os valores de F1 das abordagens propostas ficaram todos na faixa entre $94{,}42 \pm 07{,}54\%$ (VggFCN-$fc7$) e $96{,}15 \pm 04{,}57\%$ (ResNet-50), conforme representa a Tabela 5.4. Essa pouca variação $\approx 1{,}73\%$ entre os resultados mostra um comportamento similar entre as abordagens propostas, de forma geral, neste *dataset*. Entretanto, uma maior variação ($\approx 12{,}51\%$) pode ser observada entre os *baselines*, com novamente o TVM obtendo o menor valor de F1.

Tabela 5.3: Resultados de segmentação da íris no *dataset* **CasiaI3** pelos 3 *baselines* e pelas 5 abordagens propostas.

Método	Precision %	Recall %	F1 %	IoU %	E %
OSIRISv4.1	93,17 ± 04,94	86,57 ± 08,39	89,49 ± 05,78	81,44 ± 08,85	05,35 ± 02,40
IrisSeg	97,03 ± 03,38	92,54 ± 05,05	94,61 ± 03,28	89,93 ± 05,28	02,85 ± 01,62
Haindl	– – –	– – –	– – –	– – –	– – –
TVM	87,22 ± 20,14	46,75 ± 32,55	54,23 ± 33,12	43,99 ± 30,13	16,39 ± 09,47
VggFCN-$fc7$	97,71 ± 01,08	98,10 ± 01,17	97,90 ± 00,68	95,89 ± 01,29	01,15 ± 00,37
VggFCN-*pool*5	97,40 ± 01,12	97,93 ± 01,14	97,66 ± 00,67	95,43 ± 01,26	01,28 ± 00,37
ResNet-50	97,84 ± 01,27	98,00 ± 01,28	97,91 ± 00,73	95,91 ± 01,38	01,14 ± 00,38
ResNet-101	98,01 ± 01,16	98,21 ± 01,10	98,10 ± 00,59	96,28 ± 01,13	01,04 ± 00,31
ResNet-152	97,40 ± 01,17	93,89 ± 07,21	95,46 ± 01,95	91,57 ± 06,90	02,34 ± 01,88



(a) (b)

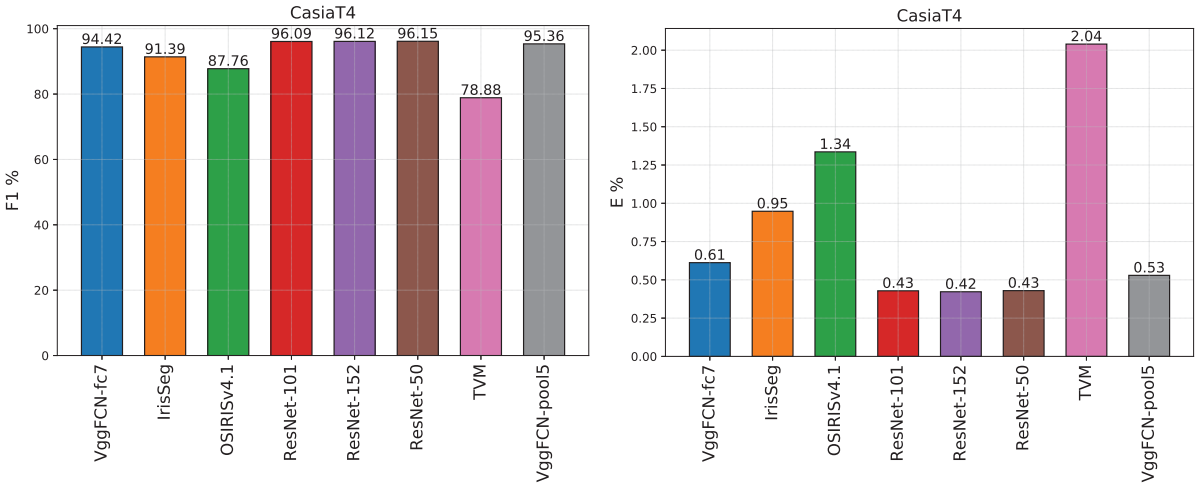
Figura 5.3: Resultados dos métodos no *dataset* CasiaI3. (a) representa a métrica F1, enquanto (b) representa o E.

Tabela 5.4: Resultados de segmentação da íris no *dataset* **CasiaT4** pelos 3 *baselines* e pelas 5 abordagens propostas.

Método	Precision %	Recall %	F1 %	IoU %	E %
OSIRISv4.1	88,15 ± 08,64	88,88 ± 11,38	87,76 ± 08,01	78,97 ± 10,98	01,34 ± 00,64
IrisSeg	90,70 ± 09,31	92,55 ± 08,55	91,39 ± 08,13	84,83 ± 09,58	00,95 ± 00,54
Haindl	– – –	– – –	– – –	– – –	– – –
TVM	92,14 ± 14,67	71,01 ± 19,15	78,88 ± 17,69	67,75 ± 18,32	02,04 ± 01,48
VggFCN-$fc7$	93,93 ± 08,26	95,11 ± 07,01	94,42 ± 07,54	90,00 ± 08,08	00,61 ± 00,58
VggFCN-*pool*5	94,88 ± 02,40	95,88 ± 01,86	95,36 ± 01,62	91,17 ± 02,89	00,53 ± 00,17
ResNet-50	96,21 ± 02,22	96,40 ± 05,74	96,15 ± 04,57	92,83 ± 05,82	00,43 ± 00,41
ResNet-101	96,34 ± 01,98	96,32 ± 06,89	96,09 ± 06,44	92,89 ± 06,86	00,43 ± 00,50
ResNet-152	96,26 ± 02,34	96,50 ± 06,97	96,12 ± 06,76	92,88 ± 07,07	00,42 ± 00,51

Outra observação interessante é vista na Figura 5.4b, onde os ResNet-50, ResNet-101 e ResNet-152 alcançaram praticamente o mesmo valor de E (0,43%, 0,43% e 0,42%

respectivaente), enquanto VggFCN-$fc7$ e VggFCN-*pool*5 obtiveram E um pouco maior (respectivamente 0,61% e 0,53%).



(a) (b)

Figura 5.4: Resultados dos métodos no *dataset* CasiaT4. (a) representa a métrica F1, enquanto (b) representa o E.

Por fim, os resultados dos *baselines* e das abordagens propostas no ultimo *dataset* (IITD-1), com imagens NIR avaliado individualmente, são apresentados na Tabela 5.5. Observe que todas as 5 abordagens propostas obtiveram F1 de $\approx 97\%$, variando apenas 0,62%, enquanto os resultados dos *baselines* tiveram maior variação (56,49%) entre os 3.

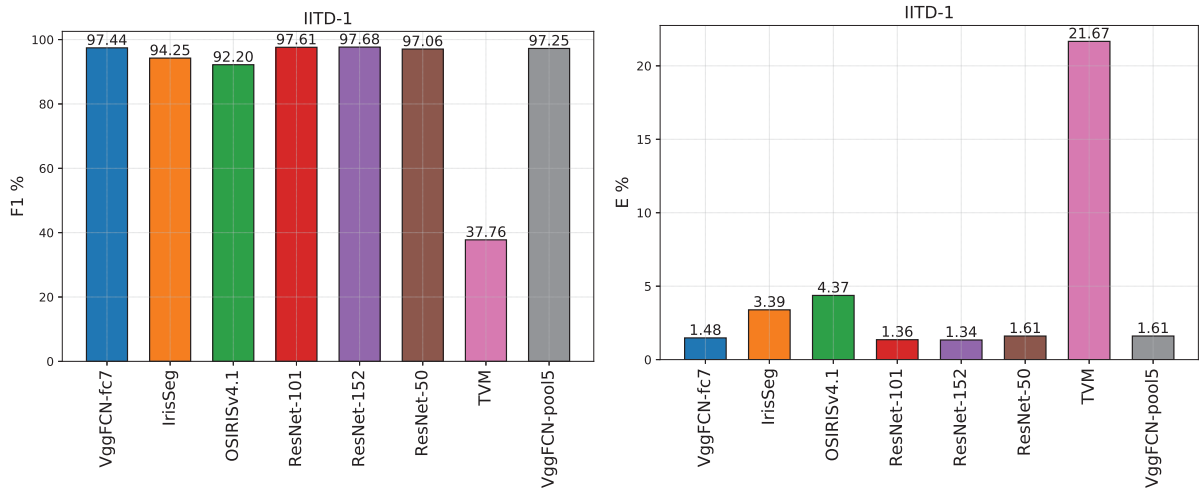
Tabela 5.5: Resultados de segmentação da íris no *dataset* **IITD-1** pelos 3 *baselines* e pelas 5 abordagens propostas.

Método	**Precision %**	**Recall %**	**F1 %**	**IoU %**	**E %**
OSIRISv4.1	$91{,}93 \pm 06{,}51$	$93{,}09 \pm 07{,}97$	$92{,}20 \pm 06{,}07$	$86{,}03 \pm 08{,}93$	$04{,}37 \pm 02{,}69$
IrisSeg	$93{,}25 \pm 06{,}06$	$95{,}61 \pm 04{,}02$	$94{,}25 \pm 03{,}89$	$89{,}36 \pm 06{,}34$	$03{,}39 \pm 02{,}16$
Haindl	– – –	– – –	– – –	– – –	– – –
TVM	$86{,}47 \pm 23{,}83$	$32{,}49 \pm 34{,}76$	$37{,}76 \pm 36{,}08$	$30{,}35 \pm 32{,}07$	$21{,}67 \pm 11{,}08$
VggFCN-$fc7$	$97{,}47 \pm 01{,}58$	$97{,}48 \pm 02{,}91$	$97{,}44 \pm 01{,}78$	$95{,}06 \pm 02{,}99$	$01{,}48 \pm 01{,}01$
VggFCN-*pool*5	$96{,}91 \pm 01{,}79$	$97{,}63 \pm 01{,}72$	$97{,}25 \pm 01{,}23$	$94{,}67 \pm 02{,}22$	$01{,}61 \pm 00{,}75$
ResNet-50	$97{,}40 \pm 01{,}48$	$96{,}96 \pm 05{,}46$	$97{,}06 \pm 03{,}51$	$94{,}47 \pm 05{,}22$	$01{,}61 \pm 01{,}05$
ResNet-101	$97{,}67 \pm 01{,}42$	$97{,}61 \pm 02{,}50$	$97{,}61 \pm 01{,}45$	$95{,}37 \pm 02{,}59$	$01{,}36 \pm 00{,}56$
ResNet-152	$97{,}66 \pm 01{,}35$	$97{,}75 \pm 02{,}51$	$97{,}68 \pm 01{,}40$	$95{,}49 \pm 02{,}46$	$01{,}34 \pm 00{,}56$

Para as abordagens propostas, de forma semelhante os resultados de E foram de $\approx 1\%$ com variação de 0,27% , conforme ilustra a Figura 5.5b.

5.2.2 Resultados nos *Datasets Visible*

Nos *datasets* sVIS todos os 4 *baselines* funcionaram sem problemas, ao contrário do *dataset* NIR, onde o método Haindl não funcionou. O primeiro *dataset* VIS analisado foi o NICE.I*, dividido de forma diferente (conforme descrito na Seção 4.3.1) da distribuição



(a) (b)

Figura 5.5: Resultados dos métodos no *dataset* IITD-1. (a) representa a métrica F1, enquanto (b) representa o E.

Tabela 5.6: Resultados de segmentação da íris no *dataset* **NICE.I*** pelos 4 *baselines* e pelas 5 abordagens propostas.

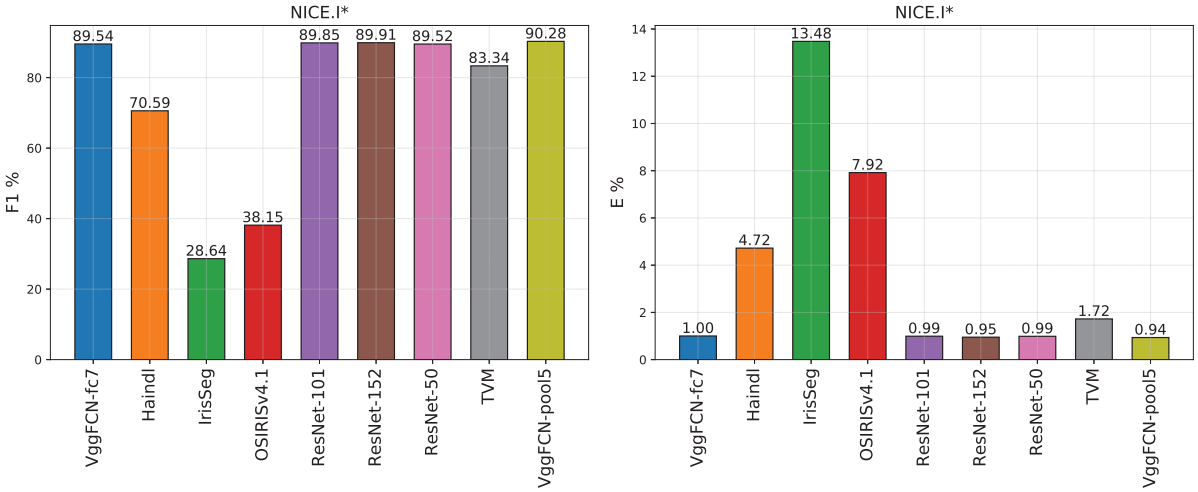
Método	**Precision %**	**Recall %**	**F1 %**	**IoU %**	**E %**
OSIRISv4.1	$38{,}70 \pm 33{,}55$	$42{,}13 \pm 36{,}30$	$38{,}15 \pm 33{,}61$	$29{,}52 \pm 28{,}93$	$07{,}92 \pm 06{,}20$
IrisSeg	$27{,}88 \pm 35{,}19$	$31{,}66 \pm 36{,}58$	$28{,}64 \pm 35{,}14$	$22{,}88 \pm 30{,}29$	$13{,}48 \pm 12{,}36$
Haindl	$69{,}05 \pm 28{,}55$	$77{,}12 \pm 22{,}13$	$70{,}59 \pm 26{,}11$	$60{,}03 \pm 27{,}50$	$04{,}72 \pm 05{,}87$
TVM	$83{,}43 \pm 19{,}27$	$84{,}64 \pm 19{,}39$	$83{,}34 \pm 18{,}63$	$74{,}41 \pm 18{,}96$	$01{,}72 \pm 01{,}38$
VggFCN-$fc7$	$88{,}90 \pm 14{,}64$	$90{,}61 \pm 14{,}08$	$89{,}54 \pm 13{,}79$	$82{,}75 \pm 13{,}85$	$01{,}00 \pm 00{,}70$
VggFCN-*pool*5	$89{,}87 \pm 14{,}17$	$91{,}06 \pm 14{,}37$	$90{,}28 \pm 13{,}74$	$83{,}96 \pm 13{,}64$	$00{,}94 \pm 00{,}72$
ResNet-50	$88{,}89 \pm 14{,}36$	$90{,}48 \pm 14{,}00$	$89{,}52 \pm 13{,}74$	$82{,}71 \pm 13{,}71$	$00{,}99 \pm 00{,}61$
ResNet-101	$88{,}87 \pm 14{,}26$	$91{,}21 \pm 14{,}26$	$89{,}85 \pm 13{,}76$	$83{,}26 \pm 13{,}74$	$00{,}99 \pm 00{,}91$
ResNet-152	$89{,}08 \pm 14{,}38$	$91{,}04 \pm 13{,}84$	$89{,}91 \pm 13{,}71$	$83{,}33 \pm 13{,}61$	$00{,}95 \pm 00{,}59$

original da competição NICE.I. A Tabela 5.6 representa os resultados de todos os *baselines* e de todas as abordagens propostas no *dataset* NICE.I*.

Observe que ambos OSIRISv4.1 e IrisSeg obtiveram os menores valores ($38{,}15 \pm 33{,}61\%$ e $28{,}64 \pm 35{,}14\%$, respectivamente) de F1 entre os 4 *baselines*. Isso já era esperado pois foram desenvolvidos para imagens NIR. Já Haindl e TVM obtiveram os maiores resultados entre os 4, sendo respectivamente $70{,}59 \pm 26{,}11\%$ e $83{,}34 \pm 18{,}63\%$ de F1.

Entretanto, semelhante aos resultados nos *datasets* NIR, todas as abordagens propostas obtiveram os maiores valores. Note que houve uma pequena variação, entre esses resultados, de apenas 0,76%, mostrando um comportamento muito similar entre as abordagens. Analisando o E, na Figura 5.6, a pequena variação de apenas 0,06% foi obtida pelas abordagens, sendo que a que alcançou maior F1 e menor E foi a VggFCN-*pool*5, ($90{,}28 \pm 13{,}74\%$ e $00{,}94 \pm 00{,}72\%$) respectivamente.

Os resultados no *dataset* CrEye-Iris são apresentados conforme a Tabela 5.7, onde é possível observar o valor de F1 mais alto para o método VggFCN-$fc7$ de $97{,}04 \pm 01{,}21\%$, destacado pela linha verde da tabela. Novamente a variação entre esses valores foi pequena (apenas 0,66%).



(a) (b)

Figura 5.6: Resultados dos métodos no *dataset* NICE.I. (a) representa a métrica F1, enquanto (b) representa o E.

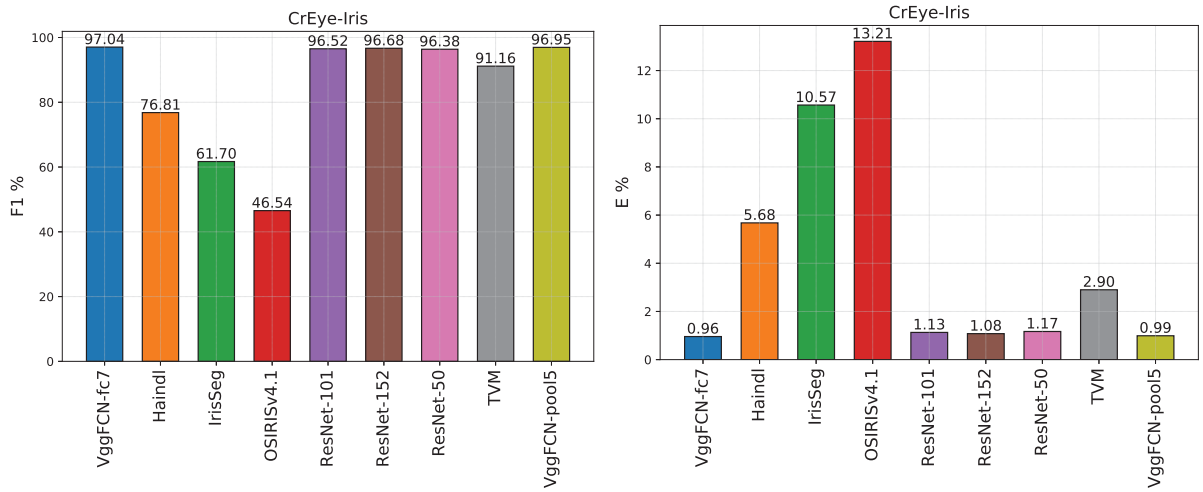
Tabela 5.7: Resultados de segmentação da íris no *dataset* **CrEye-Iris** pelos 4 *baselines* e pelas 5 abordagens propostas.

Método	**Precision %**	**Recall %**	**F1 %**	**IoU %**	**E %**
OSIRISv4.1	$57{,}86 \pm 30{,}41$	$41{,}55 \pm 29{,}53$	$46{,}53 \pm 29{,}25$	$35{,}29 \pm 26{,}42$	$13{,}22 \pm 06{,}33$
IrisSeg	$72{,}94 \pm 36{,}90$	$56{,}67 \pm 31{,}80$	$61{,}72 \pm 33{,}55$	$52{,}03 \pm 30{,}75$	$10{,}58 \pm 10{,}38$
Haindl	$94{,}48 \pm 09{,}29$	$69{,}59 \pm 25{,}98$	$76{,}81 \pm 23{,}73$	$67{,}14 \pm 25{,}09$	$05{,}69 \pm 04{,}58$
TVM	$91{,}10 \pm 06{,}83$	$91{,}81 \pm 04{,}71$	$91{,}16 \pm 03{,}31$	$83{,}92 \pm 05{,}33$	$02{,}90 \pm 00{,}98$
VggFCN-$fc7$	$96{,}85 \pm 01{,}59$	$97{,}26 \pm 01{,}77$	$97{,}04 \pm 01{,}21$	$94{,}28 \pm 02{,}25$	$00{,}96 \pm 00{,}36$
VggFCN-*pool5*	$96{,}55 \pm 02{,}21$	$97{,}39 \pm 01{,}14$	$96{,}95 \pm 01{,}27$	$94{,}12 \pm 02{,}35$	$00{,}99 \pm 00{,}38$
ResNet-50	$96{,}19 \pm 01{,}97$	$96{,}62 \pm 02{,}46$	$96{,}38 \pm 01{,}56$	$93{,}05 \pm 02{,}82$	$01{,}17 \pm 00{,}43$
ResNet-101	$96{,}45 \pm 01{,}78$	$96{,}62 \pm 01{,}81$	$96{,}52 \pm 01{,}26$	$93{,}30 \pm 02{,}30$	$01{,}13 \pm 00{,}38$
ResNet-152	$96{,}47 \pm 01{,}83$	$96{,}93 \pm 01{,}76$	$96{,}68 \pm 01{,}17$	$93{,}60 \pm 02{,}17$	$01{,}08 \pm 00{,}34$

Entre os *baselines*, o TVM obteve maior valor de F1 ($91{,}16 \pm 03{,}31\%$), enquanto o OSIRISv4.1 obteve o resultado mais baixo ($46{,}53 \pm 29{,}25\%$). Os valores de E são ilustrados na Figura 5.7b, na qual a ResNet-50, ResNet-101 e ResNet-152 alcançaram respectivamente 1,17%, 1,13% e 1,08%, enquanto VggFCN-$fc7$ obteve o menor entre todos (0,96%).

O *dataset* mais desafiador dentre os apresentados até agora foi o MICHE-I pois os resultados de F1 obtidos pelas abordagens propostas oscilaram entre 78,28% (ResNet-101) até $\approx 82\%$ (VggFCN-$fc7$), conforme apresentado na Tabela 5.8. Entre os *baselines*, o Haindl e o IrisSeg obtiveram resultados de $63{,}12 \pm 33{,}30\%$ e $19{,}34 \pm 33{,}03\%$. Note que o desvio-padrão da distribuição das imagens apresentou, para F1, valores entre $\pm 33{,}03\%$ e $\pm 40{,}10\%$ entre esses *baselines*, enquanto nas abordagens propostas esse desvio oscilou entre $\pm 20{,}28\%$ até $\pm 27{,}97\%$. Esse alto desvio padrão é justificado pelo fato que a quantidade de FPs e FNs variam muito para cada imagem, contribuindo para o aumento desses desvios.

Além disso, como as imagens deste *dataset* possuem alta resolução, conforme apresentado na Tabela 4.1, a quantidade de memória da GPU utilizada no treinamento das abordagens propostas não foi suficiente, sendo necessário o redimensionamento das imagens.



(a) (b)

Figura 5.7: Resultados dos métodos no *dataset* CrEye-Iris. (a) representa a métrica F1, enquanto (b) representa o E.

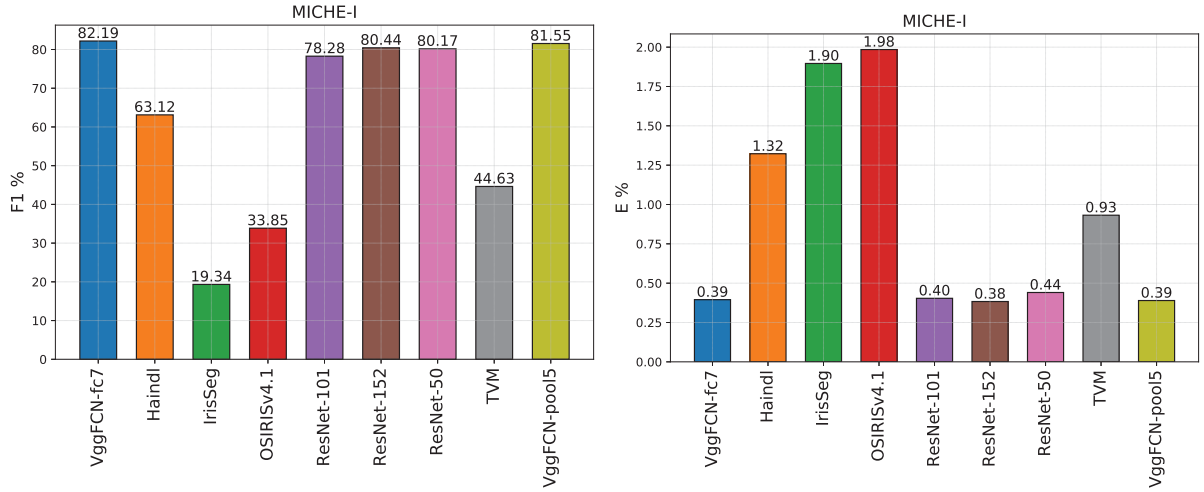
Tabela 5.8: Resultados de segmentação da íris no *dataset* **MICHE-I** pelos 4 *baselines* e pelas 5 abordagens propostas.

Método	**Precision %**	**Recall %**	**F1 %**	**IoU %**	**E %**
OSIRISv4.1	$39{,}24 \pm 37{,}55$	$35{,}81 \pm 37{,}68$	$33{,}85 \pm 35{,}86$	$26{,}90 \pm 30{,}49$	$01{,}99 \pm 02{,}90$
IrisSeg	$20{,}00 \pm 34{,}36$	$20{,}14 \pm 34{,}02$	$19{,}34 \pm 33{,}03$	$15{,}82 \pm 28{,}18$	$01{,}90 \pm 03{,}36$
Haindl	$61{,}81 \pm 34{,}79$	$69{,}46 \pm 31{,}74$	$63{,}12 \pm 33{,}30$	$53{,}71 \pm 31{,}56$	$01{,}32 \pm 02{,}10$
TVM	$45{,}94 \pm 40{,}68$	$46{,}61 \pm 42{,}60$	$44{,}63 \pm 40{,}10$	$37{,}72 \pm 35{,}04$	$00{,}93 \pm 00{,}77$
VggFCN-$fc7$	$86{,}66 \pm 16{,}80$	$81{,}42 \pm 22{,}75$	$82{,}19 \pm 20{,}28$	$73{,}44 \pm 21{,}70$	$00{,}39 \pm 00{,}45$
VggFCN-*pool5*	$87{,}01 \pm 13{,}06$	$81{,}32 \pm 24{,}50$	$81{,}55 \pm 20{,}66$	$72{,}65 \pm 22{,}23$	$00{,}39 \pm 00{,}44$
ResNet-50	$85{,}51 \pm 18{,}51$	$79{,}46 \pm 26{,}09$	$80{,}17 \pm 23{,}52$	$71{,}74 \pm 25{,}02$	$00{,}44 \pm 00{,}55$
ResNet-101	$86{,}46 \pm 23{,}15$	$76{,}16 \pm 30{,}02$	$78{,}28 \pm 27{,}97$	$70{,}66 \pm 28{,}00$	$00{,}40 \pm 00{,}47$
ResNet-152	$87{,}55 \pm 19{,}30$	$79{,}00 \pm 28{,}20$	$80{,}44 \pm 25{,}63$	$72{,}82 \pm 26{,}14$	$00{,}38 \pm 00{,}47$

Esse redimensionamento, com base na resolução média do *subset* de treinamento, resultou na resolução 1191×1672, (largura e altura) respectivamente. Porém, o redimensionamento, para menor resolução, ocasiona perda de informações pelo agrupamento dos *pixels*, de acordo com a interpolação cúbica. Outro problema quanto ao redimensionamento, é a distorção causada na imagem, principalmente quando a proporção entre a largura e altura da imagem mudam. Portanto, esses resultados de F1 mais baixos podem ser justificados por esse redimensionamento.

A Figura 5.8 ilustra os resultados de F1 e E obtidos por todos os *baselines* e todas as abordagens propostas. Observe que o E obtido pelas abordagens propostas ficaram entre 0,38% (ResNet-152) e 0,44% (ResNet-50), sendo o menor entre todos os experimentos realizados, mesmo apesar desse *dataset* ser o mais desafiador.

Isso ocorreu pois a quantidade de *pixels* presente na MICHE-I é muito maior do que os demais *datasets*, devido sua alta resolução espacial, mesmo após o redimensionamento (ver Seção 4.3, e Equação 4.1 para maiores informações). Os experimentos realizados nos *datasets* mesclados serão apresentados na próxima seção.



(a) (b)

Figura 5.8: Resultados dos métodos no *dataset* MICHE-I. (a) representa a métrica F1, enquanto (b) representa o E.

5.3 Resultados nos *Datasets* Mesclados

A mesclagem dos *datasets* NIR_{dt}, VIS_{dt} e Todos_{dt} possibilitou uma análise da capacidade de aprendizagem/generalização das abordagens propostas, com relação a maior variabilidade das imagens. Isso ocorre pois as imagens são do mesmo domínio (NIR e/ou VIS) e provenientes de *datasets* distintos.

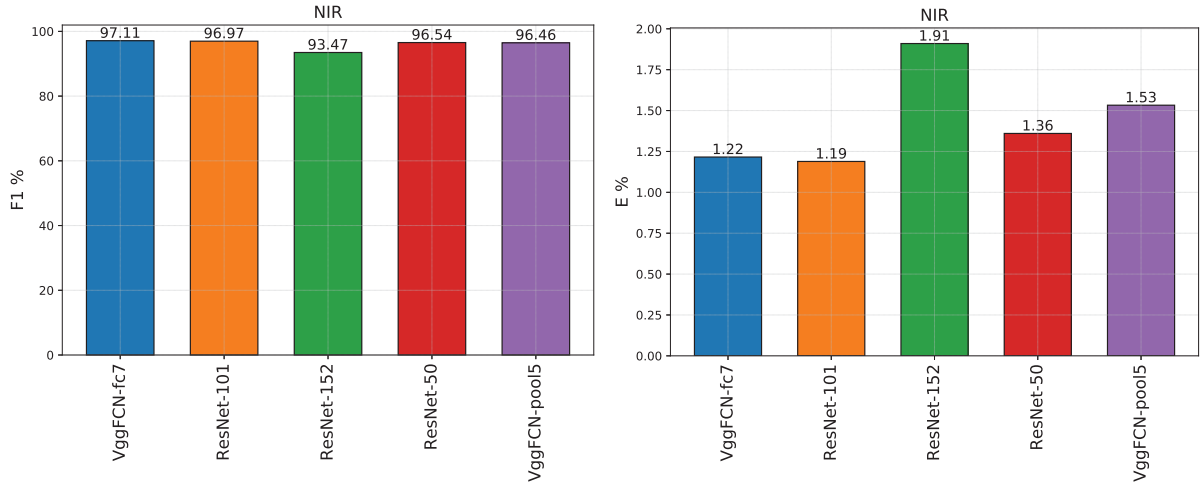
Neste caso, os experimentos com os *baselines* não foram realizados por que possuem ajustes de parâmetros referentes ao tamanho das íris, conforme apresentado na Seção 4.4. Além disso, o método específico para NIR ou VIS seria prejudicado, portanto uma comparação nestes 3 *datasets* (NIR_{dt}, VIS_{dt} e Todos_{dt}) não é justa.

As abordagens propostas alcançaram pouca variação dos valores de F1 para os 5 modelos, conforme apresentado na Tabela 5.9. O VggFCN-$fc7$ (linha em verde na tabela) apresentou o maior resultado ($97{,}11 \pm 01{,}48\%$), em comparação com as demais. O menor resultado (93,47%) foi obtido pela ResNet-152, com desvio padrão de $\approx \pm 10\%$. Ao analisar visualmente as imagens segmentadas pela ResNet-152, observou-se que a quantidade de FPs, nas imagens provenientes do *dataset* CasiaT4, foi bem maior que nas demais, principalmente em regiões de pele, bem fora do olho.

Tabela 5.9: Resultados de segmentação da íris no *dataset* $\mathbf{NIR}_{dt}$ pelas 5 abordagens propostas.

Método	**Precision %**	**Recall %**	**F1 %**	**IoU %**	**E %**
VggFCN-$fc7$	$97{,}03 \pm 01{,}91$	$97{,}23 \pm 02{,}11$	$97{,}11 \pm 01{,}48$	$94{,}41 \pm 02{,}69$	$01{,}22 \pm 00{,}58$
VggFCN-*pool*5	$96{,}00 \pm 02{,}68$	$96{,}97 \pm 01{,}35$	$96{,}46 \pm 01{,}54$	$93{,}20 \pm 02{,}78$	$01{,}53 \pm 00{,}72$
ResNet-50	$96{,}52 \pm 02{,}77$	$96{,}76 \pm 05{,}12$	$96{,}54 \pm 03{,}71$	$93{,}50 \pm 05{,}23$	$01{,}36 \pm 00{,}73$
ResNet-101	$96{,}79 \pm 02{,}77$	$97{,}27 \pm 03{,}61$	$96{,}97 \pm 02{,}77$	$94{,}24 \pm 04{,}21$	$01{,}19 \pm 00{,}58$
ResNet-152	$94{,}04 \pm 14{,}28$	$94{,}16 \pm 04{,}38$	$93{,}47 \pm 10{,}07$	$89{,}04 \pm 13{,}79$	$01{,}91 \pm 01{,}45$

Entretanto observa-se pouca variação nos valores referentes ao E, conforme ilustrado na Figura 5.9. Essa variação está entre 01,19% (menor erro para ResNet-101, ilustrada pela barra laranja) e maior erro 01,91% (ResNet-152, barra verde).



(a) (b)

Figura 5.9: Resultados das abordagens propostas no *dataset* NIR_{dt}. (a) representa a métrica F1, enquanto (b) representa o E.

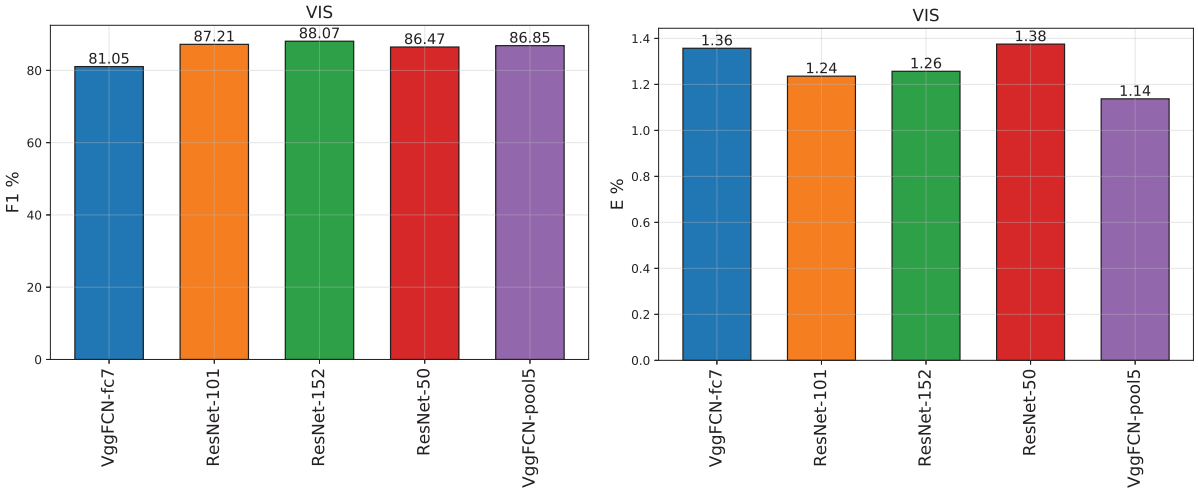
Os resultados no *dataset* VIS_{dt} são apresentados conforme a Tabela 5.10. Por ser mais desafiador, as abordagens propostas apresentaram valores um pouco abaixo do que os obtidos no NIR_{dt}. O ResNet-152 alcançou o maior valor de F1 ($88{,}07 \pm 13{,}18\%$), enquanto que o VggFCN-$fc7$ obteve o menor valor ($81{,}05 \pm 25{,}10\%$).

Tabela 5.10: Resultados de segmentação da íris no *dataset* $\textbf{VIS}_{dt}$ pelas 5 abordagens propostas.

Método	**Precision %**	**Recall %**	**F1 %**	**IoU %**	**E %**
VggFCN-$fc7$	$86{,}74 \pm 16{,}27$	$81{,}22 \pm 27{,}65$	$81{,}05 \pm 25{,}10$	$73{,}44 \pm 25{,}41$	$01{,}36 \pm 01{,}25$
VggFCN-*pool*5	$90{,}99 \pm 11{,}91$	$85{,}56 \pm 16{,}84$	$86{,}85 \pm 16{,}84$	$79{,}43 \pm 18{,}36$	$01{,}14 \pm 00{,}91$
ResNet-50	$84{,}71 \pm 14{,}66$	$90{,}13 \pm 14{,}01$	$86{,}47 \pm 14{,}06$	$78{,}06 \pm 15{,}75$	$01{,}38 \pm 00{,}96$
ResNet-101	$87{,}41 \pm 13{,}83$	$89{,}09 \pm 16{,}65$	$87{,}21 \pm 15{,}78$	$79{,}62 \pm 16{,}80$	$01{,}24 \pm 00{,}87$
ResNet-152	$87{,}62 \pm 11{,}28$	$90{,}41 \pm 15{,}01$	$88{,}07 \pm 13{,}18$	$80{,}38 \pm 14{,}71$	$01{,}26 \pm 00{,}88$

A Figura 5.10 ilustra os resultados com base na métrica F1 e E para as abordagens propostas. Observe que o VggFCN-*pool*5 obteve o menor valor de E (1,14%) entre todas as abordagens, enquanto a ResNet-50 obteve 1,38%. Entretanto, a variação do E foi de apenas (0,24%), mostrando um comportamento similar entre os métodos.

Por fim, o *dataset* Todos_{dt} possui a combinação de ambos os domínios (NIR e VIS) é considerado o mais desafiador de todos (deste trabalho) em virtude disso. Os resultados nesse *dataset* são apresentados na Tabela 5.11. O VggFCN-$fc7$ e VggFCN-*pool*5 obtiveram resultados de F1 similares ($92{,}18 \pm 14{,}70\%$ e $92{,}79 \pm 12{,}76\%$ respectivamente), sendo este último, o maior entre todos. ResNet-101 e ResNet-152 foram respectivamente $90{,}00 \pm 16{,}39\%$ e $90{,}74 \pm 15{,}70\%$, sendo também semelhantes. O menor resultado foi obtido por ResNet-50 de apenas $87{,}72 \pm 19{,}91\%$. Isso pode ser justificado pois com uma análise visual da segmentação, observou-se maior parte dos erros de segmentação estão nas imagens VIS, principalmente no *dataset* MICHE-I. Acredita-se que seja devido a quantidade de imagens NIR ser muito maior do que VIS, conforme apresentado na Tabela 4.1. Isso influencia no treinamento das abordagens de forma similar ao problema de desbalanceamento das classes.



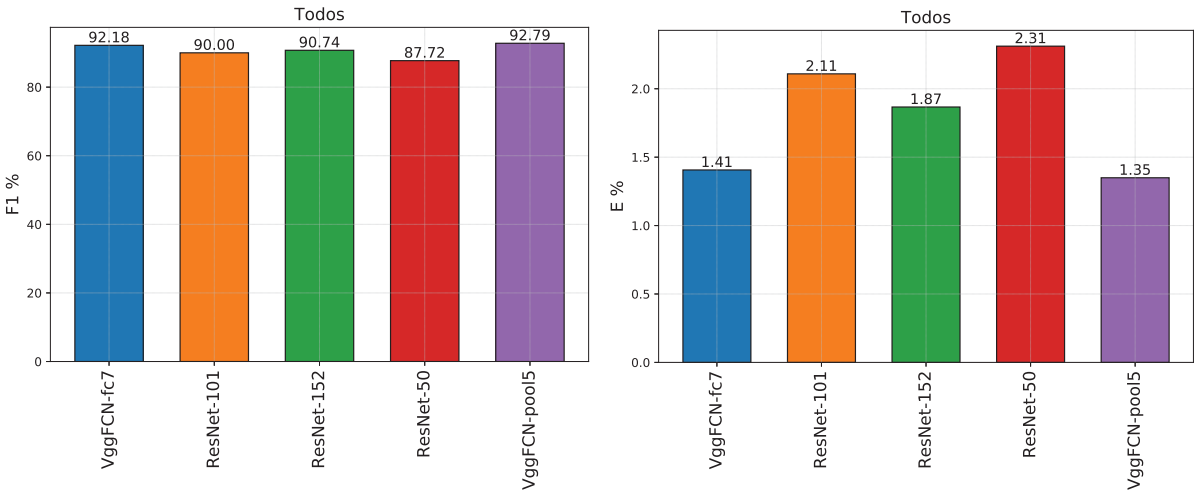
(a) (b)

Figura 5.10: Resultados dos métodos no *dataset* VIS_{dt}. (a) representa a métrica F1, enquanto (b) representa o E.

Tabela 5.11: Resultados de segmentação da íris no *datasets* Todos_{dt} pelas 5 abordagens propostas.

Método	**Precision %**	**Recall %**	**F1 %**	**IoU %**	**E %**
VggFCN-$fc7$	$91{,}82 \pm 13{,}97$	$93{,}88 \pm 13{,}72$	$92{,}18 \pm 14{,}70$	$87{,}70 \pm 16{,}36$	$01{,}41 \pm 00{,}95$
VggFCN-*pool*5	$92{,}70 \pm 11{,}13$	$94{,}08 \pm 12{,}83$	$92{,}79 \pm 12{,}76$	$88{,}28 \pm 14{,}83$	$01{,}35 \pm 00{,}74$
ResNet-50	$90{,}36 \pm 12{,}57$	$88{,}66 \pm 22{,}29$	$87{,}72 \pm 19{,}91$	$81{,}78 \pm 20{,}95$	$02{,}31 \pm 02{,}03$
ResNet-101	$90{,}66 \pm 10{,}72$	$91{,}72 \pm 18{,}06$	$90{,}00 \pm 16{,}39$	$84{,}35 \pm 17{,}19$	$02{,}11 \pm 01{,}29$
ResNet-152	$92{,}99 \pm 09{,}33$	$90{,}82 \pm 17{,}52$	$90{,}74 \pm 15{,}70$	$85{,}42 \pm 16{,}83$	$01{,}87 \pm 01{,}26$

Quanto ao E, observe na Figura 5.11, que o menor valor foi obtido por VggFCN-*pool*5 (apenas 01,35%), ilustrado pela barra na cor lilás. Vale lembrar que esse E, é



(a) (b)

Figura 5.11: Resultados dos métodos no *dataset* Todos_{dt}. (a) representa a métrica F1, enquanto (b) representa o E.

influenciado pela quantidade de *pixels* presente no *dataset* MICHE-I e pela quantidade de imagens testadas, dando a falsa impressão que o erro é baixo.

5.4 Análise Estatística dos Resultados

Um dos objetivos deste trabalho é avaliar o comportamento das abordagens propostas, que envolvem modelos de DCNNs, aplicados em segmentação de íris, em cenários e domínios distintos. Analisar apenas os resultados das métricas nesses cenários específicos não representa, de fato, qual abordagem obteve os melhores resultados. Além disso, não há como assumir que essas abordagens são estatisticamente melhores, piores ou iguais. Conforme discutido na Seção 4.6, o teste de Friedman (1937) é utilizado para tal finalidade, com pós-teste de Nemenyi (1963), a fim de identificar o quão diferente (*critical difference* (CD)) os métodos são entre si, e assim, constatar quais são os melhores ou piores.

Inicialmente aplicamos o teste-t, pois o primeiro experimento foi realizado apenas no *dataset* NICE.I com a distribuição original da competição, (ver Seção 4.3 para maiores detalhes), e o teste de Friedman não se enquadra neste cenário (único *dataset*). Para isso o teste-t compara todos os métodos entre si (todos contra todos), em forma de pares, gerando um respectivo p (p-*value*). A Figura 5.12 apresenta os p-*value* correspondentes a cada uma dessas comparações. Observe que todos os p ilustrados em azul claro são menores que $\alpha = 0{,}05$, portanto a hipótese nula H_0 é rejeitada e assume-se que os métodos referentes a p são estatisticamente diferentes, conforme discutido no capítulo anterior.

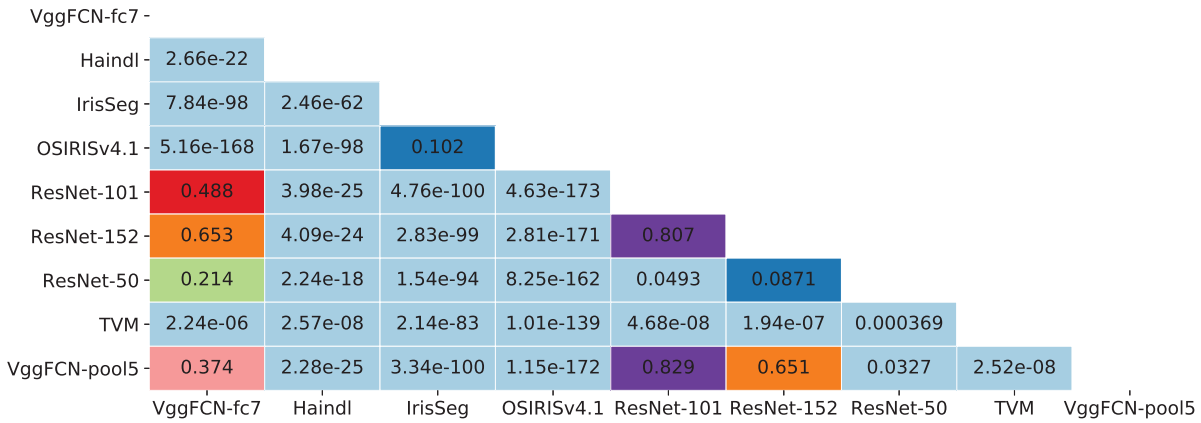


Figura 5.12: p-*value* obtidos na comparação dos *baselines* e das abordagens propostas (todos contra todos) para o *dataset* NICE.I, conforme o protocolo da competição. $p < \alpha$ representa os métodos que são estatisticamente diferentes, enquanto que $p > \alpha$ os métodos não são estatisticamente diferentes, ambos para $\alpha = 0{,}05$. Valores nas posições equivalentes foram omitidos pois a matriz é identidade.

Todas as abordagens propostas são estatisticamente diferentes dos *baselines*. Entretanto, entre essas abordagens, apenas dois pares apresentaram diferença estatística, são eles VggFCN-*pool*5 e ResNet-50, com $p = 0{,}0327$ e ResNet-50 e ResNet-101, com $p = 0{,}0493$, rejeitando novamente a hipótese nula H_0.

Posteriormente, os demais experimentos realizados foram validados por meio do teste de Friedman. A Tabela 5.12 apresenta os *ranks* médios para ambos os métodos (tanto abordagens propostas, quanto *baselines*) e *datasets*. Note que as abordagens propostas

VggFCN-$fc7$, ResNet-101, ResNet-152, ResNet-50 e VggFCN-*pool*5 obtiveram os maiores *ranks*, implicando no melhor desempenho em comparação com os *baselines*.

Tabela 5.12: *Ranks* médios para todos os métodos e todos os *datasets*, com base no teste de Friedman. p representa o *p-value* obtido nesse teste, a partir de $\alpha = 0{,}05$. (- - -) significa que o método não funcionou ou não foi utilizado no teste estatístico.

Métodos	**Datasets**		
	BioSec CasiaI3 CasiaT4 IITD-1 $p = 2{,}25 \cdot 10^{-7}$	NICE.I* CrEye-Iris MICHE-I $p = 5{,}02 \cdot 10^{-7}$	NIR_{dt} VIS_{dt} $Todos_{dt}$ $p = 0{,}8650$
VggFCN-$fc7$	3,75	2,33	2,66
Haindl	- - -	6,66	- - -
IrisSeg	6,00	8,66	- - -
OSIRISv4.1	7,00	8,33	- - -
ResNet-101	2,00	2,66	2,66
ResNet-152	2,25	3,00	3,00
ResNet-50	2,75	5,00	4,00
TVM	8,00	6,33	- - -
VggFCN-*pool*5	4,25	2,00	2,66

Além disso, os valores de $p = 2{,}25 \cdot 10^{-7}$ (BioSec, CasiaI3, CasiaT4 e IITD-1) e $p = 5{,}02 \cdot 10^{-7}$ (NICE.I*, CrEye-Iris e MICHE-I) são menores que $\alpha = 0{,}05$. Portanto, a condição de hipótese nula H_0 foi rejeitada, indicando que existe diferença estatística entre as abordagens propostas e os *baselines*. Entretanto, o valor de $p = 0{,}8650$, nos *datasets* mesclados (NIR_{dt}, VIS_{dt} e $Todos_{dt}$), é maior que $\alpha = 0{,}05$, indicando que a condição de hipótese nula H_0 deve ser aceita, não existindo diferença estatística entre as 5 abordagens propostas.

A Figura 5.13 ilustra a CD dos métodos testados nos *datasets* BioSec, CasiaI3, CasiaT4 e IITD-1, conforme a presença de diferença estatística discutida anteriormente. O eixo x representa os *ranks* médios obtidos por cada método (conforme a Tabela 5.12), com $CD \approx 3{,}71$. Linhas em negrito indicam grupos de métodos que não são significativamente diferentes (suas classificações médias diferem em menos que o valor de CD).

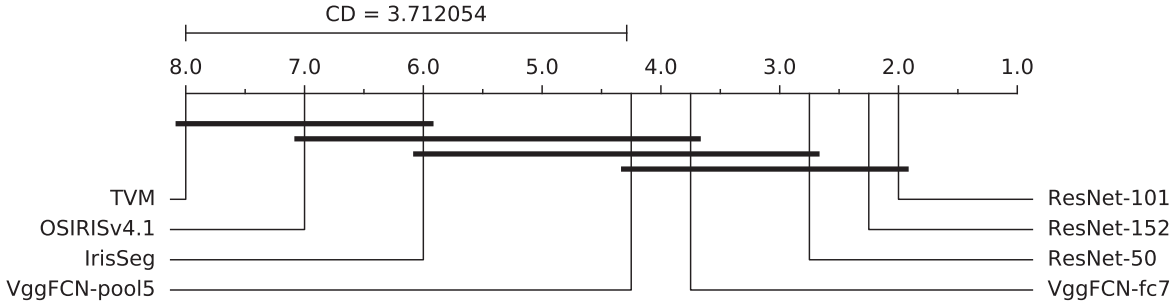


Figura 5.13: Diagrama de CD para os testes realizados nos *datasets* BioSec, CasiaI3, CasiaT4 e IITD-1, com $p = 2{,}25 \cdot 10^{-7}$ e $\alpha = 0{,}05$.

Observe que a abordagem ResNet-101 obteve o melhor *rank* (2,00) em comparação a todos os métodos. Entretanto, existe uma ligação (pela linha em negrito) entre esse método e o VggFCN-*pool*5, que representa que ambos são significativamente diferentes, mesmo VggFCN-*pool*5 estando agrupado com os piores métodos. Isso ocorre pois os *ranks*

médios dos dois são menores que a CD. Os métodos TVM e OSIRISv4.1 obtiveram os piores resultados, em comparação aos demais.

A Figura 5.14 ilustra a CD dos métodos testados nos *datasets* NICE.I*, CrEye-Iris e MICHE-I. Observe que a CD é de ≈ 4,00, sendo um pouco maior que a anterior. Neste caso, o melhor método foi VggFCN-*pool*5, que obtere *rank* médio de 2,00 nos três *datasets*. Em comparação entre VggFCN-$fc7$, ResNet-101 e ResNet-152, a CD foi bem próxima, aproximadamente 1, por isso estão agrupados à direita do diagrama.

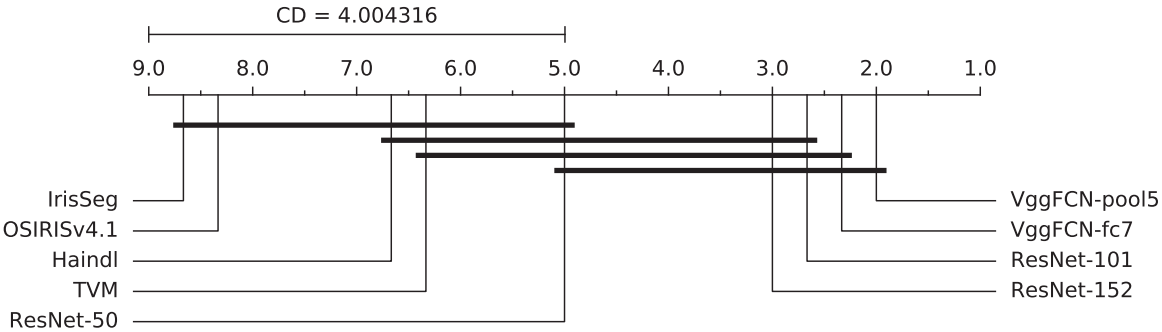


Figura 5.14: Diagrama de CD para os testes realizados nos *datasets* NICE.I*, CrEye-Iris e MICHE-I, com $p = 5{,}02 \cdot 10^{-7}$ e $\alpha = 0{,}05$.

Note também, que VggFCN-*pool*5 está conectado, pela linha em negrito, com o ResNet-50, indicando que as CD entre eles é menor que ≈ 4,00. O pior método é o IrisSeg, com *rank* médio de 8,66, e o segundo pior é o OSIRISv4.1, com 8,33. Esse pior resultado, para ambos os métodos pode ser justificado pelo fato de que eles não foram desenvolvidos para imagens VIS, apenas para NIR.

Conforme comentado anteriormente, a comparação entre apenas as abordagens propostas nos *datasets* mesclados (NIR_{dt}, VIS_{dt} e Todos_{dt}) apresentou $p = 0{,}8650$, e portanto a hipótese nula H_0 foi aceita, indicando que não há diferença estatística entre eles. Mesmo aceitando a hipótese nula, o diagrama CD foi calculado, para analisar a distribuição dessas abordagens, conforme ilustrado na Figura 5.15. Observe que as abordagens propostas são distribuídas conforme CD≈ 2,72, mostrando dois agrupamentos (VggFCN-$fc7$ e ResNet-101 à direita do diagrama), enquanto as demais, ficaram à esquerda.

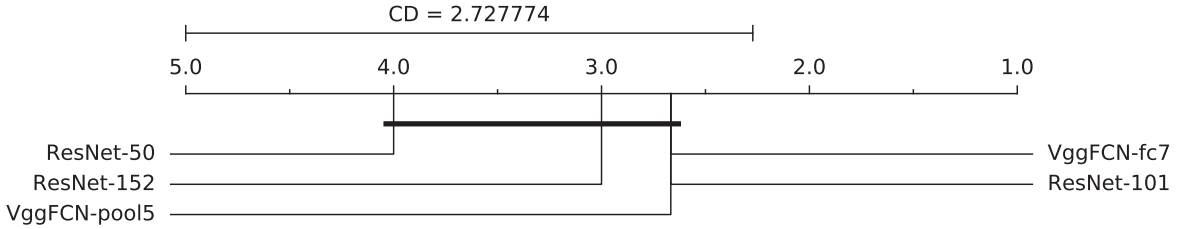


Figura 5.15: Diagrama de diferença crítica (CD) para os testes realizados nos *datasets* NIR_{dt}, VIS_{dt} e Todos_{dt}, com $p = 0{,}8650$ e $\alpha = 0{,}05$.

As abordagens VggFCN-$fc7$, ResNet-101 e VggFCN-*pool*5 ficaram empatadas, com o mesmo *rank* médio (2,66), enquanto a ResNet-152 e ResNet-50 obtiveram 3,0 e 4,0 respectivamente. Esse comportamento semelhante das 3 abordagens (VggFCN-$fc7$, ResNet-101 e VggFCN-*pool*5) é justificado pois, conforme as Tabelas 5.9, 5.10 e 5.11, cada abordagem alcançou maiores resultados, respectivamente.

Vale lembrar que cada abordagem obteve maior resultado em um domínio específico, ou seja, ao mesclar os *datasets*, o comportamento das abordagens foi similar. Note que a VggFCN-$fc7$ e VggFCN-*pool*5 possuem a arquitetura semelhante, oriundas do modelo VGG, enquanto a ResNet-101 e ResNet-152 provém do modelo de redes residuais ResNet. A seguir, uma análise qualitativa da segmentação é realizada com o intuito de ilustrar a segmentação das imagens, de acordo com cada método (*baselines* e abordagens propostas).

5.5 Análise Qualitativa da Segmentação

O erro de segmentação da íris obtida pelos *baselines* e pelas abordagens propostas, nos *datasets* NIR (CasiaI3 e CasiaT4) é apresentado na Figura 5.16. A imagem de cada um desses *datasets* apresentados foi escolhida com base na análise visual do erro entre os *baselines* e as abordagens propostas, de forma que nenhum método seja desfavorecido ou favorecido. Os *pixels* verdes representam os FPs enquanto os *pixels* vermelhos representam FNs, ou seja, ambos representam o erro de segmentação, conforme descrito na Seção 4.2.

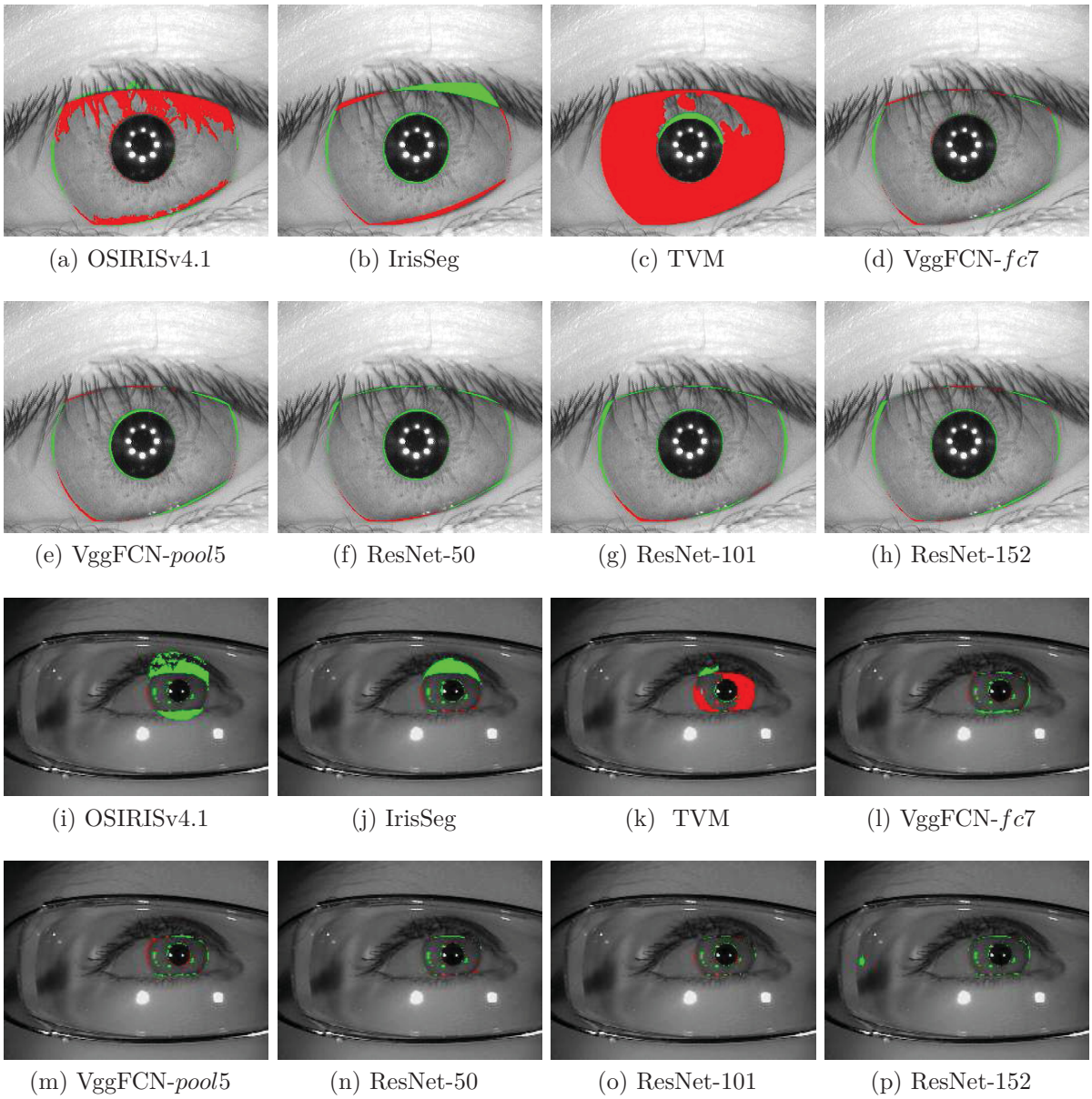
Observe que o OSIRISv4.1 e o IrisSeg segmentaram de forma errônea a região dos cílios, não distinguindo essa região de ruído, mesmo sendo desenvolvidos para este tipo de imagem (NIR). Entretanto as abordagens propostas apresentaram boa capacidade de distinção entre os cílios e as pálpebras. A ResNet-152 apresentou uma quantidade de FPs na borda da lente do óculos (Figura 5.16p), confundindo a leve distorção da lente do óculos, com provavelmente o contorno da íris. Note também, que os erros de segmentação apresentados por essas abordagens ocorrem, em sua maioria, nos pontos de transição entre pupila e íris, e íris e esclera/pálpebra. Isso ocorre pela redução de informação, durante a execução de *downsampling* no codificador, que não é recuperada pelo *upsampling* no decodificador (ver Capítulo 3 para maiores detalhes).

No domínio das imagens VIS os erros de segmentação são apresentados de forma similar, conforme discutido anteriormente, na Figura 5.17 para todos os *baselines* e abordagens propostas, nos *datasets* CrEye-Iris e MICHE-I. Observe que ambos OSIRISv4.1 e IrisSeg apresentaram maior quantidade de FPs e FNs, inclusive mostrando não serem tão sensíveis aos reflexos presentes em imagens VIS, como era de se esperar. Entretanto, mesmo sendo desenvolvidos para imagens VIS, TVM e Haindl apresentaram presença significativa de FPs e FNs.

Já as abordagens propostas, especificamente VggFCN-$fc7$ e VggFCN-*pool*5 foram mais precisas (pouca presença de FNs) na região superior da íris (mais escura devido a sombra causada pelos cílios), conforme as Figuras 5.17n e 5.17o. Entretanto, as abordagens ResNet-50, ResNet-101 e ResNet-152 apresentaram maior presença de FNs nessas regiões de sombra, de forma similar uma das outras. Além disso, todas as abordagens propostas mostraram boa distinção entre os pontos de reflexo na íris, causados pela variação de iluminação do ambiente.

Outro fator observado foi que os erros de segmentação nas imagens referentes ao *dataset* MICHE-I no *dataset* mesclado Todos$_{dt}$, apresentou maior quantidade de FPs (*pixels* em verde) e FNs (*pixels* em vermelho) conforme ilustrado na Figura 5.18. Conforme discutido na Seção 5.2 (resultados *dataset* MICHE-I) o redimensionamento das imagens, em virtude da limitação de *hardware*, afeta diretamente no desempenho das abordagens. As imagens da Figura 5.18a e 5.18b possuem originalmente resolução de 1536×2048 (largura e altura), enquanto 5.18c e 5.18d possuem 2448×3264.

Para o treinamento das abordagens propostas, o redimensionamento foi realizado (479×455), com base nas dimensões médias de todo o *subset* de treinamento do *dataset*



(a) OSIRISv4.1 (b) IrisSeg (c) TVM (d) VggFCN-$fc7$

(e) VggFCN-*pool*5 (f) ResNet-50 (g) ResNet-101 (h) ResNet-152

(i) OSIRISv4.1 (j) IrisSeg (k) TVM (l) VggFCN-$fc7$

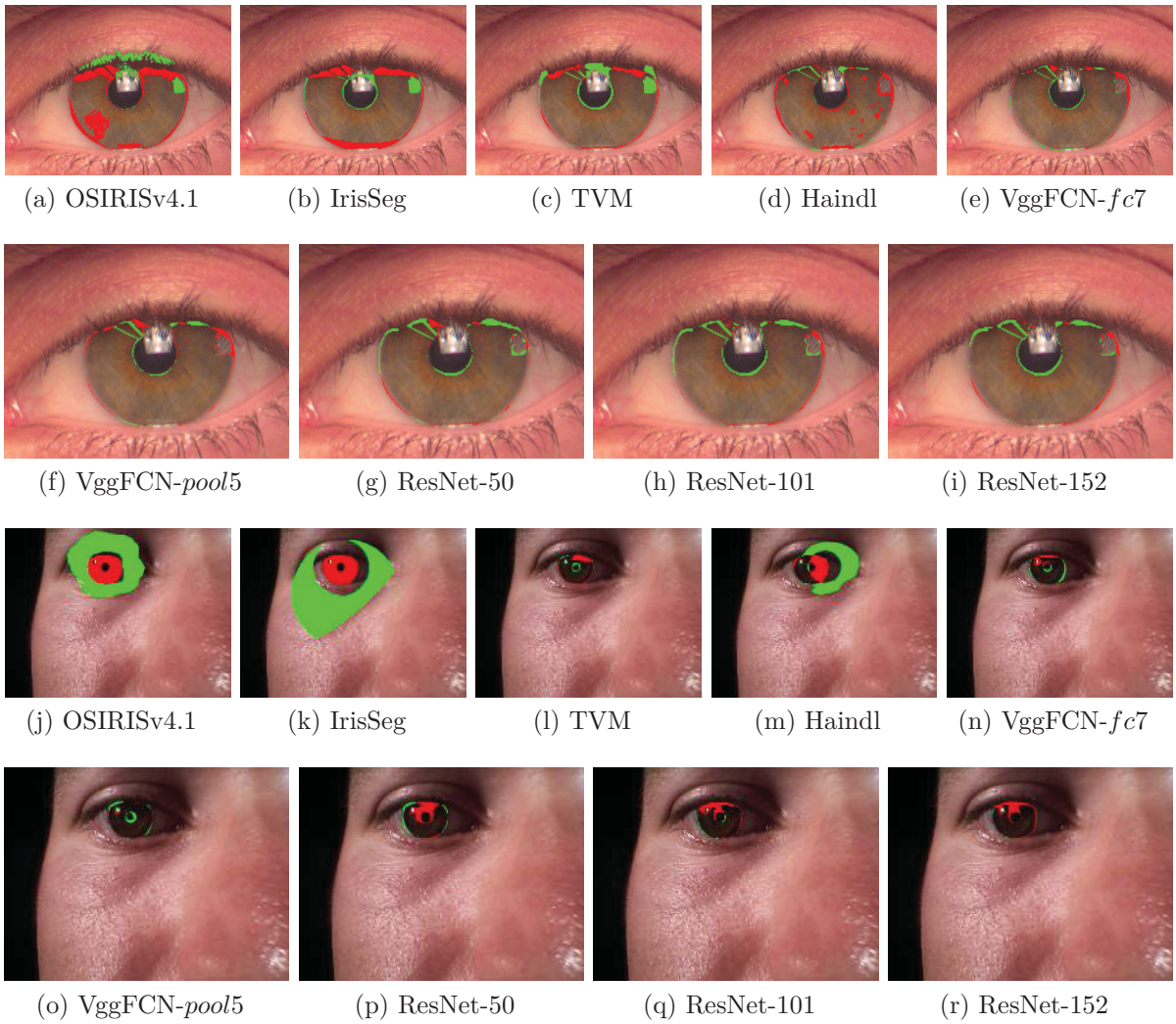
(m) VggFCN-*pool*5 (n) ResNet-50 (o) ResNet-101 (p) ResNet-152

Figura 5.16: Resultados qualitativos obtidos pelos *baselines* (OSIRISv4.1, IrisSeg e TVM) e pelas abordagens propostas (VggFCN-$fc7$, VggFCN-*pool*5, ResNet-50, ResNet-101 e ResNet-152) nos *datasets* NIR CasiaI3 (a)-(h) e CasiaT4 (i)-(p). Os *Pixels* em verde e em vermelho representam os FPs e FNs, respectivamente.

Todos$_{dt}$. Nos demais *datasets* esse redimensionamento não afetou drasticamente as imagens, como na MICHE-I, pois conforme a Tabela 4.1, a resolução desses *datasets* não foi muito reduzida. Na etapa de *upsampling* do decodificador, a imagem resultante da segmentação é reconstruída para sua resolução original.

5.6 Considerações Finais

Este Capítulo apresentou todos os resultados obtidos por meio dos experimentos realizados, representados em tabelas e gráficos de barras. Apresentou também uma análise



(a) OSIRISv4.1 (b) IrisSeg (c) TVM (d) Haindl (e) VggFCN-$fc7$

(f) VggFCN-*pool*5 (g) ResNet-50 (h) ResNet-101 (i) ResNet-152

(j) OSIRISv4.1 (k) IrisSeg (l) TVM (m) Haindl (n) VggFCN-$fc7$

(o) VggFCN-*pool*5 (p) ResNet-50 (q) ResNet-101 (r) ResNet-152

Figura 5.17: Resultados qualitativos obtidos por todos os *baselines* e pelas abordagens propostas nos *datasets* VIS CrEye-Iris (a)-(i) e MICHE-I (j)-(r). Os *Pixels* em verde e em vermelho representam os FPs e FNs, respectivamente.

estatística para validação das abordagens propostas, em relação aos *baselines*, por meio de teste-t e Friedman, este ultimo com pós teste de Nemenyi e diagramas de CD, no qual foi possível constatar estatisticamente a diferença significativa dessas abordagens, mostrando o quanto elas foram melhores. Além disso, os erros de segmentação foram apresentados, sendo possível uma análise visual desses resultados em algumas imagens, mostrando também o quanto o redimensionamento das imagens interfere nesse erro.

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

Figura 5.18: Erros de segmentação quando a imagem sofre um redimensionamento de mais de 60%. Os *Pixels* em verde e em vermelho representam os FPs e FNs, respectivamente.

6 Conclusões e Trabalhos Futuros

A segmentação da íris é a tarefa mais desafiadora em um sistema biométrico usando esta forma de biometria, devido à grande quantidade de ruídos presentes nas imagens, principalmente em bases de dados adquiridas sob condições não controladas (Jan, 2017; Santos et al., 2015). Diversas técnicas propõem a segmentação da íris usando abordagens geométricas por exemplo. Entretanto, essas abordagens são limitadas ao tipo da base de dados (VIS, NIR, e se as imagens são adquiridas em condições controladas ou não), não possuindo uma capacidade de generalização independente do *dataset* utilizado.

O uso de CNNs tem permitido a obtenção de resultados estado-da-arte em diversos problemas de visão computacional como segmentação, reconhecimento e classificação em aplicações de biometria, imagens médicas, sistemas de segurança, monitoramento, etc. (LeCun et al., 2015; Ahuja et al., 2017; Dumoulin e Visin, 2016; Krizhevsky et al., 2012; Oquab et al., 2014; Teichmann et al., 2016; Tian et al., 2017; Brandao et al., 2017).

Partindo desse pressuposto, este trabalho apresentou 5 abordagens robustas que utilizam DCNN, seguindo uma arquitetura codificador-decodificador para segmentação de íris em imagens NIR, VIS, obtidas em condições controladas e não controladas. Essas abordagens baseiam-se nos modelos VGG16 (VggFCN-$fc7$ e VggFCN-*pool*5) e ResNet (ResNet-50, ResNet-101 e ResNet-152) em seu codificador, combinados com FCN no decodificador. Os resultados de segmentação da DCNN foram comparados com 4 *frameworks* da literatura (OSIRISv4.1, IrisSeg, Haindl e TVM). Com essa comparação, constatou-se que as abordagens propostas obtiveram os melhores resultados em todos os 7 *datasets* distintos, utilizados nos experimentos.

Além disso, com base em análise estatística a melhor abordagem proposta nos *datasets* NIR e VIS foram respectivamente ResNet-101 e VggFCN-*pool*5, com $rank = 2{,}00$ conforme a Tabela 5.12. Já para os *datasets* mesclados (NIR_{dt}, VIS_{dt} e Todos_{dt}) as abordagens propostas não obtiveram diferença estatística, entretanto VggFCN-$fc7$, ResNet-50 e VggFCN-*pool*5 obtiveram $rank = 2{,}66$, ficando empatadas.

A transferência de aprendizagem, oriundos do *dataset* ILSVRC, e o *fine-tunning* foram essenciais para alcançar esses melhores resultados, uma vez que a quantidade de imagens para o treinamento das abordagens propostas é considerada pequena. Além disso, uma combinação entre os *datasets* NIR e VIS foi realizada, possibilitando analisar a capacidade de generalização e robustez das abordagens propostas. Durante o desenvolvimento deste trabalho, não foi encontrada na literatura nenhuma abordagem que explora a combinação dos domínios NIR e VIS, da maneira que exploramos.

Portanto o uso de modelos pré-treinados provenientes de outros *datasets*, traz excelentes benefícios para a aprendizagem das DCNNs. Além disso, as abordagens propostas mostraram ser mais discriminantes em relação a variações de ruídos, como reflexos, mudanças de iluminação e oclusões por exemplo, e tem uma dependência menor de etapas de pré e pós processamento com relação as diversas técnicas apresentadas na literatura.

Também, foram rotuladas manualmente 2.431 imagens de íris, dos *datasets* CasiaT4, CrEye-Iris e MICHE-I, que serão disponibilizados para a comunidade acadêmica, a fim de auxiliar no desenvolvimento e avaliação de novas técnicas de segmentação de íris.

6.1 Trabalhos Futuros

Apesar das abordagens propostas terem alcançado os melhores resultados, conforme comentado anteriormente, pretende-se na forma de trabalhos futuros:

1. Realizar a segmentação em duas etapas, i.e., primeiro detectar a íris e então segmentá-la, apenas nessa região detectada, com o intuito de reduzir a quantidade de FPs nas regiões de fora da íris, e evitar um redimensionamento muito radical das imagens, melhorando os resultados.

2. Criar uma etapa de pós-processamento, com o intuito de refinar a previsão das abordagens propostas (DCNNs), uma vez que elas apresentaram uma leve presença de erros de segmentação em regiões de transição das íris, como por exemplo, entre a íris e a pupila, íris e esclera/pálpebra.

3. Definir uma abordagem geral (em dois passos), que primeiro classifica o tipo de sensor ou imagem, e em seguida realiza a segmentação usando um modelo de DCNN específico e adaptado para cada tipo específico.

Referências

Abate, A. F., Frucci, M., Galdi, C. e Riccio, D. (2015). Bird: Watershed based {IRis} detection for mobile devices. *Pattern Recognition Letters*, 57:43–51.

Ahuja, K., Islam, R., Barbhuiya, F. A. e Dey, K. (2017). Convolutional neural networks for ocular smartphone-based biometrics. *Pattern Recognition Letters*, 91:17–26.

Alkassar, S., Woo, W. L., Dlay, S. S. e Chambers, J. A. (2016). Robust sclera recognition system with novel sclera segmentation and validation techniques. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, páginas 1–13.

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

Arsalan, M., Naqvi, R. A., Kim, D. S., Nguyen, P. H., Owais, M. e Park, K. R. (2018). Irisdensenet: Robust iris segmentation using densely connected fully convolutional networks in the images by visible light and near-infrared light camera sensors. *Sensors*, 18(5):1501.

Badejo, J. A., Atayero, A. A. e Ibiyemi, T. S. (2016). A robust preprocessing algorithm for iris segmentation from low contrast eye images. Em *Future Technologies Conference*, páginas 567–576.

Badejo, J. A., Majekodunmi, T. O. e Atayero, A. A. (2011). Development of cuiris: A dark-skinned african iris dataset for enhancement of image analysis and robust personal recognition. *Proceedings of the World Congress on Engineering and Computer Science*, páginas 624–629.

Badrinarayanan, V., Kendall, A. e Cipolla, R. (2015). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *CoRR*, abs/1511.00561.

Ballard, D. H. (1981). Generalizing the hough transform to detect arbitrary shapes. *Pattern Recognition*, 13(2):111–122.

Barra, S., Casanova, A., Narducci, F. e Ricciardi, S. (2015). Ubiquitous iris recognition by means of mobile devices. *Pattern Recognition Letters*, 57:66–73.

Bashir, F., Casaverde, P., Usher, D. e Friedman, M. (2008). Eagle-eyes: a system for iris recognition at a distance. Em *Conference on Technologies for Homeland Security*, páginas 426–431. IEEE.

Bazrafkan, S., Thavalengal, S. e Corcoran, P. (2018). An end to end deep neural network for iris segmentation in unconstrained scenarios. *Neural Networks*, 106:79 – 95.

Bengio, Y., Courville, A. e Vincent, P. (2013). Representation learning: A review and new perspectives. *Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1798–1828.

Bezerra, C. S., Laroca, R., Lucio, D. R., Severo, E., Oliveira, L. F., Britto Jr., A. S. e Menotti, D. (2018). Robust iris segmentation based on fully convolutional networks and generative adversarial networks. Em *Conference on Graphics, Patterns and Images (SIBGRAPI)*, páginas 281–288.

Bezerra, C. S. e Menotti, D. (2017). Fully convolutional neural network for ocular iris semantic segmentation. Em *XIII Brazilian Computer Vision Workshop*, páginas 1–6.

Brandao, P., Mazomenos, E., Ciuti, G., Caliò, R., Bianchi, F., Menciassi, A., Dario, P., Koulaouzidis, A., Arezzo, A. e Stoyanov, D. (2017). Fully convolutional neural networks for polyp segmentation in colonoscopy. *Proc.SPIE*, 10134:10134–10134–7.

Canny, J. (1986). A computational approach to edge detection. *Transactions on pattern analysis and machine intelligence*, páginas 679–698.

Chan, T. F. e Vese, L. A. (2001). Active contours without edges. *Transactions on Image Processing*, 10(2):266–277.

Chen, A., Zhou, T., Icke, I., Parimal, S., Dogdas, B., Forbes, J., Sampath, S., Bagchi, A. e Chin, C.-L. (2018). Transfer learning for the fully automatic segmentation of left ventricle myocardium in porcine cardiac cine MR images. Em *Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges*, páginas 21–31, Cham. Springer International Publishing.

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

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

Cheng, M. M., Mitra, N. J., Huang, X., Torr, P. H. S. e Hu, S. M. (2015). Global contrast based salient region detection. *Transactions on Pattern Analysis and Machine Intelligence*, 37(3):569–582.

Cover, T. e Hart, P. (1967). Nearest neighbor pattern classification. *Transactions on Information Theory*, 13(1):21–27.

Crihalmeanu, S., Ross, A., Schuckers, S. e Hornak, L. (2007). A protocol for multibiometric data acquisition storage and dissemination. *Dept. Comput. Sci. Elect. Eng., West Virginia Univ., Morgantown, WV, USA, Tech. Rep.*

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

Daugman, J. (2001). Statistical richness of visual phase information: Update on recognizing persons by iris patterns. *International Journal of Computer Vision*, 45(1):25–38.

Daugman, J. (2003). The importance of being random: statistical principles of iris recognition. *Pattern Recognition*, 36(2):279–291.

Daugman, J. (2004). How iris recognition works. *IEEE Transactions on circuits and systems for video technology*, 14(1):21–30.

Daugman, J. (2007). New methods in iris recognition. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, 37(5):1167–1175.

De Marsico, M., Nappi, M. e Daniel, R. (2010). Isis: Iris segmentation for identification systems. Em *International Conference on Pattern Recognition*, páginas 2857–2860.

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

Demšar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *J. Mach. Learn. Res.*, 7:1–30.

Dumoulin, V. e Visin, F. (2016). A guide to convolution arithmetic for deep learning. *ArXiv e-prints*.

Everingham, M., Eslami, S. M. A., Van Gool, L., Williams, C. K. I., Winn, J. e Zisserman, A. (2015). The pascal visual object classes challenge: A retrospective. *International Journal of Computer Vision*, 111(1):98–136.

Fawcett, T. (2006). An introduction to roc analysis. *Pattern Recognition Letters*, 27(8):861–874.

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

Fischler, M. A. e Bolles, R. C. (1981). Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395.

Flach, P. (2012). *Machine Learning: The Art and Science of Algorithms That Make Sense of Data*. Cambridge University Press, New York, NY, USA.

Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association*, 32(200):675–701.

Fritsch, J., Kühnl, T. e Geiger, A. (2013). A new performance measure and evaluation benchmark for road detection algorithms. Em *International Conference on Intelligent Transportation Systems*, páginas 1693–1700.

Gangwar, A., Joshi, A., Singh, A., Alonso-Fernandez, F. e Bigun, J. (2016). IrisSeg: A fast and robust iris segmentation framework for non-ideal iris images. Em *International Conference on Biometrics*, páginas 1–8.

Ge, F., Wang, S. e Liu, T. (2006). Image-segmentation evaluation from the perspective of salient object extraction. Em *Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, páginas 1146–1153.

Glorot, X. e Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. Em *International Conference on Artificial Intelligence and Statistics*, volume 9, páginas 249–256, Chia Laguna Resort, Sardinia, Italy. PMLR.

Goodfellow, I., Bengio, Y. e Courville, A. (2016). *Deep Learning*. MIT Press. `http://www.deeplearningbook.org`.

Haindl, M. (2012). Visual data recognition and modeling based on local markovian models. Em *Mathematical Methods for Signal and Image Analysis and Representation*, páginas 241–259, London. Springer.

Haindl, M. e Krupicka, M. (2015). Unsupervised detection of non-iris occlusions. *Pattern Recognition Letters*, 57:60–65.

Hastie, T., Tibshirani, R. e Friedman, J. (2001). *The Elements of Statistical Learning*. Springer Series in Statistics. Springer New York Inc., New York, NY, USA.

He, K., Zhang, X., Ren, S. e Sun, J. (2016). Deep residual learning for image recognition. Em *Conference on Computer Vision and Pattern Recognition*, páginas 770–778.

He, Z., Tan, T., Sun, Z. e Qiu, X. (2009). Toward accurate and fast iris segmentation for iris biometrics. *IEEE transactions on pattern analysis and machine intelligence*, 31(9):1670–1684.

ho Cho, D., Park, K. R., Rhee, D. W., Kim, Y. e Yang, J. (2006). Pupil and iris localization for iris recognition in mobile phones. Em *International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing*, páginas 197–201.

Hofbauer, H., Alonso-Fernandez, F., Wild, P., Bigun, J. e Uhl, A. (2014). A ground truth for iris segmentation. Em *International Conference on Pattern Recognition*, páginas 527–532.

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

Hu, Y., Sirlantzis, K. e Howells, G. (2015). Improving colour iris segmentation using a model selection technique. *Pattern Recognition Letters*, 57:24–32.

Jain, A. K., Hong, L. e Bolle, R. (1997). On-line fingerprint verification. *Transactions on Pattern Analysis and Machine Intelligence*, 19(4):302–314.

Jain, A. K., Nandakumar, K. e Ross, A. (2016). 50 years of biometric research: Accomplishments, challenges, and opportunities. *Pattern Recognition Letters*, 79:80–105.

Jain, A. K., Ross, A. e Prabhakar, S. (2004). An introduction to biometric recognition. *Transactions on circuits and systems for video technology*, 14(1):4–20.

Jalilian, E. e Uhl, A. (2017). *Iris Segmentation Using Fully Convolutional Encoder-Decoder Networks*, páginas 133–155. Springer International Publishing, Cham.

Jan, F. (2017). Segmentation and localization schemes for non-ideal iris biometric systems. *Signal Processing*, 133:192–212.

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

Jillela, R. e Ross, A. (2015). Segmenting iris images in the visible spectrum with sequeira,applications in mobile biometrics. *Pattern Recognition Letters*, 57:4–16.

Jillela, R., Ross, A., Boddeti, V. N., Kumar, V., Hu, X., Plemmons, R. e Pauca, P. (2013). Iris segmentation for challenging periocular images. Em *Handbook of Iris Recognition*, páginas 281–308. Springer.

Jillela, R. e Ross, A. A. (2016). Methods for iris segmentation. Em *Handbook of Iris Recognition*, páginas 137–184, London. Springer London.

Kang, B. J. e Park, K. R. (2007). A robust eyelash detection based on iris focus assessment. *Pattern Recognition Letters*, 28(13):1630–1639.

Kendall, A., Badrinarayanan, V. e Cipolla, R. (2015). Bayesian segnet: Model uncertainty in deep convolutional encoder-decoder architectures for scene understanding. *CoRR*, abs/1511.02680.

Krizhevsky, A., Sutskever, I. e Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. Em *Advances in Neural Information Processing Systems 25*, páginas 1097–1105. Curran Associates, Inc.

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

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

Laroca, R., Severo, E., Zanlorensi, L. A., Oliveira, L. S., Gonçalves, G. R., Schwartz, W. R. e Menotti, D. (2018). A robust real-time automatic license plate recognition based on the yolo detector. Em *2018 International Joint Conference on Neural Networks (IJCNN)*, páginas 1–10.

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

Li, H., Sun, Z. e Tan, T. (2012). Accurate iris localization using contour segments. Em *International Conference on Pattern Recognition-ICPR*, páginas 3398–3401.

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

Liu, N., Li, H., Zhang, M., Liu, J., Sun, Z. e Tan, T. (2016a). Accurate iris segmentation in non-cooperative environments using fully convolutional networks. Em *Int. Conf. on Biometrics*, páginas 1–8.

Liu, N., Li, H., Zhang, M., Liu, J., Sun, Z. e Tan, T. (2016b). Accurate iris segmentation in non-cooperative environments using fully convolutional networks. Em *International Conference on Biometrics*, páginas 1–8.

Liu, X., Bowyer, K. W. e Flynn, P. J. (2005). Experiments with an improved iris segmentation algorithm. Em *Workshop on Automatic Identification Advanced Technologies*, páginas 118–123. IEEE.

Lucio, D. R., Laroca, R., Severo, E., Britto Jr., A. S. e Menotti, D. (2018). Fully convolutional networks and generative adversarial networks applied to sclera segmentation. Em *IEEE International Conference on Biometrics: Theory, Applications and Systems*. CoRR: abs/1806.08722 `http://arxiv.org/abs/1806.08722`.

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

Marsico, M., Nappi, M. e Proença, H. (2017). Results from MICHE II -mobile iris challenge evaluation II. *Pattern Recognition Letters*, 91:3–10.

Minaee, S., Abdolrashidiy, A. e Wang, Y. (2016). An experimental study of deep convolutional features for iris recognition. Em *2016 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, páginas 1–6.

Nemenyi, P. (1963). *Distribution-free Multiple Comparisons*. Tese de doutorado, Princeton University.

Newman, M. E. J. (2005). Power laws, pareto distributions and zipf's law. *Contemporary Physics*, 46(5):323–351.

Nosaka, R., Suryanto, C. H. e Fukui, K. (2013). Rotation invariant co-occurrence among adjacent lbps. Em *International Workshop on Computer Vision*, páginas 15–25, Berlin, Heidelberg. Springer.

Oquab, M., Bottou, L., Laptev, I. e Sivic, J. (2014). Learning and transferring mid-level image representations using convolutional neural networks. Em *Conference on Computer Vision and Pattern Recognition*, páginas 1717–1724.

Othman, N., Dorizzi, B. e Garcia-Salicetti, S. (2016). OSIRIS: An open source iris recognition software. *Pat. Rec. Letters*, 82:124–131.

Park, W. e Chirikjian, G. S. (2007). Interconversion between truncated cartesian and polar expansions of images. *Transactions on Image Processing*, 16(8):1946–1955.

Phillips, P. J., Bowyer, K. W., Flynn, P. J., Liu, X. e Scruggs, W. T. (2008). The iris challenge evaluation 2005. Em *International Conference on Biometrics: Theory, Applications and Systems*, páginas 1–8.

Podder, P., Khan, T. Z., Khan, M. H., Rahman, M. M., Ahmed, R. e Rahman, M. S. (2015). An efficient iris segmentation model based on eyelids and eyelashes detection in iris recognition system. Em *International Conference on Computer Communication and Informatics*, páginas 1–7.

Powers, D. M. W. (2011). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness & correlation. *Journal of Machine Learning Technologies*, 2(1):37–63.

Proença, H. e Alexandre, L. A. (2005). Ubiris: A noisy iris image database. Em *International Conference on Image Analysis and Processing-ICIAP*, páginas 970–977, Berlin, Heidelberg. Springer.

Proença, H. e Alexandre, L. A. (2006). Iris segmentation methodology for non-cooperative recognition. *Proceedings-Vision, Image and Signal Processing*, 153(2):199–205.

Proença, H. e Alexandre, L. A. (2007). The NICE.I: Noisy iris challenge evaluation-part I. Em *International Conference on Biometrics: Theory, Applications, and Systems*, páginas 1–4. IEEE.

Proença, H. e Alexandre, L. A. (2010). Introduction to the special issue on the segmentation of visible wavelength iris images captured at-a-distance and on-the-move. *Image and Vision Computing*, 28(2):213–214.

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

Proença, H., Filipe, S., Santos, R., Oliveira, J. e Alexandre, L. A. (2010). The UBIRIS.v2: A database of visible wavelength images captured on-the-move and at-a-distance. *Trans. PAMI*, 32(8):1529–1535.

Pundlik, S. J., Woodard, D. L. e Birchfield, S. T. (2008). Non-ideal iris segmentation using graph cuts. Em *Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, páginas 1–6.

Raja, K. B., Raghavendra, R., Vemuri, V. K. e Busch, C. (2015). Smartphone based visible iris recognition using deep sparse filtering. *Pattern Recognition Letters*, 57:33–42.

Rankin, D. M., Scotney, B. W., Morrow, P. J., McDowell, D. R. e Pierscionek, B. K. (2010). Dynamic iris biometry: a technique for enhanced identification. *Biology and Medicine Central Research Notes*, 3(1):1–7.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C. e Fei-Fei, L. (2015). Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3):211–252.

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

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

Sequeira, A., Chen, L., Wild, P., Ferryman, J., Alonso-Fernandez, F., Raja, K. B., Raghavendra, R., Busch, C. e Bigun, J. (2016). Cross-Eyed-cross-spectral iris/periocular recognition database and competition. Em *International Conference of the Biometrics Special Interest Group*, páginas 1–5.

Serra, J. (1983). *Image Analysis and Mathematical Morphology*. Academic Press, Inc., Orlando, FL, USA.

Severo, E., Laroca, R., Bezerra, C. S., Zanlorensi, L. A., Weingaertner, D., Moreira, G. e Menotti, D. (2018). A benchmark for iris location and a deep learning detector evaluation. Em *2018 International Joint Conference on Neural Networks (IJCNN)*, páginas 1–7.

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

Shelhamer, E., Long, J. e Darrell, T. (2015). Fully convolutional networks for semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(4):640–651.

Simonyan, K. e Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556.

Sutra, G., Dorizzi, B., Garcia-Salicetti, S. e Othman, N. (2012). *A biometric reference system for iris OSIRIS version 4.1*.

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

Tan, T. e Sun, Z. (2002). CASIA-Iris-Databases. *Chinese Academy of Sciences Institute of Automation, Tech. Rep.* `http://english.ia.cas.cn/db/201610/t20161026_169399.html`.

Tan, T. e Sun, Z. (2005a). CASIA-IrisV3. *Chinese Academy of Sciences Institute of Automation, Tech. Rep.* `http://www.cbsr.ia.ac.cn/IrisDatabase.htm`.

Tan, T. e Sun, Z. (2005b). CASIA-IrisV4. *Chinese Academy of Sciences Institute of Automation, Tech. Rep.* `http://biometrics.idealtest.org/dbDetailForUser.do?id=4.htm`.

Tarawneh, A. S., Chetverikov, D. e Hassanat, A. B. (2018). Pilot comparative study of different deep features for palmprint identification in low-quality images. *CoRR*, abs/1804.04602.

Teichmann, M., Weber, M., Zoellner, M., Cipolla, R. e Urtasun, R. (2016). Multinet: Real-time joint semantic reasoning for autonomous driving. *arXiv preprint arXiv:1612.07695*.

Thoma, M. (2016). A survey of semantic segmentation. *CoRR*, abs/1602.06541.

Tian, Z., Liu, L. e Fei, B. (2017). Deep convolutional neural network for prostate MR segmentation. *Proc. SPIE*, 10135:101351–101351–6.

Turk, M. e Pentland, A. (1991). Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86.

Vera, D. F., Cadena, D. M. e Ramirez, J. M. (2015). Iris recognition algorithm on beaglebone black. Em *International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications*, volume 1, páginas 282–286.

Viola, P. e Jones, M. J. (2001). Rapid object detection using a boosted cascade of simple features. Em *Proceedings of the Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, páginas 511–518. IEEE.

Viola, P. e Jones, M. J. (2004). Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154.

Wildes, R. P. (1997). Iris recognition: an emerging biometric technology. *Proceedings*, 85(9):1348–1363.

Zanlorensi, L. A., Luz, E., Laroca, R., Britto Jr., A. S., Oliveira, L. S. e Menotti, D. (2018). The impact of preprocessing on deep representations for iris recognition on unconstrained environments. Em *31th Conference on Graphics, Patterns and Images (SIBGRAPI)*. CoRR: abs/1808.10032 `http://arxiv.org/abs/1808.10032`.

Zhao, Z. e Kumar, A. (2015). An accurate iris segmentation framework under relaxed imaging constraints using total variation model. Em *International Conference on Computer Vision*, páginas 3828–3836.

Zhou, Z., Du, Y., Thomas, N. L. e Delp III, E. J. (2010). Multimodal eye recognition. Em *SPIE Defense, Security, and Sensing*, páginas 770806–770806. International Society for Optics and Photonics.