



UNIVERSIDADE FEDERAL DE CAMPINA GRANDE
CENTRO DE ENGENHARIA ELÉTRICA E INFORMÁTICA
Coordenação de Pós-Graduação em Ciência da Computação

RM-net: Uma Rede Neural Convolutiva, com Função de
Custo no Domínio da Frequência, para Realce de Imagens de
Impressões Digitais Latentes

Nailson dos Santos Cunha

Tese submetida à Coordenação do Curso de Pós-Graduação em
Ciência da Computação da Universidade Federal de Campina
Grande - Campus I como parte dos requisitos necessários para
obtenção do grau de Doutor em Ciência da Computação.

Prof. Dr. Herman Martins Gomes
(Orientador)

Prof. Dr. Leonardo Vidal Batista
(Co-orientador)

Campina Grande, Paraíba, Brasil
Agosto, 2023

UNIVERSIDADE FEDERAL DE CAMPINA GRANDE
CENTRO DE ENGENHARIA ELÉTRICA E INFORMÁTICA
Coordenação de Pós-Graduação em Ciência da Computação

RM-net: Uma Rede Neural Convolucional, com Função de
Custo no Domínio da Frequência, para Realce de Imagens de
Impressões Digitais Latentes

Nailson dos Santos Cunha

Tese submetida à Coordenação do Curso de Pós-Graduação em
Ciência da Computação da Universidade Federal de Campina
Grande - Campus I como parte dos requisitos necessários para
obtenção do grau de Doutor em Ciência da Computação.

Área de concentração: Ciência da Computação

Linha de Pesquisa: Modelos Computacionais e Cognitivos

Prof. Dr. Herman Martins Gomes
(Orientador)

Prof. Dr. Leonardo Vidal Batista
(Co-orientador)

Campina Grande, Paraíba, Brasil
Agosto, 2023

C972r

Cunha, Nailson dos Santos.

RM-net: uma rede neural convolucional, com função de custo no domínio da frequência, para realce de imagens de impressões digitais latentes / Nailson dos Santos Cunha. - Campina Grande, 2023.

126 f. : il. color.

Tese (Doutorado em Ciência da Computação) - Universidade Federal de Campina Grande, Centro de Engenharia Elétrica e Informática, 2023.

"Orientação: Prof. Dr. Herman Martins Gomes, Prof. Dr. Leonardo Vidal Batista."

Referências.

1. Ciência da Computação. 2. Aprendizagem de Máquina. 3. Impressões Digitais Latentes. 4. Biometria. 5. Autocodificadores. 6. *Deep Learning*. 7. Redes Neurais Convolucionais. I. Gomes, Herman Martins. II. Batista, Leonardo Vidal. III. Título.

CDU 004(043)



MINISTÉRIO DA EDUCAÇÃO

UNIVERSIDADE FEDERAL DE CAMPINA GRANDE

POS-GRADUACAO EM CIENCIA DA COMPUTACAO

Rua Aprígio Veloso, 882, Edifício Telmo Silva de Araújo, Bloco CG1, - Bairro Universitário, Campina Grande/PB, CEP 58429-900

Telefone: 2101-1122 - (83) 2101-1123 - (83) 2101-1124

Site: <http://computacao.ufcg.edu.br> - E-mail: secretaria-copin@computacao.ufcg.edu.br / copin@copin.ufcg.edu.br

FOLHA DE ASSINATURA PARA TESES E DISSERTAÇÕES

NAILSON DOS SANTOS CUNHA

RM-net: UMA REDE NEURAL CONVOLUCIONAL COM FUNÇÃO DE CUSTO NO DOMÍNIO DA FREQUÊNCIA PARA REALCE DE IMAGENS DE IMPRESSÕES DIGITAIS LATENTES

Tese apresentada ao Programa de Pós-Graduação em Ciência da Computação como pré-requisito para obtenção do título de Doutor em Ciência da Computação.

Aprovada em: 21/08/2023

Prof. Dr. HERMAN MARTINS GOMES, UFCG, Orientador

Prof. Dr. LEONARDO VIDAL BATISTA, UFPB, Orientador

Prof. Dr. JOSEANA MACÊDO FECHINE RÉGIS DE ARAÚJO, UFCG, Examinadora Interna

Prof. Dr. CLÁUDIO DE SOUZA BAPTISTA, UFCG, Examinador Interno

Prof. Dr. TSANG ING REN, UFPE, Examinador Externo

Prof. Dr. JEFERSSON ALEX DOS SANTOS, UFMG, Examinador Externo



Documento assinado eletronicamente por **HERMAN MARTINS GOMES, PROFESSOR 3 GRAU**, em 22/08/2023, às 07:32, conforme horário oficial de Brasília, com fundamento no art. 8º, caput, da [Portaria SEI nº 002, de 25 de outubro de 2018](#).



Documento assinado eletronicamente por **JOSEANA MACEDO FECHINE, PROFESSOR(A) DO MAGISTERIO SUPERIOR**, em 22/08/2023, às 11:01, conforme horário oficial de Brasília, com fundamento no art. 8º, caput, da [Portaria SEI nº 002, de 25 de outubro de 2018](#).



Documento assinado eletronicamente por **Leonardo Vidal Batista, Usuário Externo**, em 22/08/2023, às 11:40, conforme horário oficial de Brasília, com fundamento no art. 8º, caput, da [Portaria SEI nº 002, de 25 de outubro de 2018](#).



Documento assinado eletronicamente por **CLAUDIO DE SOUZA BAPTISTA, PROFESSOR 3 GRAU**, em 22/08/2023, às 11:44, conforme horário oficial de Brasília, com fundamento no art. 8º, caput, da [Portaria SEI nº 002, de 25 de outubro de 2018](#).



A autenticidade deste documento pode ser conferida no site <https://sei.ufcg.edu.br/autenticidade>, informando o código verificador **3713646** e o código CRC **2F5C74A5**.

AGRADECIMENTOS

Agradeço a Deus pela força ao longo desta jornada acadêmica.

À minha família, meus pais, irmã e sobrinhos, obrigado por serem minha inspiração constante.

Aos meus orientadores, Prof. Herman e Prof. Leonardo, minha gratidão pela generosidade, paciência e inestimáveis ensinamentos.

Aos meus amigos Thiago, Igor, Pedro Paulo e Michael, que sempre acreditaram em mim. Em especial à Thiago que acompanhou todo o doutorado desde o início, e me ajudou de inúmeras formas.

Agradeço a João Janduy e à equipe da VSoft pela ajuda em algumas etapas importantes da pesquisa.

Aos demais membros da minha família e a todas as pessoas que, de alguma forma, contribuíram para esta trajetória, meu sincero agradecimento.

RESUMO

Impressões digitais desempenham um papel importante na verificação e autenticação biométrica de indivíduos. Existem diferentes tipos de impressões digitais, incluindo aquelas capturadas em ambientes controlados, usando tinta ou escâner, e as impressões digitais latentes, que são deixadas involuntariamente quando os dedos entram em contato com diversas superfícies. Procedimentos de realce são comumente aplicados para melhorar a qualidade das impressões digitais e facilitar a extração de características. No entanto, estudos revelaram que métodos de realce projetados para impressões digitais capturadas por tinta ou escâner não apresentam a mesma eficácia quando aplicados a impressões digitais latentes. Atualmente, métodos baseados em *Deep Learning* têm sido amplamente adotados no processamento de imagens. Entretanto, ao tentar reconstruir partes muito ruidosas das imagens, esses métodos tendem a preencher de forma errônea e não natural, ou até amplificar o ruído existente na imagem original, resultando em um agravamento do problema em algumas partes das imagens realçadas. Diante disso, o objetivo central desta pesquisa é desenvolver um método específico de realce para imagens de impressões digitais latentes, que seja capaz de preservar a estrutura das cristas enquanto busca mitigar ou resolver esse problema. O método desenvolvido é baseado em uma arquitetura codificador-decodificador convolucional, que tem como objetivo realizar o realce de imagens de impressões digitais latentes. Essa arquitetura é projetada para receber diretamente a imagem da impressão digital latente como entrada, sem a necessidade de pré-processamento, e gerar como saída uma versão realçada da imagem. Durante o treinamento do método, foi aplicada uma função de custo que realiza a comparação de imagens no domínio da frequência, resultando na redução do problema de reconstrução não natural das estruturas de cristas em partes da imagem original que apresentam muito ruído. Dois experimentos foram conduzidos para validar o método desenvolvido: um relacionado à correspondência latente-para-sensor e outro latente-para-latente. As bases de dados utilizadas para esses experimentos foram a MOLF e a IIIT-D *latent fingerprint database*. Os resultados obtidos demonstraram uma melhora nos testes de identificação e correspondência das impressões digitais no conjunto de testes, quando comparados aos cenários em que o realce não foi aplicado e também comparativamente a métodos encontrados na literatura.

Palavras-chave: Impressões digitais latentes, Biometria, Autocodificadores, *Deep Learning*, Redes neurais convolucionais

ABSTRACT

Fingerprint impressions play a crucial role in the verification and biometric authentication of individuals. There are different types of fingerprints, including those captured in controlled environments using ink or scanners, and latent fingerprints, which are unintentionally left behind when fingers touch various surfaces. Enhancement procedures are commonly applied to improve the quality of ridges and facilitate the extraction of minutiae in fingerprints. However, studies have revealed that enhancement methods designed for inked or scanned fingerprints are not equally effective when applied to latent fingerprints. Currently, deep learning-based methods have been widely adopted in image processing. However, when attempting to reconstruct highly noisy parts of the images, these methods tend to fill in erroneously and non-naturally, or even amplify the existing noise in the original image, resulting in a worsening of the problem in some parts of the enhanced images. In view of this, the main objective of this research is to develop a specific enhancement method for latent fingerprint images that is capable of preserving ridge structures while seeking to mitigate or resolve this issue. The developed method is based on a convolutional encoder-decoder architecture, aiming to perform enhancement of latent fingerprint images. This architecture is designed to directly take the latent fingerprint image as input, without the need for preprocessing, and generate an enhanced version of the image as output. During the training of the method, a cost function was applied, which compares images in the frequency domain, resulting in a reduction of the problem of non-natural reconstruction of ridge structures in parts of the original image that exhibit high noise. Two experiments were conducted to validate the developed method: one related to latent-to-sensor matching and another related to latent-to-latent matching. The MOLF and IIIT-D latent fingerprint databases were used for these experiments. The obtained results demonstrated an improvement in fingerprint identification and matching tests in the test set, when compared to scenarios where enhancement was not applied, as well as in comparison to methods found in the literature.

Keywords: Latent fingerprints, Biometrics, Autoencoders, Deep learning, Convolutional neural networks

LISTA DE FIGURAS

1	Representação da pele de cristas de fricção. A epiderme é parcialmente elevada da derme, expondo as papilas dérmicas. Fonte: adaptado de Jain et al. (2006).	23
2	Exemplos de impressões digitais reais. (a) é uma imagem de uma impressão digital latente, (b) é uma imagem de uma impressão digital rolada coletada usando tinta e (c) é uma imagem de uma impressão digital plana digitalizada via escâner. Fonte: adaptado de Qian et al. (2019).	24
3	Amostras de impressões digitais latentes obtidas da base de dados NIST SD27 na primeira linha e suas respectivas versões roladas na segunda linha. Fonte: adaptado de Li et al. (2017).	27
4	Impressão digital, cristas e vales. Fonte: adaptado de Maltoni et al. (2009a).	36
5	Impressão digital com destaque para o <i>core</i> (círculo) e <i>delta</i> (triângulo). Fonte: adaptado de Msiza et al. (2011).	37
6	Amostra dos diferentes tipos de minúcias. Fonte: adaptado de Fraser e Williams (2013).	38
7	Ilustração de uma rede neural de alimentação direta. Fonte: adaptado de Azzouni et al. (2017)	41
8	Ilustração da operação de convolução bidimensional. Fonte: adaptado de Goodfellow et al. (2016).	43
9	Exemplo das operações de <i>max pooling</i> and <i>average pooling</i> realizadas em um mapa de características de 4×4 com <i>pool_size</i> = 2×2 , <i>stride</i> = 2. Fonte: adaptado de Guissous (2019).	44
10	Gráfico da função ReLU. Fonte: Ghosh et al. (2020).	46
11	Gráfico da função <i>sigmoid</i> . Fonte: Ghosh et al. (2020).	46
12	Arquitetura de um autocodificador. Fonte: adaptado de Bank et al. (2020).	47
13	Exemplo de um tipo de bloco residual. Fonte: He et al. (2016).	48
14	<i>Ground truth</i> de uma impressão digital existente no banco de dados sintético de treinamento e sua respectiva representação no domínio da frequência obtida por meio da transformada de Fourier. Fonte: De autoria própria. . .	50
15	Arquitetura proposta na pesquisa de Svoboda et al. (2017). Fonte: Svoboda et al. (2017).	56
16	Arquitetura proposta na pesquisa de Li et al. (2017). Fonte: Li et al. (2017).	57

17	Arquitetura proposta na pesquisa de Qian et al. (2019). Fonte: Qian et al. (2019).	58
18	Arquitetura proposta na pesquisa de Wong e Lai (2020). Fonte: Wong e Lai (2020).	59
19	Arquitetura proposta na pesquisa de Liu e Qian (2020). Fonte: Liu e Qian (2020).	59
20	Arquitetura proposta e fluxo de treinamento da rede neural da pesquisa de Horapong et al. (2021). (a) é o <i>locally spectral autoencoder</i> que possui o aprendizado baseado em <i>patches</i> espectrais de tamanho 64×64 . (b) é chamado de <i>regionally spectral autoencoder</i> , que combina 9 <i>locally spectral autoencoders</i> para entradas e saídas. Fonte: Horapong et al. (2021).	61
21	Arquitetura proposta na pesquisa de Dabouei et al. (2018). a) rede Gerador que recebe a impressão digital latente como entrada e gera os mapas como saída. b) mapas gerados, que são concatenados com a impressão digital latente de entrada e fornecidos como condição para a rede Discriminador. c) mapas reais são extraídos das impressões digitais originais. d) rede siamesa profunda que foi treinada separadamente para receber os mapas gerados ou reais e fornecer as representações da rede PIDI para o discriminador. e) Discriminador que objetiva distinguir entre os mapas gerados e os mapas reais usando o ID combinado e as informações dos mapas gerados Fonte: Dabouei et al. (2018).	62
22	Arquitetura proposta na pesquisa de Joshi et al. (2019). (a) rede $\mathcal{E}nh_L$ treinada para fornecer como saída o realce de uma imagem de impressão digital latente recebida como entrada. (b) rede $\mathcal{D}is_E$ treinada para diferenciar se uma imagem de entrada é a imagem realçada de uma impressão digital latente real ou se é uma imagem gerada pela rede $\mathcal{E}nh_L$. Fonte: Joshi et al. (2019).	63
23	Arquitetura proposta na pesquisa de Liu et al. (2019). (a) Gerador que mapeia a impressão digital latente para a imagem realçada. (b) arquitetura cooperativa em que os módulos específicos de tarefas orientação, segmentação e qualidade, compartilham a mesma representação. (c) Discriminador que visa distinguir se uma amostra é uma imagem real ou gerada pelo Gerador. Fonte: Liu et al. (2019).	64

24	Arquitetura proposta na pesquisa de Huang et al. (2020). (a) Gerador treinado para fornecer como saída a imagem de impressão digital realçada e o mapa de orientação estimado. (b) Discriminador treinado para distinguir se um realce e mapa de orientação recebidos são de uma impressão digital latente real ou se foi gerado pelo Gerador. Fonte: Huang et al. (2020).	65
25	Arquitetura proposta na pesquisa de Xu et al. (2020). Fonte: Xu et al. (2020).	66
26	Arquitetura do FingerNet utilizada na pesquisa de Deshpande et al. (2020). Fonte: Tang et al. (2017).	67
27	Arquitetura da FingerGAN desenvolvida na pesquisa de Zhu et al. (2023). Fonte: Zhu et al. (2023).	67
28	Arquitetura da CR-GAN desenvolvida na pesquisa de Joshi et al. (2022). Fonte: Joshi et al. (2022).	68
29	Amostras de impressões digitais capturadas por sensor existentes no banco de dados MOLF: (a), (b) e (c) são impressões digitais de 500ppi capturadas pelos sensores Secugen, Lumidigm, e CrossMatch L-Scan Patrol, respectivamente. Fonte: adaptado de Sankaran et al. (2015b).	76
30	Amostras de impressões digitais latentes simultâneas da base de dados MOLF: (a) e (b) são conjuntos de impressões digitais latentes simultâneas da mão direita e mão esquerda de um indivíduo, respectivamente. Fonte: adaptado de Sankaran et al. (2015b).	77
31	Amostras de impressões digitais latentes existentes na base de dados IIIT-D <i>latent fingerprint database</i> . Fonte: De autoria própria.	79
32	Amostras de impressões digitais existentes na base de dados NIST SD300. Fonte: De autoria própria.	79
33	Amostra de impressão digital e seu respectivo <i>ground truth</i> , da esquerda para direita, gerados pelo programa SFinGe. Fonte: De autoria própria. . .	80
34	Etapas do processo para gerar o par impressão digital e <i>ground truth</i> . Fonte: De autoria própria.	80
35	Amostra das diferentes aplicações de <i>gaussian blur</i> nas imagens do conjunto de treinamento. A primeira imagem refere-se a impressão digital sintética original gerada pelo SFinGe. As demais imagens são a impressão digital original após a aplicação dos efeitos de borrado. Fonte: De autoria própria.	81

36	Amostra da aplicação de <i>gaussian noise</i> nas imagens do conjunto de treinamento. A primeira imagem refere-se a impressão digital sintética original gerada pelo SFinGe. A segunda imagem é a impressão digital original após a aplicação do <i>gaussian noise</i> . Fonte: De autoria própria.	82
37	Amostra da aplicação de ruído <i>Salt and Pepper</i> nas imagens do conjunto de treinamento. A primeira imagem refere-se a impressão digital sintética original gerada pelo SFinGe. A segunda imagem é a impressão digital original após a aplicação do ruído <i>Salt and Pepper</i> . Fonte: De autoria própria.	83
38	Amostra da aplicação de redução de contraste nas imagens do conjunto de treinamento. A primeira imagem refere-se a impressão digital sintética original gerada pelo SFinGe. A segunda imagem é a impressão digital original após a redução do contraste. Fonte: De autoria própria.	83
39	Amostra da aplicação de redução de brilho nas imagens do conjunto de treinamento. A primeira imagem refere-se a impressão digital sintética original gerada pelo SFinGe. A segunda imagem é a impressão digital original após a redução do brilho. Fonte: De autoria própria.	84
40	Amostra da combinação de ruídos e borrado nas imagens do conjunto de treinamento. A primeira imagem refere-se a impressão digital sintética original gerada pelo SFinGe. A segunda imagem é a impressão digital original após a combinação de <i>gaussian noise</i> e <i>gaussian blur</i> e a terceira imagem é após a combinação de ruído <i>Salt and Pepper</i> e <i>gaussian blur</i> . Fonte: De autoria própria.	85
41	Amostras de impressões digitais latentes sintéticas geradas a partir de imagens da base de dados NIST SD300. Fonte: De autoria própria.	86
42	Arquitetura da M-net. Fonte: Mehta e Sivaswamy (2017).	86
43	Arquitetura da rede neural RM-Net proposta, destacando em vermelho os blocos resnet adicionados como principal alteração na arquitetura da M-net original. Fonte: De autoria própria.	87
44	Configuração dos blocos resnet adicionados à arquitetura da M-net. Fonte: De autoria própria.	88
45	Curvas da evolução da função de custo durante o treinamento da RM-net. Fonte: De autoria própria.	91
46	Curvas da evolução do SSIM durante o treinamento da RM-net. Fonte: De autoria própria.	92

47	Amostras de reconstruções do conjunto sintético de testes e valores do SSIM obtidos em comparação aos <i>ground truths</i> . Fonte: De autoria própria. . . .	93
48	Curvas CMC comparando-se os métodos de realce. Fonte: De autoria própria.	96
49	Comparação entre as imagens de saída da RM-net treinada sob diferentes condições: (b) utilizando a função de custo MAE e (c) utilizando a função de custo no domínio da frequência. (a) é a impressão digital latente original. Fonte: De autoria própria.	97
50	Comparação das pontuações de qualidade das imagens com e sem a aplicação do realce, de acordo com o NFIQ. Fonte: De autoria própria. . . .	98
51	Imagens do realce que receberam pontuação 5, de acordo com o NFIQ. Fonte: De autoria própria.	99
52	Casos em que o realce das imagens obteve uma identificação precisa no teste de casamento. Isto é, a impressão digital latente realçada foi corretamente associada à sua correspondente na galeria de testes DB1. Fonte: De autoria própria.	100
53	Casos em que o modelo preenche áreas muito deterioradas da impressão digital original com estrutura de cristas imprecisa na reconstrução. Fonte: De autoria própria.	101
54	Amostras de impressões digitais latentes da base de dados III-D <i>latent fingerprint database</i> na primeira linha, juntamente com seus respectivos realces realizados pela RM-net na segunda linha.	103
55	Curvas CMC referentes aos <i>ranks</i> do experimento latente-para-latente, comparando as imagens raw e também os métodos estado-da-arte publicados na literatura.	104
56	Tela inicial do SFinGe. Fonte: De autoria própria.	120
57	SFinGe, telas <i>Step</i> 1 (a), em que é feita a escolha referente ao dedo que a imagem da impressão digital gerada representará, e <i>Step</i> 2 (b), em que é realizada a escolha da classe da impressão digital. Fonte: De autoria própria.	121
58	SFinGe, tela <i>Step</i> 3, onde é possível definir a densidade das cristas da impressão digital. Fonte: De autoria própria.	122
59	SFinGe, telas <i>Step</i> 4 (a), em que é definida a existência de arranhões ou cicatrizes na impressão digital, e <i>Step</i> 6 (b), onde pode ser definido se existem pontos de diferentes níveis de pressionamento do dedo. Fonte: De autoria própria.	122

- 60 SFinGe, telas *Step* 6. Opções de pressão do dedo em alguns pontos da impressão digital. Fonte: De autoria própria. 123
- 61 SFinGe, telas *Step* 8 (a), em que é possível selecionar diferentes níveis de ruídos nas cristas e nos vales da impressão digital, e *Step* 10 (b), onde é realizada a escolha do tipo de plano de fundo da impressão digital. Fonte: De autoria própria. 124

LISTA DE TABELAS

1	Acurácias obtidas no experimento latente-para-sensor utilizando o VeriFinger como extrator de minúcias e algoritmo de casamento.	95
2	Acurácias obtidas no experimento latente-para-sensor. Comparação dos resultados obtidos pelos modelos M-net padrão, RM-net treinada com a função de custo MAE e a RM-net treinada com a função de custo no domínio da frequência.	95
3	Acurácias do experimento latente-para-latente usando o VeriFinger como extrator de minúcias e algoritmo de casamento.	102
4	Acurácias <i>rank</i> -n obtidos no experimento latente-para-latente, comparando os resultados da rede neural treinada utilizando apenas a função de perda MAE e a função de perda no domínio da frequência.	103
5	Acurácias dos experimentos conduzidos considerando a adição de blocos residuais	126
6	Acurácias dos experimentos conduzidos considerando diferentes combinações das funções de custo	127

LISTA DE QUADROS

1	Descrição da pontuação de qualidade que pode ser atribuída pelo NFIQ para as impressões digitais	54
2	Resumo dos trabalhos relacionados	69
3	Resumo de informações conceituais sobre os trabalhos apresentados nesta seção.	70
3	Resumo de informações conceituais sobre os trabalhos apresentados nesta seção (continuação).	71
3	Resumo de informações conceituais sobre os trabalhos apresentados nesta seção (continuação).	72
4	Diferentes subconjuntos da base de dados MOLF, juntamente com o tipo de impressão digital, protocolo de captura e suas propriedades	78

GLOSSÁRIO

ACE-V	<i>Analysis, Comparison, Evaluation, and Verification</i>
Adam	<i>Adaptive Momentum Estimation</i>
AE	<i>Autoencoder</i>
AFIS	<i>Automated Fingerprint Identification System</i>
BN	<i>Batch Normalization</i>
CAE	<i>Convolutional Autoencoder</i>
cGAN	<i>Conditional Generative Adversarial Network</i>
CMC	<i>Cumulative Match Characteristic</i>
CNN	<i>Convolutional Neural Network</i>
DCT	<i>Discrete Cosine Transform</i>
DFT	<i>Discrete Fourier Transform</i>
DL	<i>Deep Learning</i>
FFT	<i>Fast Fourier Transform</i>
FN	Falso Negativo
FOMFE	<i>Fingerprint Orientation Model Based on 2D Fourier Expansion</i>
FP	Falso Positivo
GAN	<i>Generative Adversarial Network</i>
IA	Inteligência Artificial
IAFIS	<i>Integrated Automated Fingerprint Identification System</i>
MAE	<i>Mean Absolute Error</i>
MCC	<i>Minutia Cylinder Codes</i>
ML	<i>Machine Learning</i>
MLP	<i>Multilayer Perceptron</i>
MOLF	<i>Multisensor Optical And Latent Fingerprint</i>
MSE	<i>Mean Squared Error</i>
NBIS	<i>NIST Biometric Image Software</i>
NFIQ	<i>NIST Fingerprint Image Quality</i>
NIST	<i>National Institute of Standards and Technology</i>
NN	<i>Neural Network</i>

PGAN *Progressive Generative Adversarial Network*

PIDI *Perceptual ID Information*

ResNet *Residual Network*

SSIM *Structural Similarity Index Measure*

ReLU *Rectifier Linear Unit*

SFinGe *Synthetic Fingerprint Generator*

TVD *Total Variation Decomposition*

VN Verdadeiro Negativo

VP Verdadeiro Positivo

Conteúdo

1	Introdução	22
1.1	Motivação	22
1.2	Definição do problema	26
1.3	Objetivos	28
1.4	Contribuições da pesquisa	29
1.5	Publicações	30
1.6	Organização do documento	30
2	Fundamentação Teórica	32
2.1	Biometria	32
2.1.1	Biometria na ciência forense	35
2.1.2	Impressão digital	35
2.1.3	Impressão digital latente	37
2.2	Aprendizagem de máquina profunda	39
2.2.1	Redes neurais convolucionais	41
2.2.2	Autocodificadores	47
2.2.3	ResNet	48
2.3	Domínio espacial e domínio da frequência	48
2.3.1	Transformada de Fourier	49
2.3.2	Transformada discreta do cosseno	51
2.4	Métricas de desempenho	51
2.4.1	Acurácia	51
2.4.2	Índice de similaridade estrutural	53
2.4.3	<i>NIST fingerprint image quality</i>	53
2.5	Considerações finais sobre o capítulo	54
3	Pesquisas Relacionadas	55
3.1	Métodos baseados em redes convolucionais	55
3.2	Métodos baseados em redes generativas	60

3.3	Outras pesquisas relacionadas	69
3.4	Considerações finais sobre o capítulo	73
4	Materiais e Método	75
4.1	Bases de dados	75
4.1.1	<i>Multisensor Optical And Latent Fingerprint database</i>	75
4.1.2	<i>IIIT-D latent fingerprint database</i>	77
4.1.3	<i>NIST Special database 300</i>	77
4.1.4	Base de dados sintética <i>Ad hoc</i>	79
4.2	Arquitetura da rede neural	84
4.2.1	M-net	84
4.2.2	RM-net	85
4.3	Função de custo	88
4.3.1	Metodologia de treinamento	90
4.3.2	Parâmetros e hiperparâmetros	90
4.4	Considerações finais sobre o capítulo	90
5	Experimentos e Resultados	91
5.1	Treinamento da rede neural	91
5.2	Avaliação em um conjunto sintético de teste	92
5.3	Experimento latente-para-sensor	94
5.3.1	Qualidade e visualização do realce das imagens da base MOLF DB4	98
5.4	Experimento latente-para-latente	102
5.5	Considerações finais do capítulo	104
6	Considerações Finais	106
6.1	Considerações sobre a pesquisa proposta	106
6.2	Considerações sobre os resultados obtidos	107
6.3	Propostas de trabalhos futuros	109
A	Captura das Imagens de Impressões Digitais Geradas pelo SFinGe	120

B	Detalhes de Implementação da RM-net	125
C	Experimentos Preliminares Auxiliares	126
C.1	Experimentos para definição dos blocos residuais	126
C.2	Experimentos para definição da função de custo	127

1 Introdução

O objetivo deste capítulo é apresentar o problema de pesquisa abordado nesta tese. Inicialmente, serão apresentadas as motivações para a pesquisa. Em seguida, o problema a ser estudado é exposto e delimitado, seguido pela definição dos objetivos gerais e específicos almejados com a pesquisa. Por fim, descreve-se a estrutura da tese e a organização subsequente deste documento.

1.1 Motivação

Reconhecimento biométrico, ou simplesmente biometria, diz respeito à utilização de traços físicos ou comportamentais para a identificação ou verificação e autenticação de indivíduos (Aithal e Karani, 2017). A biometria está se tornando um componente essencial para a criação de soluções eficazes de identificação pessoal porque as características biométricas são únicas e individuais, representando intrinsecamente a identidade corporal do indivíduo (Maltoni et al., 2009a). Impressão digital, face, veias das mãos e dos dedos, íris e retina são exemplos de traços que podem ser utilizados para autenticação biométrica de pessoas (Chavan et al., 2015).

A pele dos dedos humanos, conhecida como pele de cristas de fricção (*friction ridge skin*), é composta por cristas e sulcos que formam as impressões digitais. Essa pele é constituída pela derme (camada interna) e pela epiderme (camada externa). As cristas emergem na epiderme, aumentando a fricção entre a pele e a superfície de contato. Cada crista de fricção é composta por unidades de crista únicas, formadas aleatoriamente, que contêm poros. Essas unidades de crista variam em forma, tamanho, alinhamento e fusão com as unidades adjacentes, tornando as impressões digitais únicas para cada pessoa (Jain et al., 2000, 2006).

Os poros, aberturas das glândulas sudoríparas subcutâneas localizadas na epiderme, penetram na derme. Durante o desenvolvimento, os poros são estabilizados nas cristas antes da conclusão do processo de formação da epiderme e derme, tornando-os imutáveis. Cada unidade de crista contém uma glândula sudorípara, e os poros são frequentemente distribuídos uniformemente ao longo das cristas. A distância entre os poros geralmente está relacionada à largura média da crista. Os poros podem ser visualizados em imagens de impressões digitais como abertos ou fechados, dependendo de sua atividade de transpiração. Essas características únicas dos poros e das unidades de crista contribuem para a singularidade das impressões digitais (Jain et al., 2006). Na Figura 1 é mostrada uma representação da estrutura da pele de fricção, evidenciando suas camadas, cristas, poros, glândulas e dutos sudoríparos.

A identificação e a verificação via impressão digital são os processos de autentica-

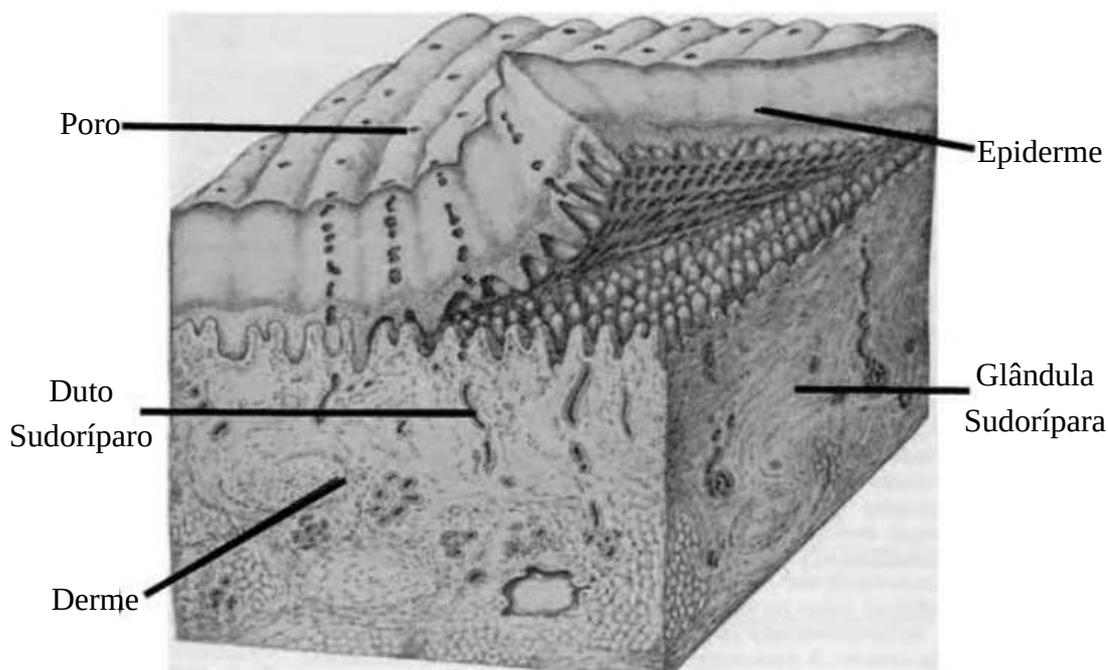


Figura 1: Representação da pele de cristas de fricção. A epiderme é parcialmente elevada da derme, expondo as papilas dérmicas. Fonte: adaptado de Jain et al. (2006).

ção biométrica mais comumente utilizados devido à sua facilidade de uso e unicidade, e também ao fato de que a estrutura e as características do padrão permanecem inalteradas ao longo da vida de uma pessoa (Chavan et al., 2015).

Com base nos tipos de aquisição, as impressões digitais podem ser categorizadas em três grupos principais: impressões digitais coletadas com tinta, impressões digitais escaneadas e impressões digitais latentes. As impressões digitais coletadas com tinta e escaneadas são geralmente obtidas em ambientes controlados, permitindo uma coleta mais precisa e monitoramento da qualidade da amostra. Nessas situações, é possível obter impressões digitais mais fidedignas e de melhor definição. Por exemplo, um agente, ou o próprio processo de aquisição do sistema, pode monitorar a qualidade das amostras e orientar os indivíduos a deixar as impressões digitais adequadas (Dabouei et al., 2018).

A técnica de aquisição em ambientes controlados também pode variar. As impressões digitais podem ser coletadas de forma plana ou rolada. As impressões digitais planas são obtidas pressionando-se o dedo contra uma superfície plana, mantendo-o paralelo à superfície de coleta. Já as impressões digitais roladas são obtidas pressionando-se o dedo contra uma superfície de captura e rolando-o de uma extremidade a outra ao longo da superfície, a fim de coletar uma impressão digital completa.

Por outro lado, impressões digitais latentes são aquelas depositadas involuntariamente quando o suor, aminoácidos, proteínas e secreções naturais presentes na superfície da pele entram em contato com uma superfície externa. Essas impressões digitais ge-

ralmente não são diretamente visíveis aos olhos e necessitam de procedimentos especiais para que possam tornar-se visíveis e posteriormente coletadas. Normalmente, uma impressão digital latente é uma imagem que possui um ruído de fundo, com áreas ausentes e, portanto, contendo uma quantidade menor de informações de cristas e vales em comparação com as impressões digitais coletadas com tinta ou por meio de escâner (Sankaran et al., 2014). Na Figura 2 é exibido um exemplo de uma impressão digital latente, uma impressão digital rolada coletada por tinta e uma impressão digital plana digitalizada via escâner.

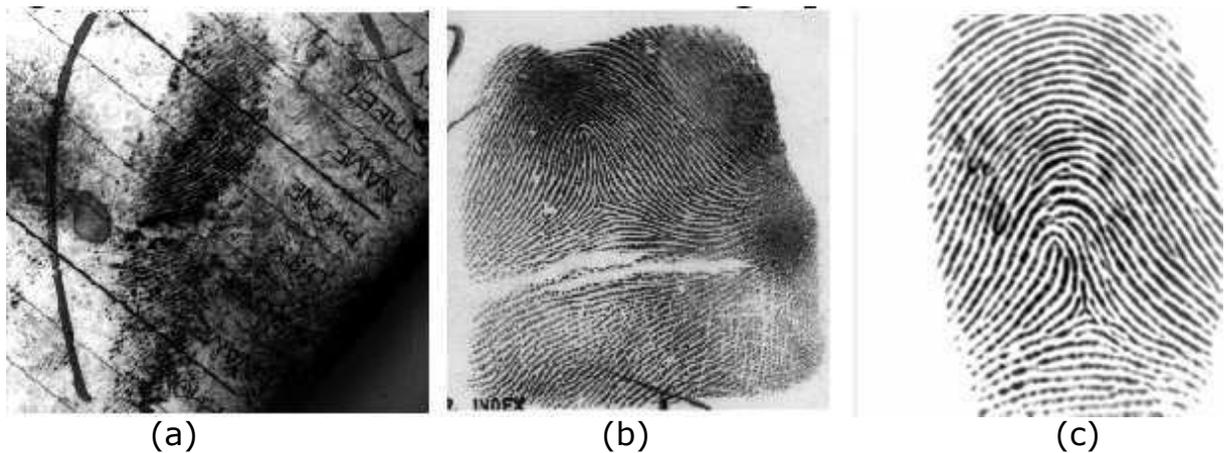


Figura 2: Exemplos de impressões digitais reais. (a) é uma imagem de uma impressão digital latente, (b) é uma imagem de uma impressão digital rolada coletada usando tinta e (c) é uma imagem de uma impressão digital plana digitalizada via escâner. Fonte: adaptado de Qian et al. (2019).

Na ciência forense, a impressão digital é uma ferramenta importante para identificar vítimas e suspeitos em cenas de crimes (Vasanth et al., 2019). Desde que as impressões digitais latentes foram apresentadas pela primeira vez como evidência para condenar um suspeito na Argentina em 1893, elas se tornaram uma das fontes de evidência mais amplamente utilizadas nas agências policiais e perícias criminalísticas em todo o mundo (Hawthorne, 2017).

Em geral, características encontradas em uma impressão digital latente coletada de uma cena de crime são anotadas manualmente por especialistas forenses. Logo após, um sistema integrado de identificação automatizada de impressão digital (IAFIS, do inglês *Integrated Automated Fingerprint Identification System*) compara as anotações da impressão digital latente com um banco de dados de referência e fornece a lista com as k correspondências prováveis, sendo os mais comuns $k = 50$ ou $k = 100$. A lista é então verificada manualmente por um especialista forense para determinar se existe individualização (identificação ou correspondência) (Houck, 2006).

Na prática, este procedimento de análise, comparação, avaliação e verificação (ACE-V, do inglês *Analysis, Comparison, Evaluation, and Verification*) de impressões

digitais latentes é um processo muito exigente e demorado. Com o intuito de tornar a identificação mais eficiente, os especialistas forenses tendem a restringir a população com a qual são realizadas as comparações (por exemplo, selecionando suspeitos apontados por testemunhas ou outras evidências). Entretanto, isso pode reduzir a probabilidade de identificar efetivamente um indivíduo e tornar o processo geral menos confiável (Jain e Feng, 2010).

Portanto, a fim de reduzir erros de individualização e tornar todo o processo mais confiável, foi criado o conceito de “Sistemas *Lights-Out*” para correspondência de impressões digitais latentes (Dvornychenko e Garris, 2006). Um Sistema *Lights-Out* tem como característica principal ser um processo de identificação totalmente automático, sem necessidade de intervenção humana. Esse sistema é capaz de extrair automaticamente recursos das impressões digitais latentes de consulta e compará-las com uma galeria de impressões digitais, a fim de obter um conjunto de possíveis correspondências. No entanto, atualmente, devido às limitações dos algoritmos disponíveis, o que é mais comumente utilizado, especialmente para impressões digitais latentes, são os “Sistemas *Semi-Lights-Out*”. Estes são sistemas que permitem alguma intervenção humana durante a extração de características (Singh e Dimri, 2020).

Embora muito progresso tenha acontecido no aprimoramento da velocidade e precisão dos sistemas de identificação automática de impressões digitais (AFIS, do inglês *Automated Fingerprint Identification System*), esses sistemas costumam funcionar muito bem em cenários onde a correspondência é realizada entre imagens de impressão digital planas ou roladas, porém não entregam um bom desempenho de classificação quando se trata de impressões digitais latentes (Zhang et al., 2013).

Com a finalidade de melhorar as taxas de identificação de impressões digitais latentes e permitir uma extração mais confiável de características, as imagens precisam passar por um processo de realce para melhorar a sua qualidade antes de serem fornecidas a um AFIS. O realce da impressão digital pode ajudar a melhorar a definição da estrutura de cristas, recuperar regiões corrompidas, remover o ruído estruturado e aumentar o contraste das cristas e vales (Feng et al., 2013).

Nos últimos anos, técnicas de Aprendizagem Profunda (DL, do inglês *Deep Learning*) têm sido amplamente estudadas e aplicadas com sucesso a problemas relacionados ao processamento de imagens e visão computacional, especialmente por intermédio da utilização de redes neurais convolucionais (CNN, do inglês *Convolutional Neural Network*) (Qian et al., 2019).

Nesta tese, é apresentada uma abordagem para o realce de imagens de impressões digitais latentes por meio do desenvolvimento da rede neural convolucional denominada M-net Residual, ou simplesmente RM-net. Essa arquitetura, baseada na M-net (Mehta

e Sivaswamy, 2017), foi adaptada para se adequar ao contexto específico de realce de impressões digitais latentes. Um dos principais aprimoramentos realizados foi a adição de blocos residuais à arquitetura da rede juntamente com a adaptação de uma função de custo que realiza a comparação de imagens no domínio da frequência.

Além da proposta de rede neural, também foi necessário criar uma base de dados sintética para o treinamento do modelo, uma vez que ainda não existe uma base de dados publicamente disponível contendo imagens reais de impressões digitais latentes com seus correspondentes *ground truths*. Para avaliar o desempenho da RM-net, foram conduzidos experimentos que demonstraram resultados promissores e competitivos em relação aos principais estudos encontrados na literatura.

1.2 Definição do problema

O principal desafio na identificação de impressões digitais latentes é a dificuldade de extrair informações devido à baixa qualidade das amostras por conta das diferentes e imprevisíveis condições em que são adquiridas. Impressões digitais roladas e planas são adquiridas sob supervisão cuidadosa, enquanto as impressões latentes são coletadas da superfície de objetos, em cenas de crime, por exemplo, e resultam em imagens com a presença de ruído e pouca informação da estrutura de cristas. Isso torna mais difícil o trabalho de sistemas de identificação em estabelecer uma ligação confiável entre as impressões parciais obtidas com as impressões digitais previamente registradas em um banco de dados (Deshpande et al., 2020). Na Figura 3 é visualizado um exemplo da diferença entre impressões digitais latentes e impressões digitais coletadas em ambientes com um maior controle de qualidade na coleta da amostra.

Extrair automaticamente características confiáveis de impressões digitais latentes, especialmente com baixa qualidade, ainda é uma tarefa complexa. Na prática, em muitos casos, o que acontece é a marcação manual de várias características, como região de interesse, singularidades e minúcias. O objetivo das técnicas de realce de impressões digitais é tornar mais evidente a estrutura de cristas e remover ruídos indesejados da impressão digital para auxiliar os sistemas nas etapas de análise e de extração de minúcias.

Estudos, como os conduzidos por Ulery et al. (2011, 2012), mostram preocupações com a influência do fator humano quanto à confiabilidade das informações extraídas. Um estudo reportado pelo *National Institute of Standards and Technology* (NIST), na pesquisa de Indovina et al. (2012), mostrou que a precisão de um sistema de casamento de impressões digitais latentes é muito afetada pela precisão da marcação do examinador, especialmente quando a própria imagem latente não está disponível para o sistema. Dror et al. (2012) também identificaram que quando o tempo de comparação é limitado, os examinadores das impressões digitais latentes são mais propensos a tomar uma decisão

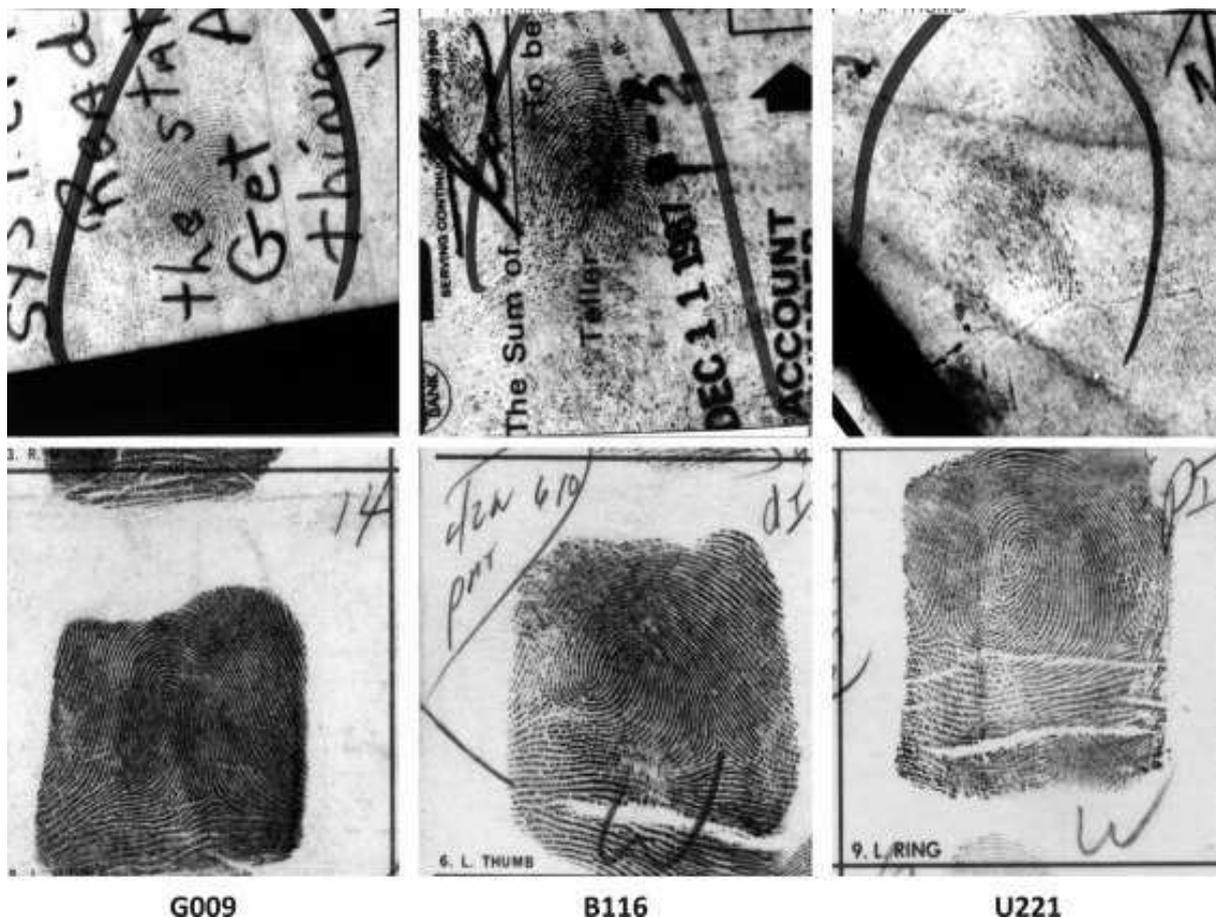


Figura 3: Amostras de impressões digitais latentes obtidas da base de dados NIST SD27 na primeira linha e suas respectivas versões roladas na segunda linha. Fonte: adaptado de Li et al. (2017).

de correspondência inconclusiva entre uma impressão latente e sua correspondente.

Um AFIS, seja ele para correspondência de impressões roladas, planas ou latentes, normalmente contém vários módulos, incluindo segmentação de região de interesse, realce, extração de minúcias e casamento. Realçar estruturas de cristas e vales e remover ruídos na imagem da impressão digital são essenciais para extrair minúcias precisas, estimar a orientação e o mapa de frequência das cristas (Cao et al., 2014).

Conforme explanado anteriormente, métodos baseados em DL vêm sendo estudados como opção para problemas que envolvem processamento de imagens. Entretanto, o treinamento de modelos de DL para realce de imagens de impressões digitais latentes normalmente se baseia na comparação das informações de cristas reconstruídas com uma saída esperada (*ground truth*). Embora essa abordagem possa fornecer resultados úteis em áreas com informações claras da estrutura de cristas, isso pode gerar padrões de cristas errôneos e artificiais em áreas com distorções severas, podendo assim resultar na criação de falsas minúcias, o que pode reduzir a precisão do processo de identificação.

É importante ressaltar que esse desafio é agravado pela escassez de bases de dados

de impressões digitais latentes reais acompanhadas de seus *ground truths*. Até o momento atual, não existe uma base de dados disponível com imagens de impressões digitais latentes reais e suas correspondentes informações corretas de cristas. Como resultado, os métodos de realce de impressões digitais latentes precisam ser treinados utilizando bases de dados sintéticas, o que introduz um desafio adicional, que é a criação dos dados que representem as características encontradas em impressões digitais latentes reais.

No campo da pesquisa em visão computacional, a tradução de imagem para imagem é uma tarefa na qual uma rede neural aprende a mapear imagens de um domínio para outro. Essa tarefa é particularmente desafiadora ao lidar com imagens de baixa qualidade, como aquelas com pouca iluminação ou esboços, porque os mapeamentos resultantes podem conter artefatos indesejados, decorrentes da amplificação de ruídos existentes no domínio original. Para atenuar esse problema, técnicas que analisam a imagem no domínio da frequência são frequentemente empregadas para remover esses artefatos.

Métodos clássicos de realce de impressões digitais também são comumente baseados em periodicidade da estrutura de cristas e podem enfrentar dificuldades similares ao lidar com imagens extremamente ruidosas. A presença de ruídos pode comprometer a precisão e a qualidade dos realces obtidos, resultando também em artefatos indesejados. Isso motiva ainda mais a investigação de como incorporar métodos de processamento de imagens no domínio da frequência em métodos de aprendizagem profunda.

Neste contexto, o problema abordado nesta de tese é o realce de imagens de impressões digitais latentes, partindo das seguintes questões:

- Q1.** Como melhorar aprendizado da estrutura de crista em modelos de aprendizagem profunda para o realce de imagens de impressões digitais latentes?
- Q2.** Técnicas de processamento de imagens no domínio da frequência usadas em problemas de tradução de imagem para imagem de contexto geral também apresentam contribuição quando aplicadas no contexto de realce de impressões digitais latentes?
- Q3.** Como reduzir a criação de padrões de cristas errôneos e artificiais em áreas em que existem distorções severas nas imagens originais?
- Q4.** Qual o impacto e resultados de incorporar técnicas de processamento de imagens no domínio da frequência na função de custo de um método de aprendizagem profunda?

1.3 Objetivos

O objetivo central e escopo desta pesquisa é a concepção e desenvolvimento de um método ponto-a-ponto para o realce de imagens de impressões digitais latentes. Isto é,

um método que recebe como entrada a imagem capturada da impressão digital latente, sem a necessidade de pré-processamento aplicado, e produz como resultado uma versão realçada da estrutura de cristas da imagem. O método também visa reduzir a reconstrução de áreas errôneas e criação de falsas minúcias nas imagens realçadas, usando técnicas de processamento de imagens no domínio da frequência. Dessa forma, os métodos de casamento de impressões digitais podem utilizar esse realce para melhorar a identificação e correspondência de impressões digitais latentes. Os objetivos específicos são:

- Criar uma base de dados sintética, que simule distorções, deteriorações e ruídos existentes em imagens de impressões digitais latentes reais, para a realização do treinamento de uma rede neural;
- Desenvolver um método baseado em uma arquitetura de codificador-decodificador que realize a tarefa de receber como entrada uma imagem de impressão digital latente e produzir como saída uma versão realçada da estrutura de cristas da impressão digital;
- Fazer o uso de técnicas de processamento de imagem no domínio da frequência para tratar problemas conhecidos, resultantes da reconstrução de áreas das imagens com a ocorrência de muito ruído;
- Aplicar o realce em bases de dados públicas de imagens de impressões digitais latentes reais disponíveis;
- Realizar testes de casamento das impressões digitais realçadas, utilizando os principais métodos de casamento relatados na literatura, e fazer a comparação dos resultados com outros métodos de realce publicados.

1.4 Contribuições da pesquisa

Nesta pesquisa, foi desenvolvido um método que integra técnicas de processamento de imagens no domínio da frequência a um método de aprendizagem profunda para a realização do treinamento de um modelo de rede neural que visa cumprir a tarefa de realce de imagens de impressões digitais latentes. Dentro deste contexto, esta pesquisa apresenta as seguintes contribuições:

- **Adaptação da arquitetura da rede neural desenvolvida:** nesta pesquisa foi desenvolvida uma arquitetura de rede neural que foi chamada de RM-net que adota uma estratégia de aprendizagem residual. A RM-net é uma adaptação de um modelo conhecido na literatura como M-net.

- **Metodologia de criação de uma base de dados de impressões digitais latentes sintéticas:** para realização do treinamento da rede neural desenvolvida, foi necessário criar uma base, que simule os desafios existentes em impressões digitais latentes reais. Dessa forma, foram realizadas a seleção e criada uma metodologia de adição de distorções e ruídos em impressões digitais criadas por um programa especializado e também em impressões digitais reais de uma base pública disponível.
- **Adaptação de uma função de custo que realiza a comparação entre duas imagens no domínio da frequência:** para que durante o processo de treinamento, a rede neural desenvolvida aprenda não apenas a reconstruir uma versão realçada da estrutura de cristas da impressão digital original, mas também aprenda a reduzir problemas decorrentes da tentativa de reconstrução de áreas com a presença de elevados níveis de ruído, foi adicionada uma função de custo que realiza correções nessas áreas baseando-se em comparações realizadas no domínio da frequência.

1.5 Publicações

O desenvolvimento desta pesquisa resultou na seguinte publicação:

- Cunha, N. D. S., Gomes, H. M., & Batista, L. V. (2022, October). Residual M-net with Frequency-Domain Loss Function for Latent Fingerprint Enhancement. In 2022 35th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI) (Vol. 1, pp. 198-203). IEEE;

Uma versão estendida do artigo, intitulado *Extended Evaluation of Latent Fingerprint Enhancement using Residual M-net with a Frequency-Domain Loss Function*, foi submetida, a convite, ao periódico *Springer Computational Statistics*, por ter sido classificada entre os melhores artigos do evento Sibgrapi 2022.

No primeiro artigo, foi apresentado um método de realce de imagens de impressões digitais latentes que se baseou na adaptação da arquitetura de uma M-net. Esse aprimoramento envolveu a incorporação de blocos residuais e a utilização de uma função de custo no domínio da frequência. Além disso, foram realizados experimentos de casamento de impressões digitais latentes aprimoradas pelo método proposto, utilizando imagens de sensores provenientes de uma base de dados de teste. A versão estendida desse artigo expandiu os experimentos para uma nova base de dados de teste, permitindo a comparação direta entre impressões digitais latentes aprimoradas pelo método proposto.

1.6 Organização do documento

A estrutura e organização dos demais capítulos desta tese são apresentadas a seguir. No Capítulo 2 são fornecidos os fundamentos dos principais conceitos necessários para

uma melhor compreensão deste trabalho de pesquisa. No Capítulo 3, são apresentados os principais estudos encontrados na literatura que abordam o tema de realce de imagens de impressões digitais latentes. No Capítulo 4, é apresentada uma descrição detalhada do método desenvolvido e dos materiais utilizados nesta pesquisa. Os resultados obtidos, juntamente com uma comparação com estudos publicados disponíveis, são apresentados no Capítulo 5. Por fim, são discutidas algumas considerações finais sobre o método e propostas de trabalhos futuros no Capítulo 6.

2 Fundamentação Teórica

Neste capítulo são abordados conceitos importantes para o desenvolvimento do método de realce de impressões digitais latentes. Inicialmente, são discutidos os conceitos relacionados à biometria e sua aplicação na ciência forense, destacando-se a importância do uso de impressões digitais como uma das principais fontes de evidência para a identificação de indivíduos. Em seguida, são apresentadas as características das impressões digitais latentes e os desafios associados ao seu realce, como a baixa qualidade das imagens e a falta de contraste entre as regiões da imagem.

Também são apresentados alguns princípios básicos do *Deep Learning* e suas aplicações em problemas de processamento de imagens, com foco nas redes neurais convolucionais e autocodificadores. Logo após, são abordados os conceitos de processamento de imagens no domínio do espaço e no domínio da frequência e as técnicas aplicadas para a implementação da função de custo no domínio da frequência utilizada para o treinamento da rede neural. Por último, são apresentadas as métricas para avaliação do desempenho da RM-net, sua eficácia no realce de impressões digitais latentes e sua contribuição na melhoria dos resultados dos algoritmos de identificação.

2.1 Biometria

Atualmente, o gerenciamento de identidades é parte crítica em muitos sistemas. Alguns exemplos de aplicações são regulação de passagens de fronteiras internacionais, restrição de acesso físico a instalações importantes, controle de acesso lógico a recursos e informações compartilhados, realização de transações financeiras remotas ou distribuição de benefícios de bem-estar social. Com o grande aumento de serviços que podem ser acessados remotamente também aumentou o risco de roubo de identidade, isso eleva a preocupação com segurança e mostra a necessidade do desenvolvimento de sistemas confiáveis de gerenciamento de identidade (Jain et al., 2022).

Uma tarefa fundamental na gestão de identidade é estabelecer a associação entre um indivíduo e sua identidade pessoal para, quando necessário, determinar ou verificar a reivindicação de identificação. Um indivíduo pode ser reconhecido com base nos seguintes três métodos básicos: (a) o que ele sabe, (b) o que ele possui extrinsecamente e (c) quem ele é intrinsecamente. O primeiro método se baseia no fato de existir uma informação secreta de conhecimento exclusivo do indivíduo, como, por exemplo, uma senha ou número de identificação pessoal. O segundo método pressupõe que há de posse do indivíduo um *token* extrínseco como, por exemplo, um cartão ou outro documento de identificação pessoal. O terceiro método estabelece a identidade da pessoa com base em traços físicos ou comportamentais inerentes, isto é chamado de reconhecimento biométrico (Jain et al., 2011).

O reconhecimento biométrico, ou simplesmente biometria, oferece uma solução mais natural e confiável para alguns problemas que podem existir nos métodos de gestão de identidade baseados em conhecimento e nos baseados em *tokens*. Por exemplo, um indivíduo pode negar o uso de um sistema ou serviço alegando que sua senha ou seu documento de identificação pessoal podem ter sido roubados. Ou ainda, indivíduos também podem ocultar sua verdadeira identidade apresentando documentos de identificação falsos ou duplicados. Como os identificadores biométricos são baseados em características inerentes a um indivíduo, fica mais difícil manipular, compartilhar ou esquecer esses traços. Assim, os traços biométricos constituem um vínculo forte e razoavelmente permanente entre uma pessoa e sua identidade.

Como exemplos de características físicas podem ser citadas a impressão digital, face, DNA, orelha, íris, retina e geometria da mão, sendo estes associados com a forma ou medidas do corpo humano. Entre as características comportamentais estão inclusas a assinatura, voz e forma de andar, sendo estas relacionadas com o comportamento ou medidas dinâmicas de um indivíduo (Sabhanayagam et al., 2018).

Cada traço biométrico possui seus prós e contras e, portanto, a escolha da utilização de uma característica biométrica para uma determinada aplicação depende de uma série de questões além de seu desempenho na correspondência. Jain et al. (1999) citam sete fatores que determinam se uma característica física ou comportamental é adequada para ser utilizada em uma aplicação biométrica. São elas:

1. **Universalidade:** todo indivíduo que será verificado pelo sistema deve possuir o traço;
2. **Unicidade:** o traço biométrico dado deve ser suficientemente diferente entre os indivíduos que compõem a população;
3. **Permanência:** o traço biométrico deve ser invariável em um longo período de tempo;
4. **Mensurabilidade:** a aquisição do traço biométrico não deve causar inconvenientes ao indivíduo e os dados coletados devem ser passíveis de processamento possibilitando extrair conjuntos de características representativas;
5. **Desempenho:** é importante que a taxa de precisão do sistema atenda a requisitos estabelecidos, considerando os recursos disponíveis para assegurar um desempenho adequado;
6. **Aceitabilidade:** os indivíduos que formarão a população que utilizará a aplicação biométrica devem estar dispostos a fornecer sua característica biométrica ao sistema;

7. **Circunvenção:** refere-se à facilidade com que o traço de um indivíduo pode ser imitado usando artefatos, no caso de traços físicos, e mimetismo, no caso de traços comportamentais.

É importante dizer que não é esperado que uma única biometria atenda efetivamente a todos os requisitos impostos por todas as aplicações. Isto é, a relevância de uma biometria específica para uma aplicação é estabelecida dependendo da natureza e dos requisitos da aplicação e das propriedades da característica biométrica (Jain et al., 2011).

Um sistema biométrico é uma aplicação de reconhecimento de padrões que adquire dados biométricos de um indivíduo, extrai um conjunto de características dos dados capturados, compara esse conjunto com outros conjuntos de características armazenados em um banco de dados e executa uma ação com base no resultado da comparação (Jain et al., 2007; Sabhanayagam et al., 2018). De uma forma geral, esses sistemas possuem quatro módulos principais:

1. **Módulo sensor ou aquisição de imagem:** Módulo de captura dos dados biométricos de um indivíduo na forma de vídeo, áudio, imagem ou algum outro sinal;
2. **Módulo de avaliação de qualidade e extração de características:** Módulo onde os dados biométricos adquiridos passam por uma avaliação de qualidade das amostras para determinar se estão em um patamar de qualidade adequado para realização do processamento posterior de extração de características;
3. **Módulo de correspondência e tomada de decisão:** As características extraídas são comparadas com os modelos armazenados e são geradas pontuações de correspondência que são calculadas para definir o nível de semelhança entre duas amostras biométricas. A pontuação computada é usada para validar uma identidade reivindicada ou fornecer uma classificação das identidades registradas para identificar um indivíduo;
4. **Módulo de banco de dados:** um repositório de informações biométricas cadastradas de usuários que armazena vários modelos de características de indivíduos.

Um sistema biométrico pode operar tanto no modo de verificação quanto de identificação, dependendo do contexto da aplicação. Um sistema de identificação verifica se um indivíduo é alguém conhecido do sistema sem reivindicar explicitamente uma identidade, sendo isto uma comparação 1 para N. Na verificação, é reivindicada uma identidade e o sistema verifica se a reivindicação é genuína, ou seja, se o indivíduo é quem ele diz ser. Nesse cenário, a consulta é comparada apenas ao modelo correspondente à identidade reivindicada, isto é, uma correspondência um para um.

Maltoni et al. (2009a) afirmam que na impressão digital existe equilíbrio entre todas as propriedades desejáveis em um traço biométrico. Além disso, os sistemas biométricos de reconhecimento de impressão digital atuais estão cada vez melhores em questão de desempenho de identificação e os escâneres de impressão digital tornaram-se compactos e acessíveis. Isso ajudou a tornar o reconhecimento de impressão digital uma das tecnologias biométricas mais maduras e adequadas para um grande número de aplicações de reconhecimento.

2.1.1 Biometria na ciência forense

Nas últimas décadas, sistemas biométricos passaram a ser adotados em vários domínios de aplicação e a tecnologia biométrica está criando um impacto significativo na sociedade. Por exemplo, a biometria passou a desempenhar um papel crítico em sistemas de aplicação da lei, servindo tanto como ferramenta de investigação para restringir a lista de suspeitos quanto como evidência forense em um tribunal (Jain e Ross, 2015).

Diferentemente da utilização típica de sistemas biométricos, na ciência forense, a tarefa fundamental é a comparação de material de fonte desconhecida com material de origem conhecida. Normalmente, existem dois cenários comuns: um em que o material desconhecido pode ser recolhido em condições controladas e outro em que o material pode ser recolhido em situações não controladas ou imprevisíveis, como, por exemplo, marcas recuperadas de cenas de crime ou material de um cadáver. Materiais coletados em condições irrestritas afetam substancialmente o desempenho dos sistemas, pois podem apresentar vários fatores de complexidade além das condições de aquisição (Tistarelli e Champod, 2017).

Sistemas biométricos forenses, na prática, são usados como dispositivos de classificação em que não existe mecanismo de decisão embutido sobre a veracidade da identificação. São ferramentas de classificação que permitem, com uma taxa média conhecida de eficácia, apresentar ao usuário do sistema uma lista contendo potenciais candidatos para uma consulta solicitada. As conclusões ficam sob responsabilidade de um perito forense após examinar cada candidato da lista (Dessimoz e Champod, 2008).

2.1.2 Impressão digital

Impressões digitais são padrões únicos, formados por cristas e sulcos, comumente chamados de vales, que existem na camada mais externa da pele dos dedos. Conforme mostrado na Figura 4, as cristas são a porção elevada da pele e os vales são a parte da pele mais baixa e entre as cristas (Hawthorne, 2017).

A diferenciação das características das impressões digitais é realizada em três níveis. O primeiro nível descreve o padrão global que é visível na ponta do dedo mesmo a olho nu.

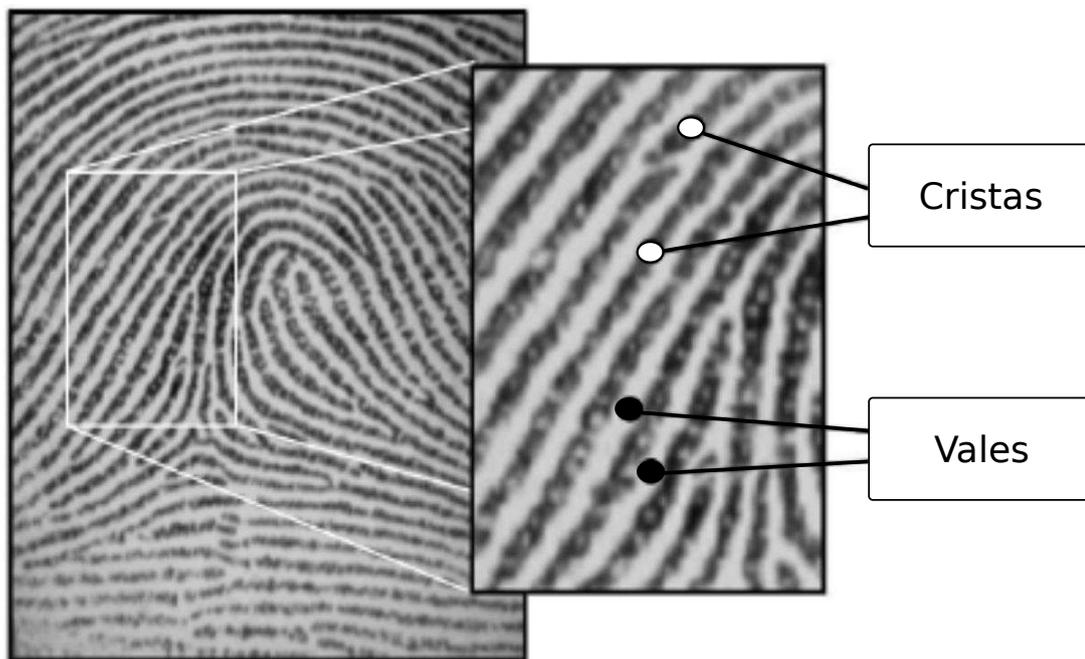


Figura 4: Impressão digital, cristas e vales. Fonte: adaptado de Maltoni et al. (2009a).

O segundo nível descreve as características locais, que são características específicas usadas principalmente para correspondência de impressões digitais, conhecidas como minúcias. O terceiro nível de características descreve detalhes microscópicos, como poros ou bordas de linhas papilares (Champod et al., 2004).

Três categorias principais são definidas pelos padrões formados pelo fluxo geral das cristas, são elas: arcos (*arches*), laços ou presilhas (*loops*) e verticilo (*whorls*). Arcos podem ser subdivididos em arcos planos (*plain arches*) e arcos tentados (*tented arches*), e laços podem ser subdivididos em presilha externa (*left loop*) e presilha interna (*right loop*). Esses padrões são formados pela junção da estrutura de cristas que são articuladas em torno de singularidades chamadas núcleo (*core*) e *delta*, conforme pode ser visto na Figura 5. O *core* de uma impressão digital pode ser definido como o ponto central mais interno onde as cristas formam um *loop*, enquanto o *delta* é o ponto onde essas cristas possuem uma forma triangular (Msiza et al., 2011).

As minúcias de uma impressão digital referem-se às diferentes formas em que uma crista é descontinuada. É observado que as cristas podem se dividir e gerar bifurcações ou podem também terminar subitamente, o que gera uma terminação. As minúcias também podem formar arranjos combinados, como os lagos, que são formados por duas bifurcações opostas, ilhas ou pontos formadas por duas terminações de cristas conectadas e também *spurs*, formados a partir de uma combinação entre terminação e bifurcação. A Figura 6 mostra os principais tipos de minúcias citados.

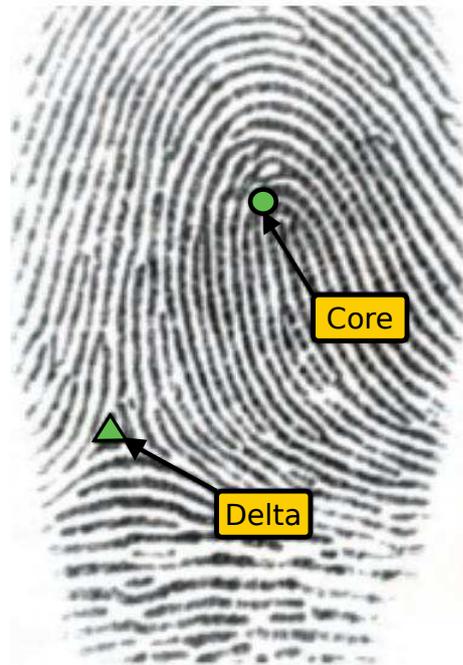


Figura 5: Impressão digital com destaque para o *core* (círculo) e *delta* (triângulo). Fonte: adaptado de Msiza et al. (2011).

2.1.3 Impressão digital latente

Quando a área de cristas da pele entra em contato direto com alguma superfície, uma representação das suas características pode ser deixada. Essas representações são chamadas de marcas (*marks*). As marcas são feitas de resíduos de suor, uma mistura complexa de compostos originários das glândulas écrinas e sebáceas. Pela forma não controlada como são depositadas, as marcas são muitas vezes de qualidade variável em comparação com outras impressões. Quando tais marcas são qualificadas como marcas de dedos (*fingerprints*) ou palma da mão (*palm prints*), significa que a posição correspondente na área da pele foi estabelecida, caso contrário, o termo por si só se referirá a uma impressão de qualquer área da pele. Essas marcas também podem ser chamadas de marcas latentes ou impressões latentes. O termo latente é usado para definir marcas que não podem ser vistas sem a aplicação de técnicas de detecção (Champod e Chamberlain, 2009).

De acordo com Tistarelli e Jain (2017), a aquisição e a análise de marcas latentes na perspectiva das impressões digitais é uma das disciplinas mais antigas da ciência forense, e esse fato abriu caminho para a utilização da impressão digital como um traço biométrico em diversas aplicações. Devido ao fato de muitas vezes não serem diretamente visíveis na superfície em que foram depositadas, as impressões latentes requerem tratamento físico ou químico para sua posterior aquisição por meio de fotografias ou outros métodos sem contato.

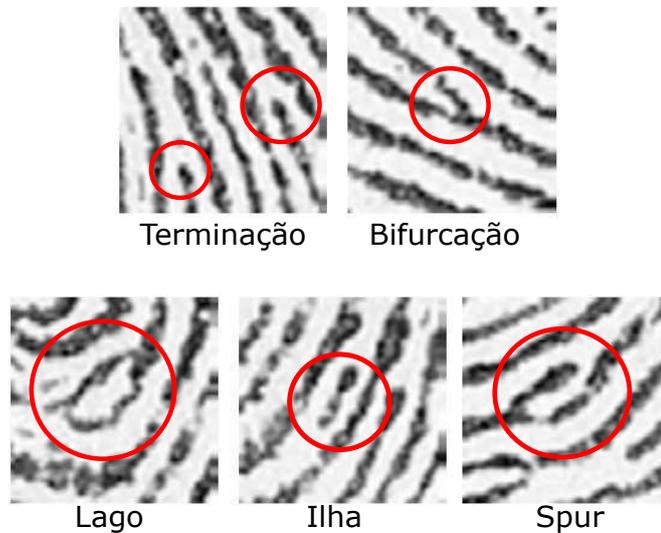


Figura 6: Amostra dos diferentes tipos de minúcias. Fonte: adaptado de Fraser e Williams (2013).

Há alguns anos todo processo de identificação e correspondência de impressões digitais latentes era feito de forma manual por especialistas forenses. Com o passar dos anos, sistemas semiautomáticos de identificação de impressões digitais latentes foram desenvolvidos, automatizando parte desses processos manuais. Impressões digitais latentes são analisadas por especialistas por meio do método ACE-V (Haber e Haber, 2007; Chugh et al., 2018):

- **Análise (A):** é a etapa inicial em que a impressão digital é avaliada para determinar sua adequação à comparação. Durante essa análise, a impressão pode receber um dos três rótulos: VID (válida para identificação), VEO (válida apenas para exclusão) ou NV (sem valor). A classificação VID indica que a impressão pode ser utilizada para identificar um indivíduo, enquanto VEO indica que ela pode ser usada somente para excluir a possibilidade de correspondência por não possuir informações suficientes para individualização. Por outro lado, o rótulo NV é atribuído quando a impressão não apresenta informações suficientes para ser classificada como VID ou VEO.
- **Comparação (C):** é a fase em que o especialista forense compara os traços característicos da impressão latente desconhecida com os das impressões exemplares conhecidas e identifica a semelhança ou diferença entre elas;
- **Avaliação (E):** dependendo dos resultados da fase de comparação, nesta fase o examinador avalia e toma uma decisão que pode ser individualização, exclusão ou inconclusivo. Uma individualização significa que a impressão latente e a impressão conhecida têm uma origem idêntica com exclusão de todos os outros potenciais candidatos. Uma exclusão significa que a impressão latente e a impressão conhecida

não têm a mesma origem. Por último, inconclusivo significa que não foi possível determinar nem a individualização nem a exclusão;

- **Verificação (V):** Nesta fase, um segundo examinador verifica independentemente a decisão do examinador original, analisando e comparando a impressão digital. Isso é feito seguindo um padrão que estabelece as regras de seleção do verificador e a resolução de conclusões conflitantes, conforme definido pelo laboratório ou órgão forense responsável pela análise.

Como consequência de um procedimento manual extenso, podem haver inconsistências nas avaliações de impressões latentes por diferentes examinadores. Isso reforça a necessidade do desenvolvimento e utilização de sistemas automatizados de identificação de impressões digitais latentes (Singla et al., 2020).

Apesar dos avanços alcançados na correspondência de impressões digitais escaneadas, a correspondência automatizada de impressões digitais latentes ainda é um problema desafiador. O pré-processamento de impressões latentes é necessário antes da marcação e extração de características. É importante que a impressão latente seja segmentada do fundo ruidoso e que a qualidade das cristas seja aprimorada. Além disso, as características da impressão latente coletada são marcadas manualmente por um examinador e fornecidas como entrada para um AFIS, que compara as características com o banco de dados e gera uma lista das melhores correspondências possíveis.

2.2 Aprendizagem de máquina profunda

Kaplan e Haenlein (2019) definem Inteligência Artificial (IA) como a capacidade de sistemas de interpretar e aprender a partir de dados externos, e de utilizar esse aprendizado para realização de tarefas através de adaptação flexível. IA é também um campo de estudos da Ciência da Computação que visa a possibilitar que computadores realizem tarefas comumente associadas a inteligência humana. A automatização de trabalhos rotineiros, reconhecimento e compreensão de fala ou de imagens, realização de diagnósticos médicos são alguns exemplos de aplicações práticas e pesquisas ativas na área de IA (Goodfellow et al., 2016).

Aprendizagem de Máquina (ML, do inglês *Machine Learning*) é vista como um campo de estudos dentro da área de IA que se preocupa com a questão de como construir sistemas que possam ir melhorando automaticamente, através da experiência e do uso de dados. Uma definição formal é dada por Michalski et al. (2013):

“Diz-se que um programa de computador aprende com a experiência E no que diz respeito a uma classe de tarefas T e medida de desempenho P , se

seu desempenho nas tarefas em T , conforme medido por P , melhora com a experiência E .”

Aprendizagem de Máquina é uma abordagem mais voltada ao desenvolvimento de algoritmos que aprendam automaticamente padrões intrínsecos nos dados, ao invés de buscar-se uma representação manual de regras para o problema.

Entretanto, uma dificuldade existente nas aplicações de IA no mundo real é que existem muitos fatores de variação que influenciam os dados que são possíveis de se observar, como, por exemplo, quando as cores de um objeto em uma imagem são muito próximas das cores do plano de fundo, ou quando a forma desse objeto depende do ângulo em que está sendo visto. Para solucionar esse problema, faz-se necessário que os fatores de variação sejam separados, e descartados aqueles considerados sem importância. Porém, nem sempre é simples abstrair tais características dos dados em diferentes níveis.

Aprendizagem de Máquina Profunda é uma sub-área particular dentro do campo de aprendizagem de máquina que possibilita que modelos computacionais compostos de várias camadas de processamento aprendam representações de dados com vários níveis de abstração (LeCun et al., 2015). DL aborda o problema central no aprendizado de representação, introduzindo representações que são expressas em termos de outras representações mais simples. Por exemplo, considere-se a tarefa de identificação de um objeto em uma imagem. Para algoritmos tradicionais, é difícil realizar o mapeamento dos dados brutos de entrada para a identidade do objeto. Através da aplicação do DL tenta-se superar essa dificuldade dividindo-se o mapeamento mais complexo em mapeamentos mais simples aninhados, cada um correspondendo a uma camada diferente do modelo. Como exemplo, a primeira camada pode ficar responsável por detectar bordas, a segunda camada pode então detectar cantos e contornos definidos por um conjunto de bordas. Sendo encontrados padrões de contornos e cantos, a terceira camada pode detectar partes inteiras de objetos e essas partes detectadas podem ser finalmente utilizadas para avaliar a existência do objeto na imagem (Goodfellow et al., 2016).

Os modelos mais fundamentais de DL são as redes neurais de alimentação direta (*deep feedforward*), também conhecidas como Perceptrons Multicamada (MLP). O objetivo dessas redes é realizar a aproximação de alguma função f . Essas redes são um modelo simplificado da estrutura da rede neural biológica e são constituídas por unidades de processamento interconectadas. As unidades de processamento recebem N valores de entrada, associam um peso a cada valor e calculam uma soma ponderada que é chamada de valor de ativação. O sinal do peso determina se a entrada é excitatória ou inibitória (Yegnanarayana, 2009). Uma pilha de nós no mesmo nível forma uma camada e sequências de camadas compõem toda a rede neural. A camada oculta é a camada entre a camada de entrada e a camada de saída. Uma ilustração desse modelo pode ser visualizada na Figura

7. Em uma rede neural de alimentação direta, não há conexões de retroalimentação entre as camadas.

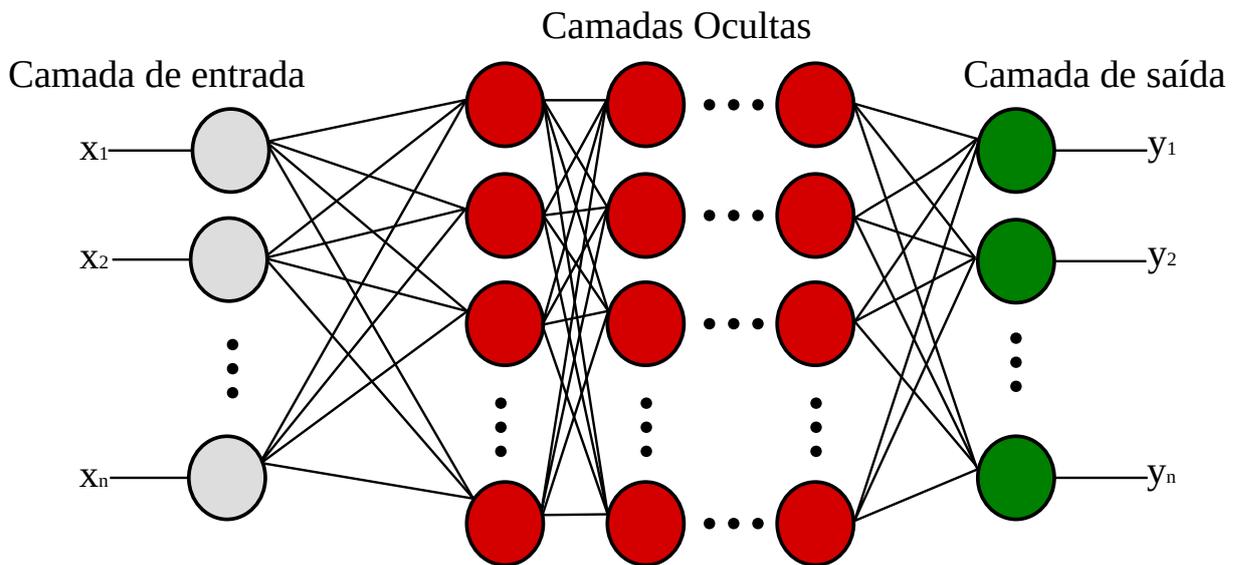


Figura 7: Ilustração de uma rede neural de alimentação direta. Fonte: adaptado de Azzouni et al. (2017)

Nas subseções a seguir, são descritos alguns tipos específicos de redes neurais (NNs, do inglês *Neural Networks*), que foram utilizados neste trabalho, e que são necessários para uma melhor compreensão da pesquisa.

2.2.1 Redes neurais convolucionais

Uma Rede Neural Convolucional (CNN, do inglês *Convolutional Neural Network*) é um tipo especial de rede neural multicamadas ou arquitetura de DL inspirada no sistema visual biológico, o que a torna um modelo adequado para diversas aplicações no campo da visão computacional (Ghosh et al., 2020). Uma CNN tradicional é composta por blocos únicos ou múltiplos de camadas de convolução e abstração (*pooling*), opcionalmente seguidos por uma ou várias camadas totalmente conectadas, e uma camada de saída. A camada convolucional é o bloco de construção central de uma CNN, cujo objetivo é aprender representações de características da entrada. Essa camada é composta por vários núcleos (*kernels*) ou filtros de convolução que podem ser aprendidos e são utilizados para calcular diferentes mapas de características.

A convolução é a operação principal de uma CNN. Em sua forma mais geral, é uma operação em duas funções de um argumento de valor real (Goodfellow et al., 2016). Considerando as funções x e w , a saída resultante da operação de convolução entre a entrada x e o filtro w no domínio contínuo, aplicada em um intervalo de tempo de medição τ - denotada por $h(t)$ - pode ser definida conforme a Equação 1:

$$h(t) = \int_{-\infty}^{\infty} x(\tau)w(t - \tau) d\tau \quad (1)$$

A operação de convolução é tipicamente representada por um asterisco, conforme pode ser visto na Equação 2:

$$h(t) = (x * w)(t) \quad (2)$$

em que x representa a entrada, w é o *kernel*, e t é o ponto onde a convolução é calculada. Se w é uma função de densidade de probabilidade válida, a convolução pode ser considerada como a média ponderada da entrada naquele ponto t . Quando os dados são discretizados, a Equação 1 pode ser reescrita como a Equação 3:

$$h(t) = \sum_{a=-\infty}^{\infty} x(\tau)w(t - \tau) \quad (3)$$

A operação de convolução também pode ser aplicada de forma simultânea em múltiplos eixos. Por exemplo, ao utilizar uma imagem bidimensional I como entrada, é recomendável empregar um *kernel* bidimensional K para a convolução. A Equação (4) define a operação de convolução bidimensional, também ilustrada na Figura 8.

$$H(i, j) = (I * K)(i, j) = \sum_m \sum_n I(m, n) \cdot K(i - m, j - n) \quad (4)$$

em que $H(i, j)$ representa o valor resultante da convolução no ponto de saída (i, j) . A convolução é realizada somando o produto dos elementos da matriz de entrada I com os elementos do *kernel* K , considerando o deslocamento $(i - m, j - n)$.

De acordo com Ghosh et al. (2020), as principais vantagens das camadas convolucionais são:

- **conectividade esparsa:** em uma rede neural totalmente conectada, cada neurônio de uma camada se conecta a cada neurônio da próxima camada, mas em uma CNN existe um número restrito de pesos entre duas camadas. Como resultado, o número de conexões ou pesos necessários é reduzido e a quantidade de memória para armazenar esses pesos também é reduzida, resultando em maior eficiência em termos de memória;
- **pesos compartilhados:** em uma CNN, cada neurônio em uma camada convolucional não possui pesos individuais exclusivos conectados a neurônios de camadas adjacentes. Em vez disso, todos os pesos são compartilhados e trabalham com cada

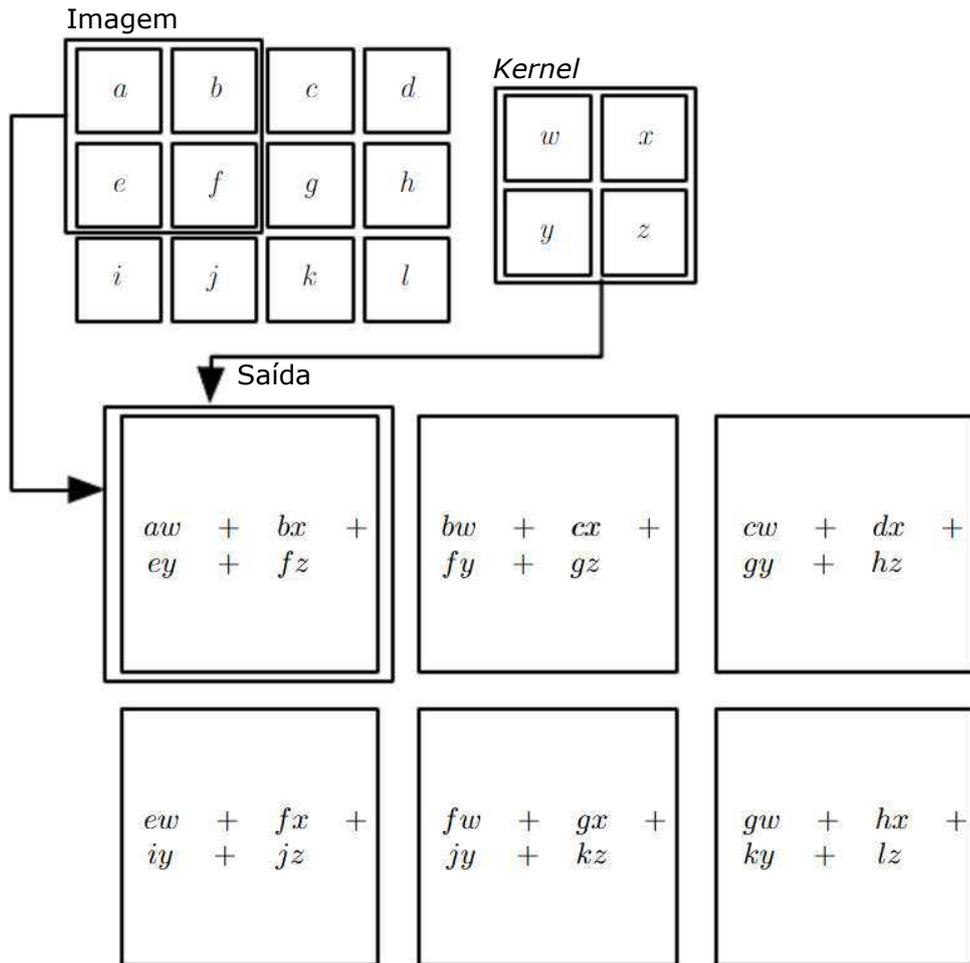


Figura 8: Ilustração da operação de convolução bidimensional. Fonte: adaptado de Goodfellow et al. (2016).

pixel da imagem de entrada. Isso significa que um conjunto de pesos é usado para processar toda a imagem, em vez de aprender pesos separados para cada neurônio, reduzindo drasticamente o tempo de treinamento, bem como outros custos envolvidos.

2.2.1.1 Pooling

As camadas de *pooling* (ou de abstração) são aplicadas para reduzir o tamanho dos mapas de características obtidos após as operações de convolução, preservando as informações mais dominantes em cada etapa do *pool*. A operação de *pooling* é executada especificando-se o tamanho do agrupamento (*pool size*) e o tamanho do passo (*stride*) da operação. Existem diferentes tipos de técnicas de *pooling*, porém as principais são o *pooling* máximo (*max pooling*) e *pooling* médio (*average pooling*). Um exemplo de ambas operações pode ser observado na Figura 9.

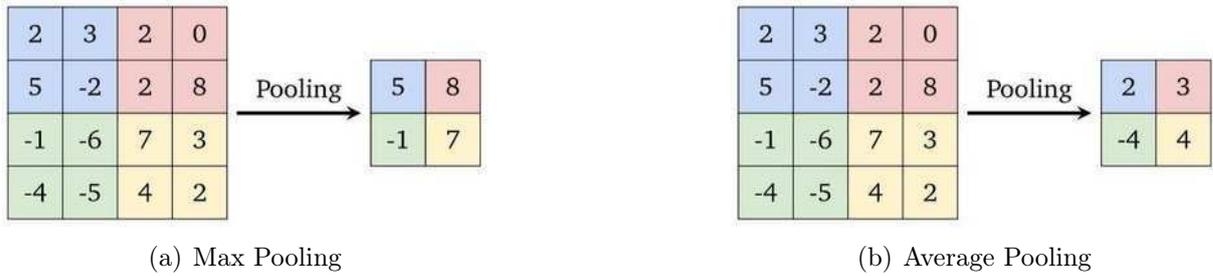


Figura 9: Exemplo das operações de *max pooling* and *average pooling* realizadas em um mapa de características de 4×4 com $pool_size = 2 \times 2$, $stride = 2$. Fonte: adaptado de Guissous (2019).

Goodfellow et al. (2016) afirmam que o *pooling* ajuda a reduzir o número de cálculos realizados na rede e torna o modelo mais robusto quanto a pequenas variações na posição dos elementos na imagem de entrada, controlando o *overfitting*. Em geral, as camadas de *pooling* são aplicadas após a função de ativação das camadas convolucionais e operam em cada mapa de características de forma independente.

2.2.1.2 Normalização em lote

Um dos fatores que torna desafiador o treinamento de redes neurais profundas é o fato de que a distribuição das entradas de cada camada muda durante o treinamento, conforme os parâmetros das camadas anteriores são atualizados. Isso retarda o treinamento, exigindo taxas de aprendizado mais baixas e uma inicialização cuidadosa dos parâmetros. Ioffe e Szegedy (2015) referem-se a esse fenômeno como mudança de covariância interna (*internal covariate shift*) e apontam como solução para esse problema a normalização das entradas da camada.

Normalização em lote (BN, do inglês *Batch Normalization*) (Ioffe e Szegedy, 2015) é uma técnica amplamente utilizada que possibilita um treinamento mais rápido e estável de redes neurais profundas. Em alto nível, BN objetiva melhorar o treinamento de redes neurais por meio da estabilização das distribuições de entradas de camadas introduzindo camadas de rede adicionais que controlam a média e a variância dessas distribuições (Santurkar et al., 2018).

Considerando camadas BN para redes convolucionais, têm-se a entrada $I_{b,c,x,y}$ e a saída $O_{b,c,x,y}$. As dimensões correspondem às amostras dentro de um lote b , canais c e duas dimensões espaciais x, y . BN aplica a mesma normalização para todas as ativações em um determinado canal conforme a Equação 5.

$$O_{b,c,x,y} \leftarrow \gamma_c \frac{I_{b,c,x,y} - \mu_c}{\sqrt{\sigma_c^2 + \epsilon}} + \beta_c \quad \forall b, c, x, y \quad (5)$$

BN subtrai a média das ativações de todas as entradas no canal c , Equação 6,

$$\mu_c = \frac{1}{|\mathcal{B}|} \sum_{b,x,y} I_{b,c,x,y} \quad (6)$$

em que \mathcal{B} é o conjunto de todas as ativações no canal c em todos os elementos do mini-lote b em todas as localizações espaciais x, y . Posteriormente, BN divide a ativação centralizada pelo desvio padrão σ_c (somando ϵ para estabilidade numérica). A normalização é seguida por uma transformação por canal parametrizada através de γ_c e β_c que são aprendidos durante o treinamento (Bjorck et al., 2018).

Ioffe e Szegedy (2015) também afirmam que BN, além de tornar o treinamento de uma rede neural profunda mais rápido e estável, também pode substituir o uso de *dropout* como técnica de regularização.

2.2.1.3 Funções de ativação

A principal tarefa de uma função de ativação em modelos de redes neurais consiste em realizar o mapeamento de uma entrada para uma saída, por meio do cálculo da soma ponderada da entrada do neurônio, acrescido de um viés, se houver. Em outras palavras, a função de ativação determina o comportamento de um neurônio, decidindo como este neurônio acionará, ou se não acionará, a partir de uma entrada específica, e, conseqüentemente, gerando a respectiva saída correspondente (Ghosh et al., 2020). Em uma arquitetura CNN, após cada camada treinável, camadas de ativação não linear são usadas. O comportamento de não linearidade dessas camadas permite que o modelo aprenda características mais complexas e crie um mapeamento não linear entre as entradas e saídas da rede neural.

Existem várias funções de ativações conhecidas e aplicadas em modelos de aprendizagem profunda. A seguir serão descritas duas das principais funções que foram utilizadas no modelo desenvolvido nesta pesquisa:

- **ReLU**: a *Rectifier Linear Unit* (ReLU) (Nair e Hinton, 2010) é a função de ativação mais comumente usada em CNNs. É uma função que converte todos os valores de entrada em números positivos e possui a vantagem de necessitar de uma carga computacional menor em comparação com outras funções de ativação. A curva da função ReLU é exibida na Figura 10.

Sua representação matemática é denotada pela Equação 7:

$$f(x)_{ReLU} = \mathbf{max}(0, x) \quad (7)$$

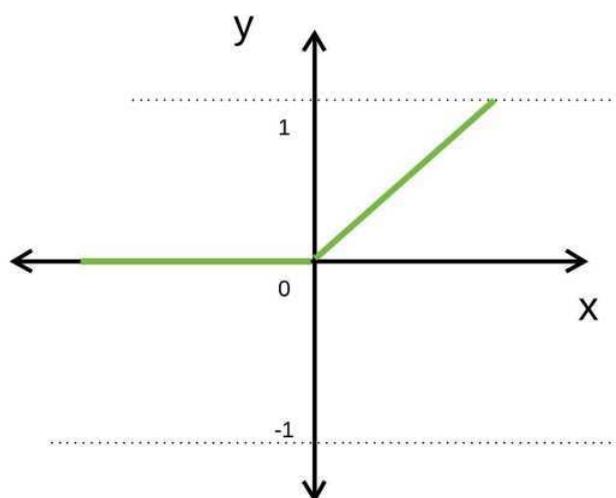


Figura 10: Gráfico da função ReLU. Fonte: Ghosh et al. (2020).

- **sigmoide**: a função de ativação sigmoide (*sigmoid*) (Rasamoelina et al., 2020) recebe números reais como entrada e entrega como saída um valor no intervalo de $[0,1]$. A curva da função sigmoide é exibida na Figura 11.

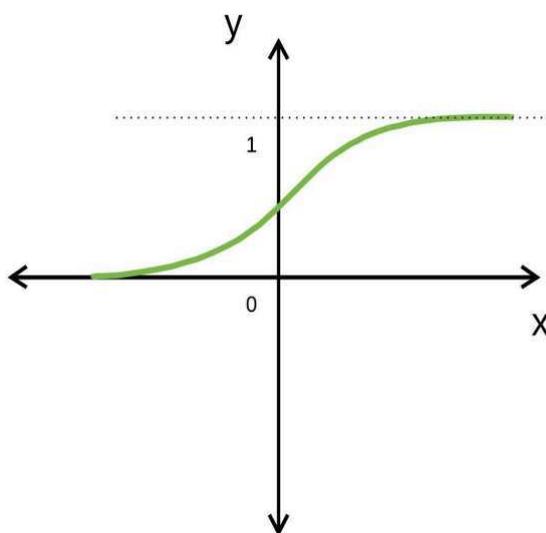


Figura 11: Gráfico da função *sigmoid*. Fonte: Ghosh et al. (2020).

Sua representação matemática é denotada pela Equação 8:

$$f(x)_{sigm} = \frac{1}{1 + e^{-x}} \quad (8)$$

2.2.1.4 Dropout

O *Dropout* é uma das técnicas de regularização mais utilizadas para prevenção de *overfitting* (Ghosh et al., 2020) e consiste em desativar alguns neurônios da rede,

aleatoriamente escolhidos, a cada época de treinamento. A desativação dessas unidades de processamento tem como propósito distribuir o poder de seleção de características para todos os neurônios igualmente e forçar o modelo a aprender várias características independentemente (Srivastava et al., 2014).

2.2.2 Autocodificadores

Um autocodificador (AE, do inglês *Autoencoder*) é um tipo específico de rede neural projetado para criar uma representação codificada da entrada original e, em seguida, decodificá-la de modo que observações da representação original da entrada possam ser reconstruídas (Tschannen et al., 2018; Bank et al., 2020). A arquitetura mais comum de um AE é ilustrada na Figura 12 e suas principais partes são:

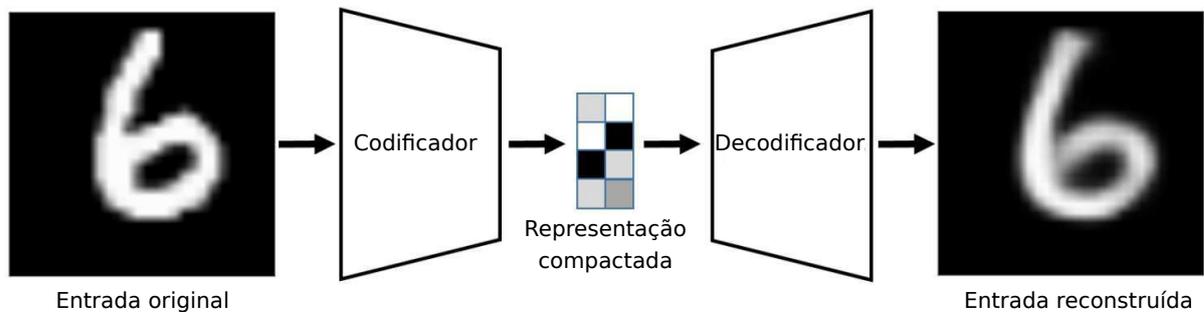


Figura 12: Arquitetura de um autocodificador. Fonte: adaptado de Bank et al. (2020).

1. **Codificador:** responsável por aprender uma representação compactada da entrada, também chamada de representação latente. Pode ser representado como uma função de codificação $h = f(x)$, onde x é a entrada e h corresponde a codificação da entrada em uma representação latente ($h : \mathbb{R}^D \rightarrow \mathbb{R}^K$, onde D e K representam a dimensionalidade da entrada e da representação latente, respectivamente).
2. **Decodificador:** responsável por mapear a representação latente h , aprendida pelo codificador, de volta a uma reconstrução ($\mathbb{R}^K \rightarrow \mathbb{R}^D$). Pode ser descrito como uma função $r = g(h)$.

Aprender a definir $g(f(x)) = x$ não é especialmente útil. Em vez disso, os AEs são projetados para criar uma reconstrução aproximada de entradas que se assemelhem aos dados de treinamento, de modo que o modelo aprenda a priorizar quais aspectos da entrada devem ser aproximados. Assim, geralmente são aprendidas propriedades úteis dos dados (Goodfellow et al., 2016).

Nos últimos anos, AEs vêm sendo aplicados com sucesso em tarefas como: redução de dimensionalidade (Petschornig et al., 2017; Wang et al., 2015), tarefas de recuperação de informação, (Pfeiffer et al., 2018), detecção de anomalias (Sakurada e Yairi, 2014), e segmentação de imagens (Baur et al., 2018; Karimpouli e Tahmasebi, 2019).

2.2.3 ResNet

Uma rede residual (ResNet, do inglês *Residual Network*) é um modelo de rede neural introduzido pela primeira vez por He et al. (2016). Tem por característica a aprendizagem de funções residuais com referência às entradas da camada. Formalmente, denotando um mapeamento subjacente desejado como $\mathcal{H}(x)$, as camadas não lineares empilhadas se ajustam a outro mapeamento de $\mathcal{F}(x) = \mathcal{H}(x) - x$. O mapeamento original é reformulado como $\mathcal{F}(x) + x$. Tal reformulação pode ser aplicada a redes neurais de alimentação direta com salto de conexões, conforme ilustrado na Figura 13. As conexões de atalho são aquelas que saltam uma ou mais camadas (*skip connections*).

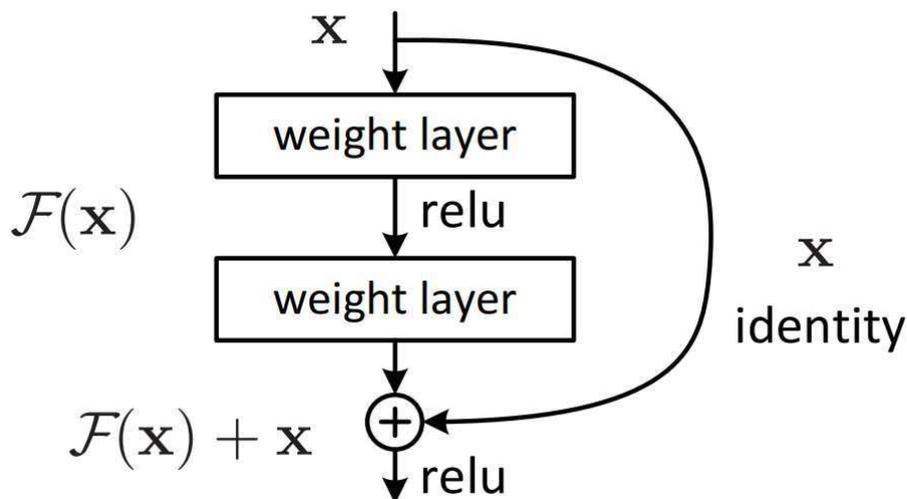


Figura 13: Exemplo de um tipo de bloco residual. Fonte: He et al. (2016).

2.3 Domínio espacial e domínio da frequência

Domínio espacial e domínio da frequência são duas formas de representar imagens no processamento digital de imagens. De acordo com Solomon e Breckon (2011), uma imagem digital pode ser considerada como uma representação discreta de dados que possuem informação espacial e de intensidade (cor).

No domínio espacial, a imagem é representada como uma matriz bidimensional de intensidades de pixels, dada por uma função $f(x, y)$, em que x e y são coordenadas no

plano da imagem e $f(x, y)$ é a intensidade do pixel naquela coordenada. No domínio da frequência, cada pixel de uma imagem representa uma determinada frequência espacial. A informação contida em um pixel nessa representação é composta por duas partes: amplitude e fase. A amplitude é a magnitude do sinal da frequência espacial correspondente ao pixel, enquanto a fase representa o deslocamento da onda senoidal que descreve a variação da amplitude do sinal (Jain, 1989; Gonzalez e Woods, 2009).

A frequência das cristas é uma propriedade intrínseca das impressões digitais que descreve a distância local entre as cristas em cada ponto da imagem. Essa informação é importante para o processamento de impressões digitais porque as cristas e sulcos aparecem como estruturas periódicas na imagem da impressão digital. Essa periodicidade se manifesta como picos no espectro de frequência da imagem (Guan et al., 2013).

O espectro de frequência é uma representação da amplitude de um sinal em função da sua frequência. O espectro de frequência de uma imagem de impressão digital pode ser usado para identificar as estruturas periódicas, como as cristas e sulcos. Por ser uma característica global da impressão digital a frequência das cristas é bastante explorada em métodos de pré-processamento de imagens usados em sistemas automáticos de identificação de impressões digitais. Isso porque as informações de frequência das cristas pode ser usada para remover ruído por filtragem, melhorar o contraste da imagem e destacar os detalhes das cristas (Orczyk e Wieclaw, 2011).

Existem diferentes tipos de transformadas que podem ser usadas para se obter a representação de uma imagem no domínio da frequência, sendo as transformadas de Fourier e a do Cosseno duas das mais comuns (Oppenheim e Schaffer, 2010). Na Figura 14 é mostrada a imagem do *ground truth* de uma impressão digital existente no banco de dados sintético, criado para o treinamento da RM-net, e seu espectro de frequências normalizado entre 0 e 255, obtido por meio da Transformada de Fourier. Nesta pesquisa, a Transformada Discreta de Fourier (DFT, do inglês *Discrete Fourier Transform*), utilizando o algoritmo da Transformada rápida de Fourier (FFT, do inglês *Fast Fourier Transform*), e a Transformada Discreta do Cosseno (DCT, do inglês *Discrete Cosine Transform*) foram as transformadas utilizadas para se obter as representações no domínio da frequência das imagens existentes na base de dados sintética de treinamento.

2.3.1 Transformada de Fourier

A transformada de Fourier (Nussbaumer e Nussbaumer, 1981) é uma técnica matemática utilizada para representar uma função ou sinal no domínio da frequência, em contraste com o domínio do tempo e que, de acordo com Bracewell (2000), é amplamente utilizada em processamento de sinais e processamento de imagem, permitindo a decomposição de um sinal em suas componentes de frequência, que em alguns casos podem ser

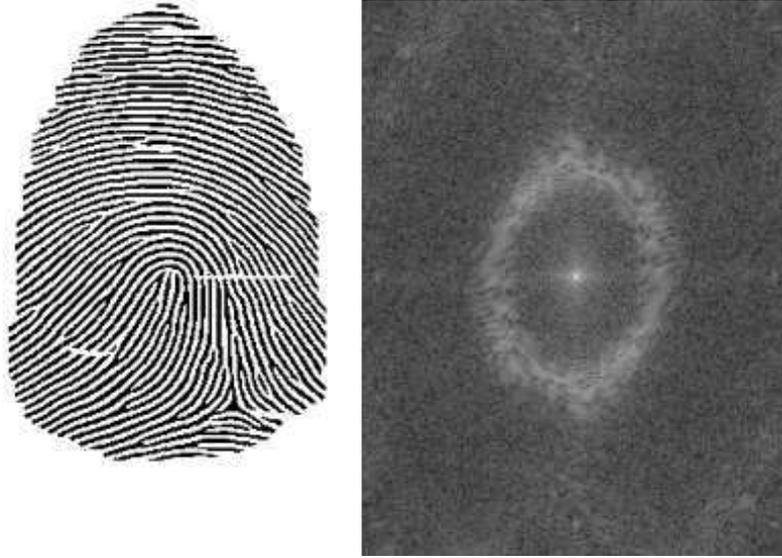


Figura 14: *Ground truth* de uma impressão digital existente no banco de dados sintético de treinamento e sua respectiva representação no domínio da frequência obtida por meio da transformada de Fourier. Fonte: De autoria própria.

analisadas e manipuladas de forma mais eficiente do que as informações no domínio do tempo. A transformada unidimensional de Fourier é definida pela Equação 9:

$$F(\omega) = \int_{-\infty}^{\infty} f(t)e^{-i\omega t} dt \quad (9)$$

em que $f(t)$ é a função no domínio do tempo, ω é a frequência e $e^{-i\omega t}$ é a função exponencial complexa. Segundo Gonzales e Woods (2018), a transformada de Fourier produz uma representação complexa no domínio da frequência.

A transformada de Fourier pode ser aplicada a imagens, permitindo a decomposição da imagem em suas componentes de frequência (Oppenheim e Schaffer, 2010). A transformada de Fourier bidimensional é definida na Equação 10:

$$F(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)e^{-i2\pi(ux+vy)} dx dy \quad (10)$$

em que $f(x, y)$ é a imagem no domínio espacial, u e v são as frequências horizontal e vertical, respectivamente. Assim como a transformada de Fourier unidimensional, a transformada de Fourier bidimensional também produz uma representação complexa no domínio da frequência.

2.3.2 Transformada discreta do cosseno

A DCT, introduzida por Ahmed et al. (1974), é uma técnica amplamente empregada no processamento de sinais e na compressão de dados. A DCT é uma transformada matemática que converte um sinal discreto de entrada em uma sequência de coeficientes, representando a contribuição de diferentes frequências presentes no sinal. A DCT unidimensional, a qual é frequentemente utilizada, é definida pela Equação (11).

$$X(k) = \sum_{n=0}^{N-1} x(n) \cdot \cos\left(\frac{\pi}{N} \cdot (n + 0.5) \cdot k\right) \quad (11)$$

em que $x(n)$ representa o valor do sinal de entrada em n , $X(k)$ é o coeficiente DCT correspondente à frequência k e N é o número total de amostras do sinal. A aplicação da DCT é amplamente difundida em áreas como processamento de imagens médicas, transmissão de vídeos e sistemas de comunicação digital (Gonzalez e Woods, 2009; Pratt, 2007).

A DCT bidimensional, aplicada em uma matriz bidimensional de dados, como uma imagem, é uma extensão da DCT unidimensional para duas dimensões e é definida conforme a Equação (12):

$$X_{ij} = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} x_{mn} \cdot \cos\left(\frac{\pi}{M} \cdot \left(m + \frac{1}{2}\right) \cdot i\right) \cdot \cos\left(\frac{\pi}{N} \cdot \left(n + \frac{1}{2}\right) \cdot j\right) \quad (12)$$

em que X_{ij} são os coeficientes de frequência resultantes da transformada, x_{mn} são os valores de entrada da matriz, M é o número de linhas e N é o número de colunas da matriz.

2.4 Métricas de desempenho

Nesta subseção serão descritas as métricas utilizadas nesta pesquisa para avaliar o desempenho da abordagem desenvolvida, tanto na reconstrução da imagens dadas como entrada, como no resultado final de identificação e correspondência das impressões digitais latentes realizadas pela rede neural.

2.4.1 Acurácia

Para um problema de classificação binária, as possíveis predições são: Verdadeiro Positivo (VP), Verdadeiro Negativo (VN), Falso Positivo (FP) e Falso Negativo (FN). A

acurácia, que é uma das métricas mais utilizadas para avaliar o desempenho de modelos de ML, mede quantas predições corretas foram realizadas dentre todas as amostras de um dado conjunto de dados. Isto é, a acurácia avalia a proporção de predições corretas. A acurácia é descrita pela Equação 13:

$$\text{Acurácia} = \frac{VP + VN}{VP + VN + FP + FN} \quad (13)$$

A acurácia tem como desvantagem não ser adequada quando se faz uso de conjuntos de dados desbalanceados. Por exemplo, supondo uma base de dados em que 90% das amostras são da classe positiva e os 10% restante são da classe negativa e um classificador que sempre entregue como classe positiva o resultado de uma predição. A acurácia geral será de 90% devido ao desbalanceamento entre as classes, porém este classificador erra 100% dos exemplos da classe negativa.

2.4.1.1 Acurácia rank- n

A acurácia *rank- n* , também chamada de acurácia *top- n* , é uma métrica frequentemente utilizada para medir o desempenho de redes neurais, principalmente em pesquisas relacionadas à visão computacional.

Considerando o contexto de identificação biométrica, o problema de identificação consiste em determinar a identidade de um indivíduo por meio de uma amostra biométrica coletada e comparada com todos os modelos em um banco de dados. O cenário em que presume-se a existência do indivíduo no banco de dados é chamado de conjunto fechado (*closed set*) (Poh et al., 2012). Portanto, em uma identificação em conjunto fechado, uma amostra de teste recebe classificação positiva quando o indivíduo real é identificado dentro de um grupo de n candidatos retornados por um sistema de identificação, disto segue a denominação *rank- n* .

A ideia geral da acurácia *rank- n* é calcular a proporção de amostras em que a classe correta está entre as n principais previsões feitas pelo modelo. Entretanto, é importante salientar que sua implementação pode variar dependendo do contexto e dos requisitos específicos do problema. Sendo assim, uma fórmula usual para calcular a acurácia *rank- n* é a seguinte, mostrada na Equação (14):

$$\text{Acurácia rank-}n = \frac{AC}{NT} \quad (14)$$

em que AC representa o número de amostras com a classe correta entre as n principais previsões e NT é o número total de amostras.

2.4.2 Índice de similaridade estrutural

O índice de similaridade estrutural (SSIM, do inglês *Structural Similarity Index Measure*) (Wang et al., 2004) é uma métrica de qualidade usada para medir a similaridade entre duas imagens. Em vez de usar métodos tradicionais de soma de erros, o SSIM foi concebido por meio da modelagem de qualquer distorção de imagem como uma combinação de três fatores que são: distorção de estrutura, distorção de luminância e distorção de contraste (Hore e Ziou, 2010). Dada uma imagem de referência f e uma imagem de teste g , ambas de tamanho $M \times N$, o SSIM entre f e g é definido conforme a Equação 15:

$$SSIM(f, g) = l(f, g)c(f, g)s(f, g) \quad (15)$$

em que as comparações individuais de proximidade das médias μ_f e μ_g de luminância l (Equação 16), da proximidade do contraste c , medido pelo desvio padrão σ_f e σ_g (Equação 17) e da proximidade de estrutura s , em que σ_{fg} é a covariância entre f e g (Equação 18), são, respectivamente:

$$l(f, g) = \frac{2\mu_f\mu_g + C_1}{\mu_f^2 + \mu_g^2 + C_1} \quad (16)$$

$$c(f, g) = \frac{2\sigma_f\sigma_g + C_2}{\sigma_f^2 + \sigma_g^2 + C_2} \quad (17)$$

$$s(f, g) = \frac{\sigma_{fg} + C_3}{\sigma_f\sigma_g + C_3} \quad (18)$$

Os valores do índice SSIM são positivos e estão em $[0, 1]$. Assim, um valor 0 significa que não há correlação entre as imagens e 1 significa que $f = g$. As constantes positivas C_1 , C_2 e C_3 são usadas para evitar denominadores nulos.

2.4.3 NIST fingerprint image quality

Foi mostrado por Tabassi et al. (2004) e Watson et al. (2004) que à medida que a qualidade da imagem da impressão digital diminui, o desempenho das técnicas de identificação e correspondência também diminui. Isso mostra que a qualidade da imagem da impressão digital é uma característica importante para determinar a eficácia das ferramentas de casamento de impressão digital nas tarefas de comparação.

Conhecer a qualidade da imagem da impressão digital pode mudar a forma como as impressões digitais são processadas. Por exemplo, várias tentativas podem ser feitas para se obter impressões digitais com melhor qualidade no momento da captura da imagem.

Ao se ter antecipadamente uma referência da qualidade de uma impressão digital, um sistema pode lidar com impressões digitais de qualidade diferente de maneiras diferentes. Tomando como exemplo um cenário em que é possível fazer uso de diferentes ferramentas de teste de correspondência, uma ferramenta mais rápida computacionalmente pode ser usada em imagens de melhor qualidade sem queda na precisão e uma outra mais elaborada pode ser usada em imagens de qualidade inferior para melhorar a precisão da correspondência.

NIST Fingerprint Image Quality (NFIQ) é uma implementação do algoritmo descrito na pesquisa de Tabassi et al. (2004). Dada uma imagem de impressão digital, o NFIQ atribui uma pontuação com o valor que pode ser de 1 a 5, sendo 1 a indicação de melhor qualidade e 5 a de pior, conforme apresentado no Quadro 1.

Quadro 1: Descrição da pontuação de qualidade que pode ser atribuída pelo NFIQ para as impressões digitais

Pontuação	Qualidade
5	Ruim
4	Razoável
3	Boa
2	Muito boa
1	Excelente

2.5 Considerações finais sobre o capítulo

Neste capítulo foram apresentados os principais conceitos necessários para uma melhor compreensão desta tese. Foram abordados os principais conceitos relacionados à biometria, com destaque para sua utilização na biometria na ciência forense. Também foram explicadas as principais características das impressões digitais latentes e as dificuldades relacionadas ao realce desse tipo de impressão digital. Foram abordados temas relacionados à aprendizagem profunda, incluindo modelos e características de redes neurais que se fazem necessários para explicar o modelo desenvolvido e aplicado nesta pesquisa. Por fim, foram apresentados detalhes sobre as métricas utilizadas para avaliar o desempenho da RM-net. No próximo capítulo são apresentadas e discutidas algumas pesquisas relacionadas com a pesquisa realizada nesta tese.

3 Pesquisas Relacionadas

Neste capítulo é apresentada uma revisão da literatura sobre pesquisas relacionadas ao problema de realce de imagens de impressões digitais latentes utilizando técnicas de *Deep Learning*. O objetivo é fornecer uma visão geral do estado da arte e das principais abordagens aplicadas no tema, com foco em publicações a partir de 2017. A revisão foi realizada com base em uma pesquisa sistemática na ferramenta Google Scholar¹ e na análise das referências bibliográficas dos artigos encontrados. Para melhor organização, as pesquisas foram agrupadas em duas categorias: as que utilizam redes neurais convolucionais e suas variações, e as que empregam redes neurais generativas.

3.1 Métodos baseados em redes convolucionais

Inspirados por casos de sucesso de aplicação de CNNs a problemas relacionados ao processamento de imagens, Svoboda et al. (2017) desenvolveram um autocodificador convolucional (CAE, do inglês *Convolutional Autoencoder*) para realizar o realce de imagens de impressão digital latente. Em comparação com pesquisas mais recentes, a arquitetura aplicada pode ser considerada simples, possui poucas camadas e, diferentemente da tendência das pesquisas mais atuais, não faz uso de conexões de atalho entre camadas do codificador e do decodificador. O CAE desenvolvido consiste em 5 camadas convolucionais, com ativação ReLU e aplicação de BN no codificador, seguidas por 5 camadas convolucionais, de maneira espelhada, substituindo-se a função de ativação ReLU pela função Leaky-ReLU, no decodificador. A arquitetura proposta pode ser visualizada na Figura 15. Os autores não relataram o uso de nenhuma técnica específica para redução de *overfitting*. Para treinamento da rede neural, foi criado um banco de dados sintético de impressões digitais latentes a partir da adição de ruído gaussiano e de diferentes planos de fundos em imagens sintéticas geradas por uma implementação de código aberto fundamentada no algoritmo do programa SFinGe (Maltoni, 2004). As bases de dados utilizadas para verificar o desempenho da abordagem proposta foram a IIIT-Delhi *latent fingerprint* (Sankaran et al., 2011) e a IIIT-Delhi MOLF (Sankaran et al., 2015b). Na primeira, o melhor resultado do teste de casamento foi de 79,04% reportado como acurácia *rank-10*. Na segunda base de dados, o melhor resultado obtido, em termos de acurácia *rank-50* foi de 22,36%. Uma questão interessante observada foi a opção dos autores pelo Erro Quadrático Médio (MSE, do inglês *Mean Squared Error*) como função de custo, uma vez que outras pesquisas têm mostrado que a utilização do Erro Médio Absoluto (MAE, do inglês *Mean Absolute Error*) tende a produzir imagens menos borradas e mais nítidas como saída.

¹<https://scholar.google.com.br/>

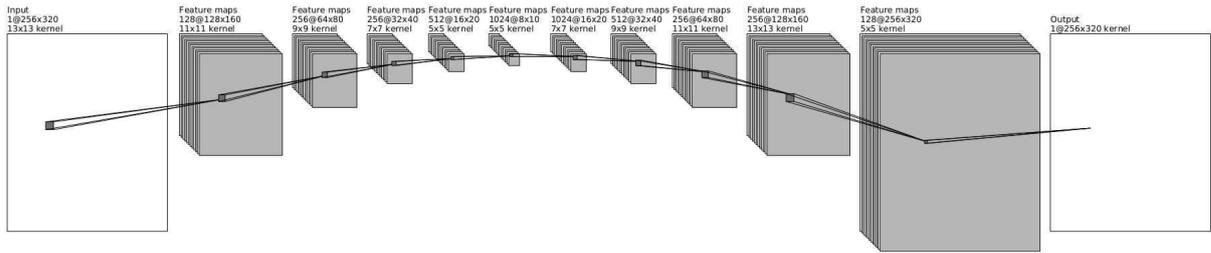


Figura 15: Arquitetura proposta na pesquisa de Svoboda et al. (2017). Fonte: Svoboda et al. (2017).

Li et al. (2017) apresentaram uma abordagem baseada em aprendizagem multi-tarefa para o processo de realce de imagens de impressão digital latente. A rede neural convolucional proposta possui duas ramificações. A primeira ramificação, chamada de *enhancement deconvolution branch*, é responsável pela remoção do ruído estruturado e pelo realce das imagens. A segunda ramificação, a *orientation deconvolution branch* ajuda a guiar o processo de realce utilizando as informações de orientação extraídas da impressão digital. Essa arquitetura é mostrada na Figura 16. Para o treinamento da rede, os autores utilizaram a base de dados de impressões digitais reais NIST SD4². Os autores consideraram esta como uma base de dados com imagens de boa qualidade. Diferentemente da utilização de uma base de dados sintética gerada por algum *software* especializado, essa abordagem necessita da confiabilidade em um algoritmo de realce de terceiros para a obtenção das saídas esperadas (*ground truths*). O algoritmo selecionado para essa tarefa foi o apresentado por Hong et al. (1998). As imagens para treinamento foram pré-processadas e delas foram feitos recortes de partes (*patches*), de tamanho 109×109 pixels, permitindo-se sobreposição. Foram adicionados ruídos como ruído gaussiano, caracteres e linhas a esses *patches* e também foi realizado um procedimento de *Total Variation Decomposition* (TVD) (Yin et al., 2007; Zhang et al., 2013) nas imagens ruidosas para se obter os componentes *cartoon* e textura. O componente textura resultante é o que de fato foi utilizado como entrada para o treinamento da rede.

TVD é um método para decompor uma imagem real na soma de dois componentes constituintes. Conforme (Yin et al., 2007), uma imagem I pode ser decomposta como sendo $I = t + c$. O componente t é considerado uma estrutura repetitiva e significativa de pequenos padrões e/ou ruídos aleatórios não correlacionados, denominada de textura. O restante da imagem, constituído por objetos e bordas, é chamado de componente *cartoon* c .

Uma proposta semelhante ao que foi apresentado na pesquisa de Li et al. (2017), no que diz respeito à preparação dos dados de treinamento, é vista na pesquisa de Qian et al. (2019). A abordagem de treinamento também se baseia na utilização de *patches*

²Atualmente a base de dados NIST SD4 não está mais disponível por não possuir a documentação exigida pelo NIST para a distribuição. Fonte: <https://www.nist.gov/srd/nist-special-database-4>

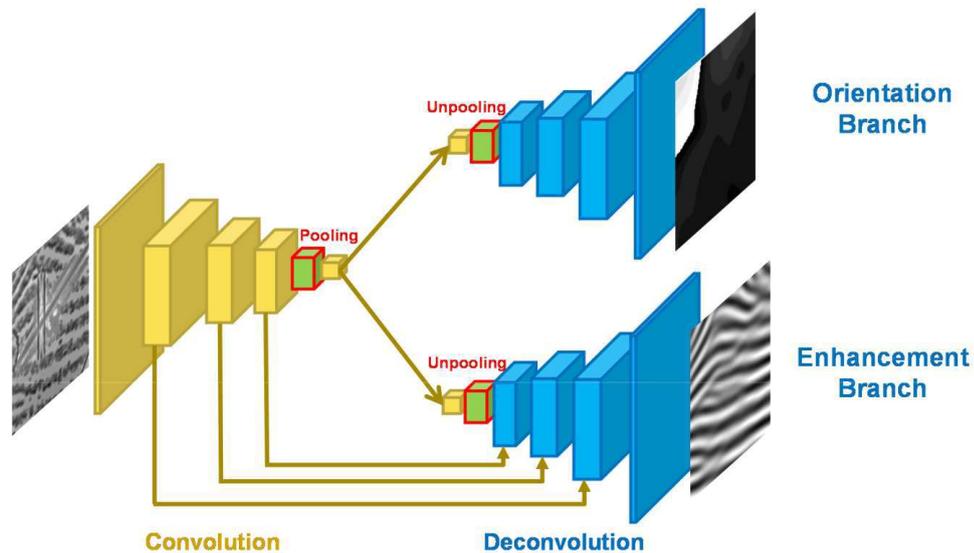


Figura 16: Arquitetura proposta na pesquisa de Li et al. (2017). Fonte: Li et al. (2017).

de imagens de impressão digital, também obtidos por meio de extração de partes das imagens da base NIST SD4. De maneira análoga, foram adicionados caracteres, linhas e ruído estrutural aos *patches* extraídos. Os ruídos estruturais foram obtidos da extração do componente *cartoon*, após a aplicação do método de composição de variação total, das imagens do banco de impressões digitais latentes NIST SD27. Com relação à preparação dos dados de treinamento, a principal diferença é que as saídas esperadas das imagens utilizadas no treinamento foram obtidas por meio da aplicação de filtros de Gabor. Como arquitetura da rede neural, que pode ser visualizada na Figura 17, os autores propuseram um modelo densamente conectado por meio da adição de blocos *dense* e *skip connections* à arquitetura original da U-net e chamaram o método de DenseUnet. A base de dados NIST SD27 também foi utilizada para realização dos testes de realce e comparação de melhorias no casamento entre as imagens de impressão digital. No artigo, foram realizadas comparações com outros métodos, focando nas melhorias em relação a métodos baseados em estimação da orientação, dicionários localizados e filtros de Gabor. O único método que foi comparado e que utiliza redes neurais foi o método proposto por Li et al. (2017).

Wong e Lai (2020) desenvolveram um método que aplica uma rede convolucional multi-tarefa para realizar a restauração de imagens de impressão digital corrompidas. A CNN proposta tem por objetivo reconstruir a estrutura de cristas da impressão digital e o mapa de orientação simultaneamente, utilizando as informações de orientação para melhorar a qualidade da reconstrução da estrutura das cristas. Para tal, a imagem de entrada passa por duas redes base independentes para gerar a estrutura de cristas limpa e o mapa de orientação. Em seguida, as saídas das duas redes são concatenadas e utilizadas para alimentar outras cinco camadas convolucionais e assim gerar a saída final.

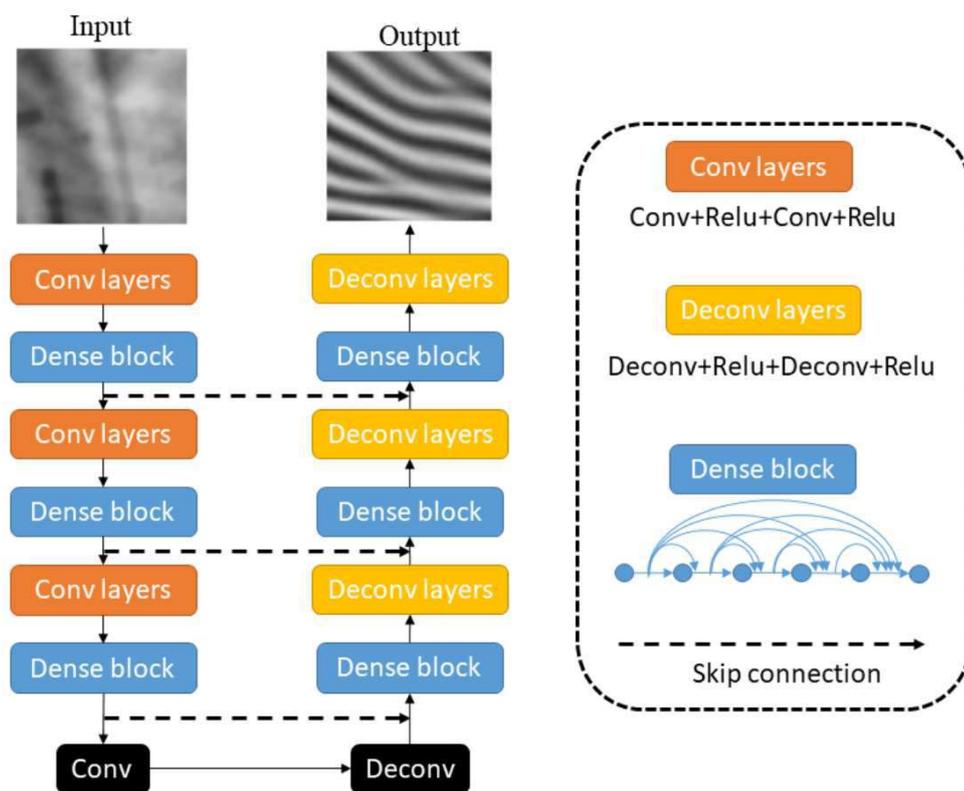


Figura 17: Arquitetura proposta na pesquisa de Qian et al. (2019). Fonte: Qian et al. (2019).

A arquitetura da rede neural proposta é mostrada na Figura 18. Para o treinamento da rede neural foi criada uma base de dados sintética, utilizando a versão de demonstração do programa SFinGe (Maltoni, 2004). O programa foi empregado para obter impressões digitais sintéticas e as respectivas saídas esperadas. Para simulação das digitais latentes foram aplicados ruídos às imagens geradas pelo SFinGe com o propósito de emular deteriorações comuns existentes em impressões digitais corrompidas. Os testes de desempenho da rede foram realizados principalmente nas bases de dados FVC2002 (Maio et al., 2002), FVC2004 (Maio et al., 2004) e FVC2006 (Fierrez et al., 2007). Também foram realizados testes com a base de dados de impressões digitais latentes IIIT-Delhi MOLF, porém utilizando uma metodologia diferente da adotada pela maior parte dos artigos sobre o tema que também utilizaram essa mesma base de dados.

O estudo conduzido por Liu e Qian (2020) propôs realizar a segmentação e realce de impressões digitais latentes por meio de uma arquitetura de redes U-net profundas aninhadas. Essa arquitetura é exibida na Figura 19. Nessa pesquisa, os autores utilizaram como base uma versão aprimorada da arquitetura U-net, conhecida por U-net++ (Zhou et al., 2018), desenvolvida com a proposta de substituir as conexões de atalho (*skip connections*) simples por caminhos de atalho (*skip pathways*) densos e aninhados. Para criar a base de dados de treinamento, foram selecionadas manualmente 300 imagens de boa qualidade da base de dados de impressões digitais NIST SD14. Entretanto, não

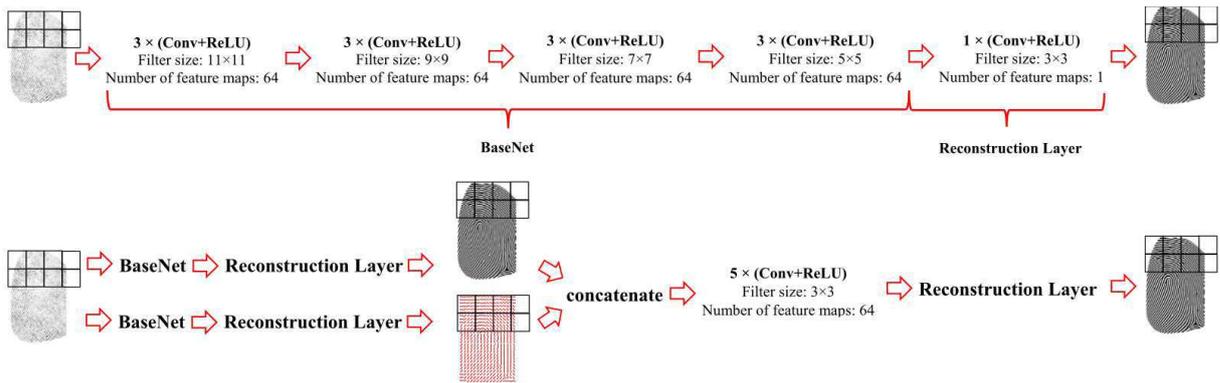


Figura 18: Arquitetura proposta na pesquisa de Wong e Lai (2020). Fonte: Wong e Lai (2020).

foi especificado pelos autores os critérios de qualidade adotados para definir a escolha dessas imagens. Algumas pesquisas optam por adotar a métrica NFIQ ou NFIQ2 para esses casos. Para as imagens selecionadas foram adicionados diferentes planos de fundo e ruído. O ruído adicionado às imagens foi obtido por meio do componente *cartoon* após a aplicação da técnica de decomposição de variação total em imagens de impressão digital latente reais da base de dados NIST SD27. As imagens de saída desejada foram obtidas por meio da aplicação de filtros de Gabor. Um total de 30,000 imagens foram usadas para o treinamento da rede. As bases de dados utilizadas para testes de desempenho do método foram a NIST SD27 e a IIIT-Delhi MOLF. Porém, também não há no artigo uma comparação com métodos de realce baseados em arquiteturas de DL, apenas a pesquisa de Li et al. (2017) é citada na comparação.

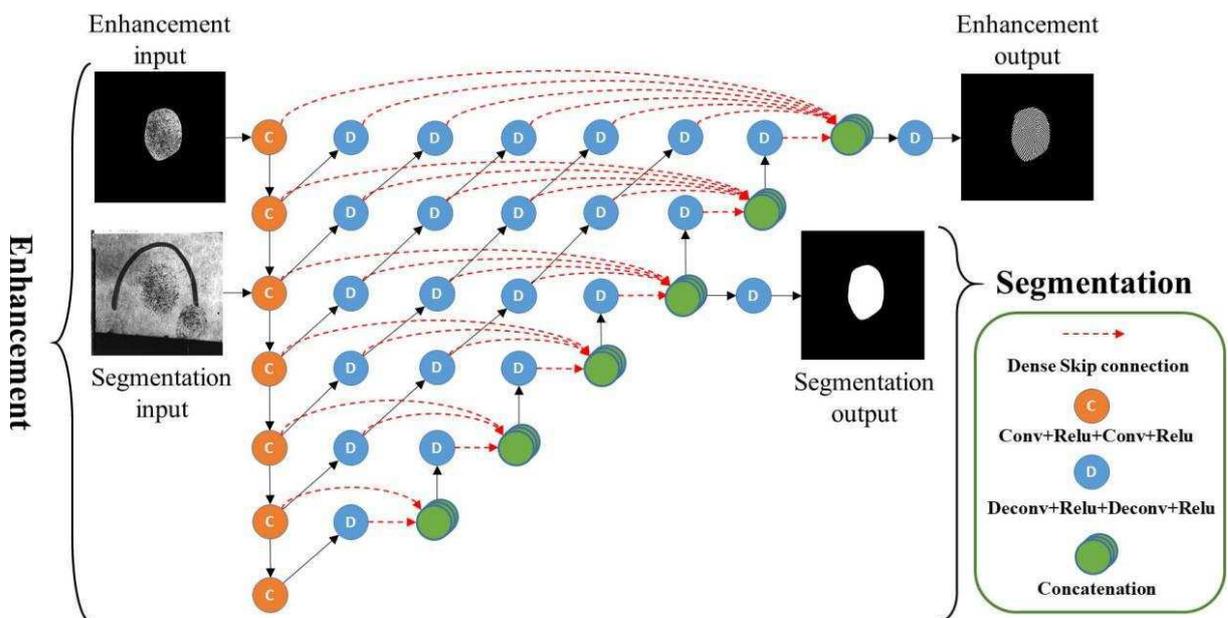


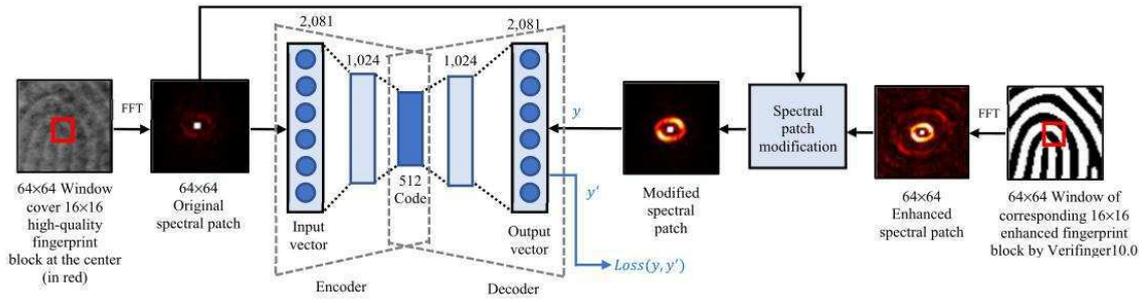
Figura 19: Arquitetura proposta na pesquisa de Liu e Qian (2020). Fonte: Liu e Qian (2020).

Um estudo recente, publicado por Horapong et al. (2021), adota um método de

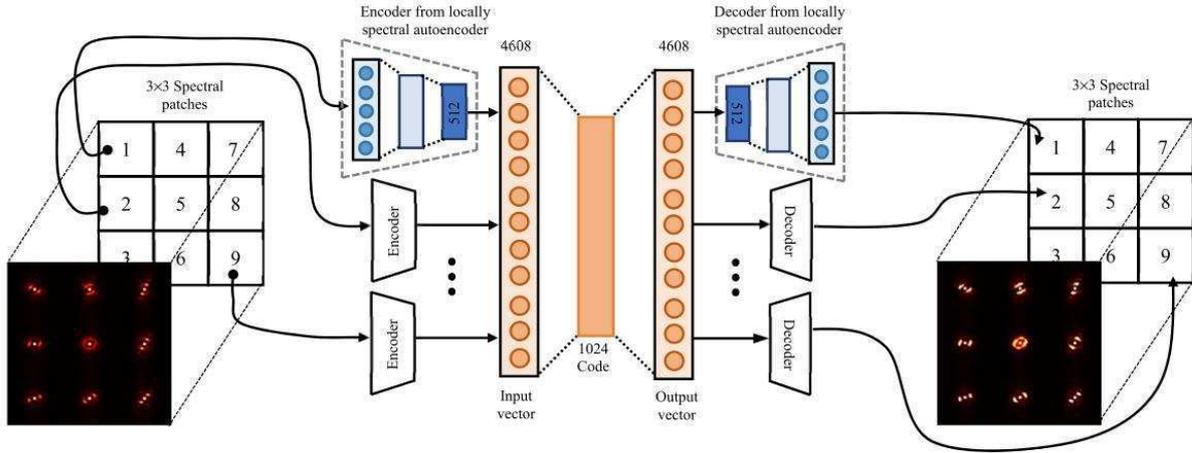
feedback para o problema de realce de imagens de impressão digital latente. O *framework* apresentado consiste em dois mecanismos principais. O primeiro mecanismo aprimora áreas de alta qualidade e realimenta as áreas aprimoradas para melhorar a qualidade das impressões digitais latentes nas áreas próximas. O segundo mecanismo tem como função verificar se o primeiro mecanismo funciona corretamente, detectando padrões de impressão digital anormalmente aprimorados. Isto é feito por meio de um autocodificador espectral que aprende fazendo uso de espectros de impressão digital no domínio da frequência. Áreas de impressão digital anômalas são enviadas de volta ao primeiro mecanismo para contribuir com a melhora do resultado aprimorado anteriormente. Entretanto, o método exposto tem como desvantagem o fato de necessitar de uma segmentação manual das impressões digitais que são dadas como entrada para o *framework*. A arquitetura desenvolvida, juntamente com o fluxo de treinamento da rede neural, são mostrados na Figura 20. O AE espectral usa um par de entrada-saída de *patches* espectrais correspondentes, de dimensão 64×64 . Os *patches* espectrais são obtidos por meio da aplicação da transformada rápida de Fourier em blocos de tamanho de 64×64 pixels de impressões digitais de alta qualidade da base de dados NIST SD14. Também não foram reportados os critérios de seleção das imagens com as melhores qualidades. Para testes foram usadas as bases NIST SD27 e IIIT-Delhi MOLF. Como resultados de teste de casamento das impressões digitais da base NIST SD27, foram reportados os valores, no *rank-30*, de 68,60% e 33,72% usando, respectivamente, os algoritmos de casamento Verifinger SDK e o *Minutia Cylinder Codes* (MCC) (Cappelli et al., 2010), enquanto que na base MOLF os valores obtidos foram de 43,84% e 27,64%, respectivamente.

3.2 Métodos baseados em redes generativas

Um modelo baseado em redes generativas adversárias (GANs, do inglês *Generative Adversarial Networks*) foi proposto por Dabouei et al. (2018). Com o intuito de fazer com que a rede neural proposta preserve ao máximo as informações de identificação contidas nas imagens das impressões digitais e também para evitar que áreas muito ruidosas das imagens de entrada sejam preenchidas com informações errôneas durante o processo de reconstrução, foram propostas duas modificações na arquitetura convencional de uma GAN condicional (cGAN, do inglês *Conditional Generative Adversarial Network*). A primeira foi treinar o gerador para criar quatro mapas: o realce, chamado pelos autores de mapa de cristas, o mapa de frequência, o mapa de orientação e o mapa de segmentação. A segunda modificação foi a adição de uma rede convolucional siamesa ao discriminador da GAN, chamada de *Perceptual ID Information* (PIDI) *extractor*. Essa rede foi treinada separadamente para a extração da informação perceptual de identificação dos mapas produzidos pelo gerador, de modo que as informações extraídas são concatenadas com as informações existentes em camadas correspondentes do discriminador para que assim sejam levadas



(a). Locally spectral autoencoder architecture with training dataflow.



(b). Regionally spectral autoencoder architecture with input-output dataflow.

Figura 20: Arquitetura proposta e fluxo de treinamento da rede neural da pesquisa de Horapong et al. (2021). (a) é o *locally spectral autoencoder* que possui o aprendizado baseado em *patches* espectrais de tamanho 64×64 . (b) é chamado de *regionally spectral autoencoder*, que combina 9 *locally spectral autoencoders* para entradas e saídas. Fonte: Horapong et al. (2021).

em consideração e enfatizadas durante a etapa de treinamento. O modelo desenvolvido é apresentado na Figura 21. A base de dados de treinamento foi criada a partir da distorção e adição de ruído e de diferentes planos de fundo em imagens de impressão digital da base de dados BioCOP 2013³. Como base de dados de teste foram utilizadas as bases IIIT-Delhi *latent database*, com a qual foi obtido um resultado de 94,67% na acurácia *rank-25*, e a IIIT-Delhi MOLF, com a qual foi obtido o resultado de 70,89%, 66,11% e 68,61% nas acurácias *rank-50* das bases de testes. Essa é a pesquisa com os melhores resultados, em termos de acurácia *rank-50*, encontrados. Entretanto, informações a respeito da base de dados utilizada para gerar os dados sintéticos para treinamento da rede, bem como as técnicas utilizadas para gerar os *ground truths* são apresentados superficialmente e de forma resumida. A abordagem de treinamento da rede PIDI *extractor* também não é relatada em detalhes no artigo. Essa informação seria de grande valia, uma vez que esse modelo necessitou de um treinamento à parte e é a rede PIDI que potencializa os resultados da cGAN proposta.

³<http://biic.wvu.edu/>

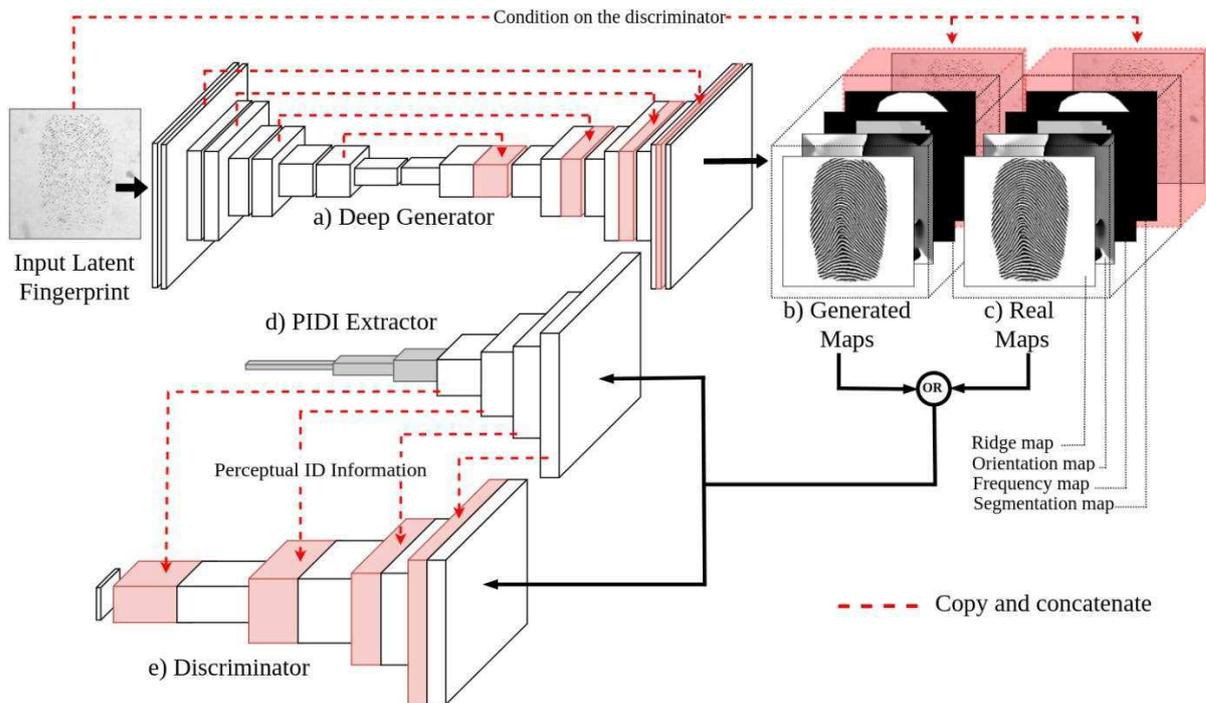


Figura 21: Arquitetura proposta na pesquisa de Dabouei et al. (2018). a) rede Gerador que recebe a impressão digital latente como entrada e gera os mapas como saída. b) mapas gerados, que são concatenados com a impressão digital latente de entrada e fornecidos como condição para a rede Discriminador. c) mapas reais são extraídos das impressões digitais originais. d) rede siamesa profunda que foi treinada separadamente para receber os mapas gerados ou reais e fornecer as representações da rede PIDI para o discriminador. e) Discriminador que objetiva distinguir entre os mapas gerados e os mapas reais usando o ID combinado e as informações dos mapas gerados Fonte: Dabouei et al. (2018).

Joshi et al. (2019) aplicaram um modelo que utiliza GANs condicionais para a realização da tarefa do realce das impressões digitais latentes. O modelo proposto consiste de duas redes que os autores denominaram como: *latent enhancer network* ($\mathcal{E}nh_L$) e *enhanced fingerprint discriminator network* (Dis_E). $\mathcal{E}nh_L$ foi treinada para fornecer como saída uma versão realçada de uma dada imagem de impressão digital latente x , enquanto que Dis_E foi treinada para diferenciar se uma imagem de entrada y é a imagem realçada de uma impressão digital latente real ou se é uma imagem gerada pelo $\mathcal{E}nh_L$. A arquitetura proposta está ilustrada na Figura 22. Para o treinamento da rede foi criada uma base de dados sintética utilizando o Anguli⁴ (Ansari, 2011), uma implementação *open source* fundamentada no algoritmo do SFinGe (Cappelli et al., 2002). Foram adicionados às imagens geradas diferentes tipos de ruídos e diferentes planos de fundo na tentativa de representar características encontradas nas imagens de impressões digitais latentes reais. Dentre os trabalhos encontrados até o presente momento da revisão bibliográfica, a abordagem

⁴<https://dsl.cds.iisc.ac.in/projects/Anguli/>

apresentada nesse artigo é a que mais focou na capacidade de generalização do modelo. Apesar da quantidade relativamente baixa de imagens utilizadas no treinamento, 8.423, a forma como o banco de dados sintético de treinamento foi criado apresentou resultados finais que ratificam a capacidade do método performar bem em impressões digitais latentes presentes em diferentes contextos. Esta foi a única pesquisa encontrada até o presente momento que apresentou resultados do seu método para a base de impressões digitais latentes multi superfície IIIT-Delhi MSLFD (Sankaran et al., 2015a). As bases de dados de testes utilizadas no trabalho foram a IIIT-Delhi MSLFD e a IIIT-Delhi MOLF com respectivas acurácias *rank-50* de 15,24% e 35,66% respectivamente.

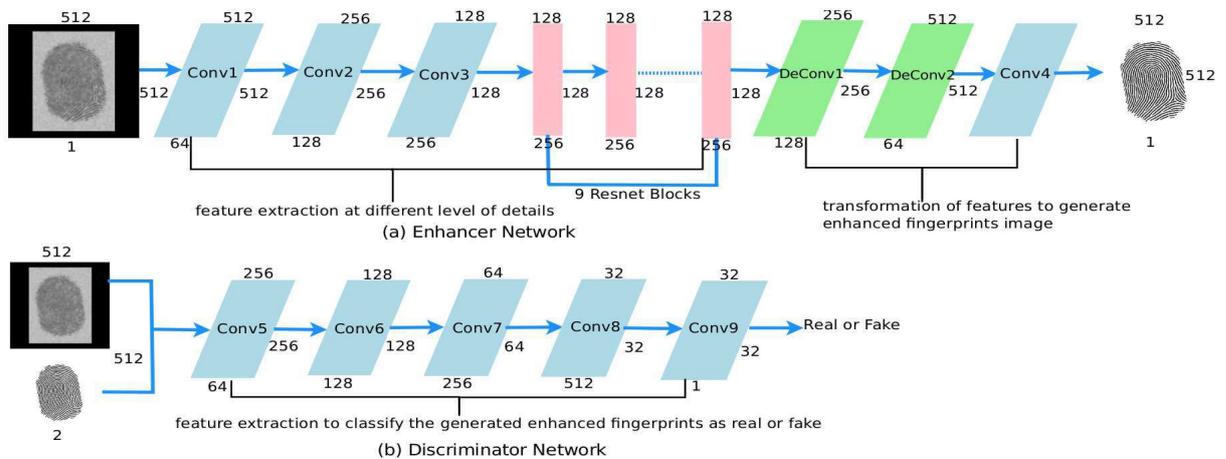


Figura 22: Arquitetura proposta na pesquisa de Joshi et al. (2019). (a) rede $\mathcal{E}nh_L$ treinada para fornecer como saída o realce de uma imagem de impressão digital latente recebida como entrada. (b) rede Dis_E treinada para diferenciar se uma imagem de entrada é a imagem realçada de uma impressão digital latente real ou se é uma imagem gerada pela rede $\mathcal{E}nh_L$. Fonte: Joshi et al. (2019).

Considerando que a orientação das cristas e vales é uma característica intrínseca de uma impressão digital, Liu et al. (2019) apresentaram uma rede denominada *cooperative orientation generative adversarial network* (COOGAN). A arquitetura da rede, ilustrada na Figura 23, foi projetada para mapear o domínio de origem das imagens de impressão digital em um espaço latente em que orientação e realce compartilham a mesma representação. Adicionalmente, com o intuito de aprimorar mais o realce, três módulos adicionais foram acrescentados ao gerador: módulo de orientação, módulo de segmentação e módulo de estimativa de qualidade. Este último possui a finalidade de remover partes irrecuperáveis nas imagens. Os dados de treinamento da rede foram criados a partir da adição de distorções em imagens de impressão digital da base NIST SD14⁵. Foi utilizado o programa Fingernet (Li et al., 2017) para obtenção das saídas esperadas desses dados. Um

⁵Atualmente a base de dados NIST SD14 não está mais disponível por não possuir a documentação exigida pelo NIST para a distribuição. Fonte: <https://www.nist.gov/srd/nist-special-database-14>

diferencial no treinamento da GAN apresentada nessa pesquisa é que, antes do fornecimento dos *ground truths* como entrada para o discriminador, é realizado o alinhamento e concatenação com suas correspondentes versões roladas, que são impressões digitais em que é feito o rolamento do dedo de uma extremidade à outra sobre a superfície de captura, no momento da aquisição. O teste de desempenho de melhoria no resultado do casamento de impressões digitais foi realizado utilizando a base de dados NIST SD27 e o algoritmo de casamento do software comercial *VeriFinger SDK*.

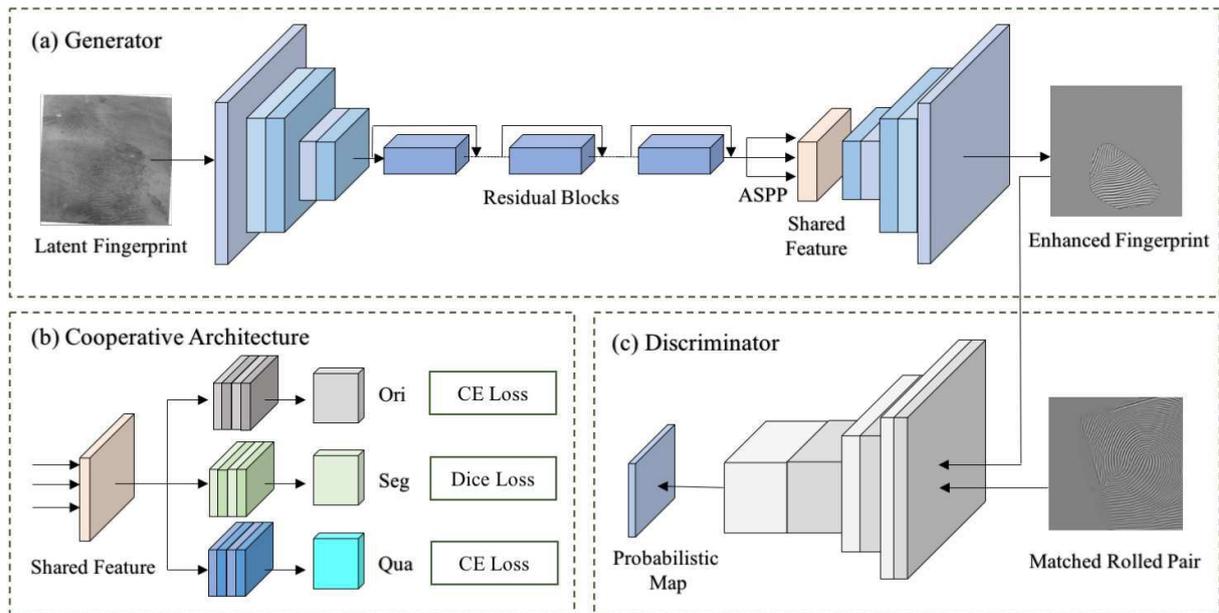


Figura 23: Arquitetura proposta na pesquisa de Liu et al. (2019). (a) Gerador que mapeia a impressão digital latente para a imagem realçada. (b) arquitetura cooperativa em que os módulos específicos de tarefas orientação, segmentação e qualidade, compartilham a mesma representação. (c) Discriminador que visa distinguir se uma amostra é uma imagem real ou gerada pelo Gerador. Fonte: Liu et al. (2019).

Um método para a realização de realce de imagens de impressão digital latente baseado em redes generativas adversárias progressivas (PGANs, do inglês *Progressive Generative Adversarial Networks*) foi proposto por Huang et al. (2020). A arquitetura desenvolvida é mostrada na Figura 24. O método desenvolvido foi dividido em dois estágios: Treinamento Progressivo *Offline* (POT) e Teste Iterativo *Online* (IOT). No estágio POT, a rede é treinada em uma base de dados criada sinteticamente com a adição de ruídos de modo a simular características de impressões digitais latentes pareadas com suas respectivas impressões digitais em alta qualidade. A base de dados NIST SD14 foi utilizada para gerar os dados sintéticos de treinamento. Seguindo a mesma tendência de vários artigos que utilizaram bases de imagens de impressão digitais reais para gerar o conjunto de treinamento, a técnica de decompor imagens de impressões digitais latentes da base NIST SD27 em componentes *cartoon* e textura, por meio de decomposição de variação total, também foi utilizada. Assim, foram escolhidas manualmente 200 imagens

da base NIST SD14 e a elas somados diferentes tipos de ruídos estruturados obtidos dos componentes *cartoons* resultantes. Os autores não forneceram informações sobre como os *ground truths* foram obtidos. No total foram geradas 30.000 pares de imagens para o treinamento da rede neural. Durante o treinamento é feito o crescimento progressivo da GAN para aumentar a estabilidade e melhorar o desempenho da rede neural. No estágio IOT, a impressão digital latente segmentada é inserida na GAN treinada no estágio anterior. Esse pode ser considerado um ponto negativo do método, uma vez que também é necessário passar como entrada para a rede o mapa de segmentação da impressão digital, e esse processo é muitas vezes realizado de forma manual. O conjunto de dados de teste foi o NIST SD27, utilizando como extrator de minúcias e algoritmo de casamento o sistema Verifinger SDK.

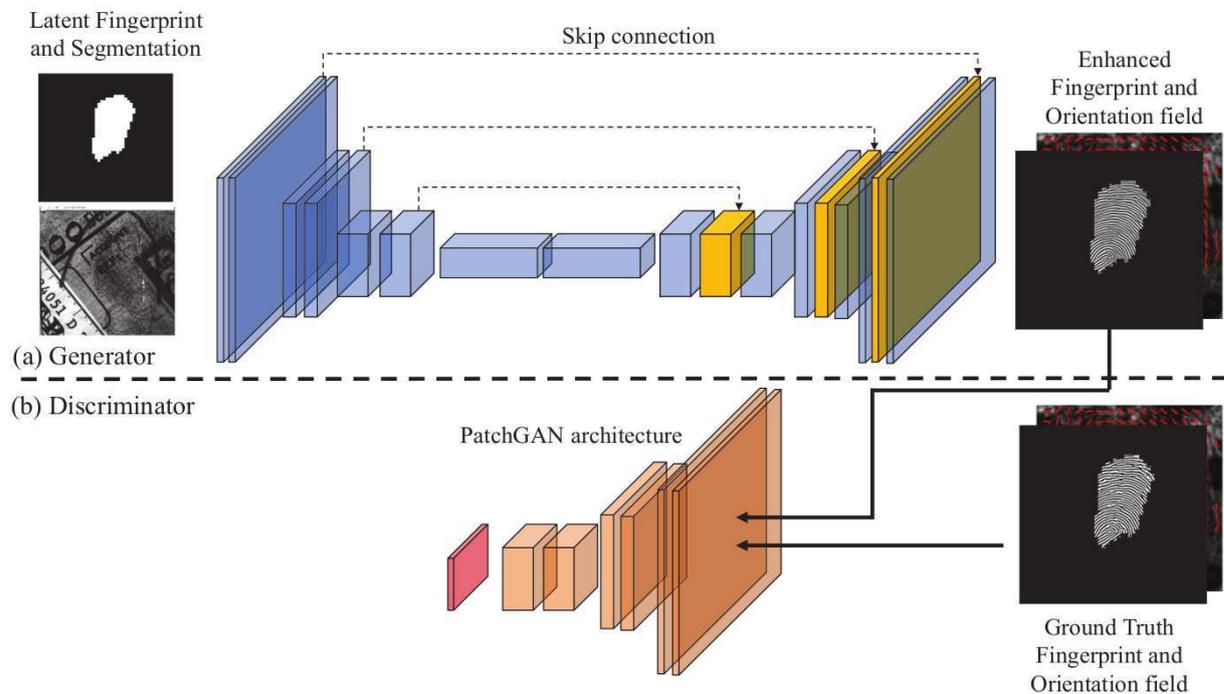


Figura 24: Arquitetura proposta na pesquisa de Huang et al. (2020). (a) Gerador treinado para fornecer como saída a imagem de impressão digital realçada e o mapa de orientação estimado. (b) Discriminador treinado para distinguir se um realce e mapa de orientação recebidos são de uma impressão digital latente real ou se foi gerado pelo Gerador. Fonte: Huang et al. (2020).

Xu et al. (2020) criaram um esquema de aumento de dados baseado em GANs para melhorar a reconstrução de imagens de impressão digital latente. Na abordagem apresentada, impressões digitais de boa qualidade são binarizadas e em seguida transformadas em uma versão de impressão digital latente correspondente. Para isso, o método modela a distribuição dos padrões das degradações encontradas em impressões digitais latentes em uma distribuição Gaussiana e gera as impressões digitais latentes sintéticas por meio da amostragem desses padrões e das impressões de boa qualidade. Também foi desenvolvida outra rede neural treinada para gerar a versão correspondente binarizada de

impressões digitais latentes. A arquitetura da GAN proposta é exibida na Figura 25. As bases de dados utilizadas para a criação dos dados de treinamento do modelo de aumento de dados foram as NIST SD14, e MOLF DB1, DB2 e DB3. A avaliação do desempenho das redes foi realizada nas bases de dados NIST SD27, IIITD e MOLF-DB4 obtendo como resultados de acurácia *rank-25* os valores de 82,17%, 95.12%, e 45,88%, respectivamente. Para todos os testes foi utilizado como extrator de minúcias e algoritmo de casamento o Verifinger SDK.

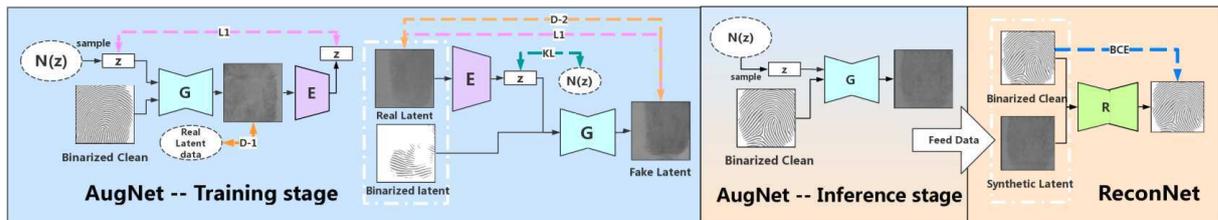


Figura 25: Arquitetura proposta na pesquisa de Xu et al. (2020). Fonte: Xu et al. (2020).

Um método ponto-a-ponto híbrido é apresentado na pesquisa de Deshpande et al. (2020). O método proposto possui três fases: um pré-processamento que utiliza a aplicação de filtros de Gabor, juntamente com algumas camadas da rede convolucional Fingernet de Tang et al. (2017), uma rede neural desenvolvida para extração de minúcias em impressões digitais, para obtenção do mapa de frequência realçado. A arquitetura FingerNet está ilustrada na Figura 26. Uma observação relevante em relação à dependência do Fingernet em termos de realce é que, nos casos em que essa rede não apresente bom desempenho, a eficácia do método proposto pode ser comprometida. O mapa de frequência aprimorado obtido é realçado ainda mais por meio da aplicação de filtros da Transformada Rápida de Fourier (FFT, do inglês *Fast Fourier Transform*). O *framework* apresentado nessa pesquisa tem a proposta de ser uma abordagem completa. Dessa forma, também foram implementados um módulo para extrair as minúcias e outro para realizar o teste de correspondência das imagens realçadas. Os testes foram realizados nas bases de dados de impressões digitais FVC2002 e FVC2004 e na base de dados de impressões digitais latentes NIST SD27. Os resultados relatados foram de 100% de acurácia no *rank-1* para as bases FVC e de 85,5% para a NIST SD27.

Nesta pesquisa, Zhu et al. (2023) propõem um método chamado FingerGAN para aprimoramento de impressões digitais latentes utilizando como imagens de referência o mapa de esqueleto da impressão digital e o mapa de orientação. Diferentemente da maioria dos métodos de realce de impressões digitais latentes, que tentam restaurar as cristas/vales cinza corrompidas, a GAN proposta foca em otimizar a informação de minúcias. O método proposto utiliza um modelo de orientação de impressão digital baseado em expansão de Fourier bidimensional (FOMFE, do inglês *Fingerprint Orientation Model Based on 2D Fourier Expansion*) Wang et al. (2007) para regularizar o campo de orientação da

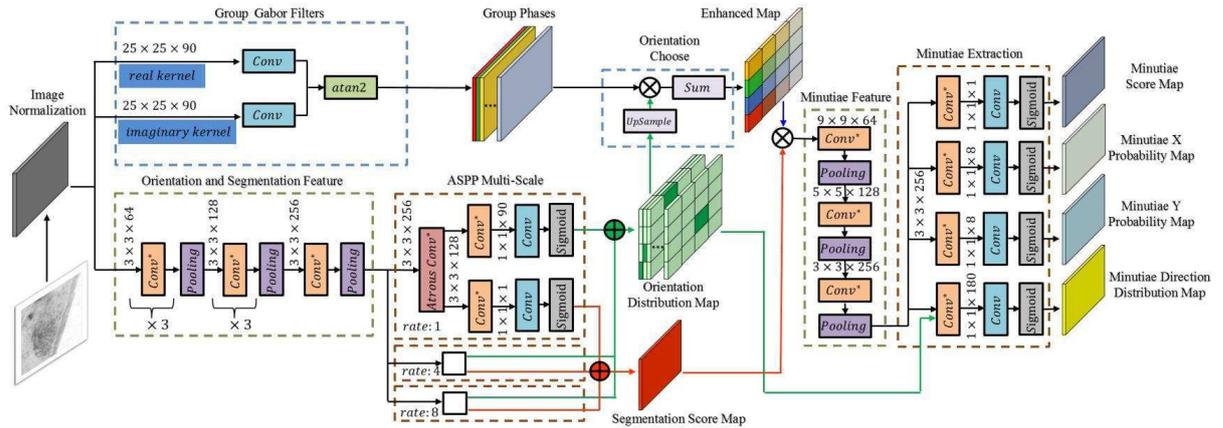


Figura 26: Arquitetura do FingerNet utilizada na pesquisa de Deshpande et al. (2020). Fonte: Tang et al. (2017).

impressão digital aprimorada. O modelo FOMFE descreve a geometria global das cristas de uma impressão digital e é usado para modelar suas orientações. De acordo com os autores, é um campo de orientação regularizado que é mais confiável contra ruído e funciona bem para impressões digitais de baixa qualidade. As funções de custo aplicadas ao treinamento da rede foram a MAE e a *adversarial loss*. O conjunto de dados de treinamento da FingerGAN foi criado utilizando imagens da base de dados NIST SD14 e as bases de dados de testes utilizadas foram a NIST SD27 e a MOLF. Na Figura 27 é mostrada a arquitetura da rede FingerGAN.

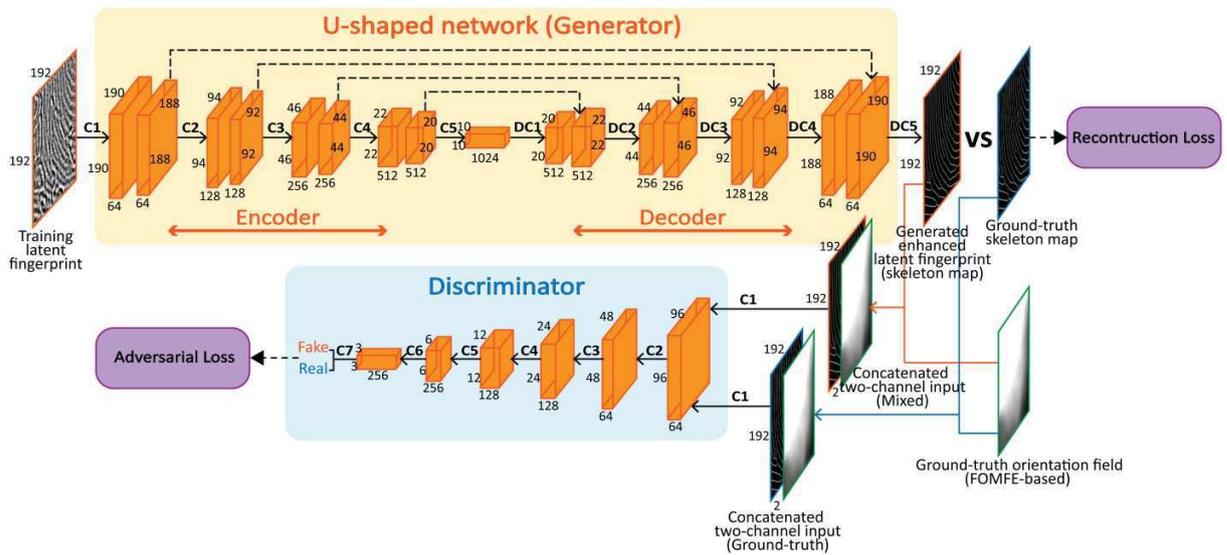


Figura 27: Arquitetura da FingerGAN desenvolvida na pesquisa de Zhu et al. (2023). Fonte: Zhu et al. (2023).

A pesquisa conduzida por Joshi et al. (2022) apresenta um mecanismo de atenção leve para refinamento de canal em modelos de restauração de impressão digital. O mecanismo proposto é baseado na ideia de reduzir a redundância entre os pesos dos canais dos kernels convolucionais. Essa redução de redundância é realizada por meio de uma

ponderação global entre os canais, com base em sua importância para a reconstrução da estrutura de cristas da impressão digital. O mecanismo de atenção proposto foi avaliado em três bancos de dados públicos de impressões digitais: IIITD-MOLF, Rural Indian Fingerprint Database e um banco de dados privado de impressões digitais rurais. Os dados de treinamento da rede proposta foram criados seguindo a pesquisa de Joshi et al. (2019). A avaliação de desempenho do método foi realizada com base nas pontuações NFIQ, para avaliação de qualidade, e utilizando os métodos de casamento de impressões digitais Bozorth e MCC. A arquitetura da CR-GAN desenvolvida é ilustrada na Figura 28.

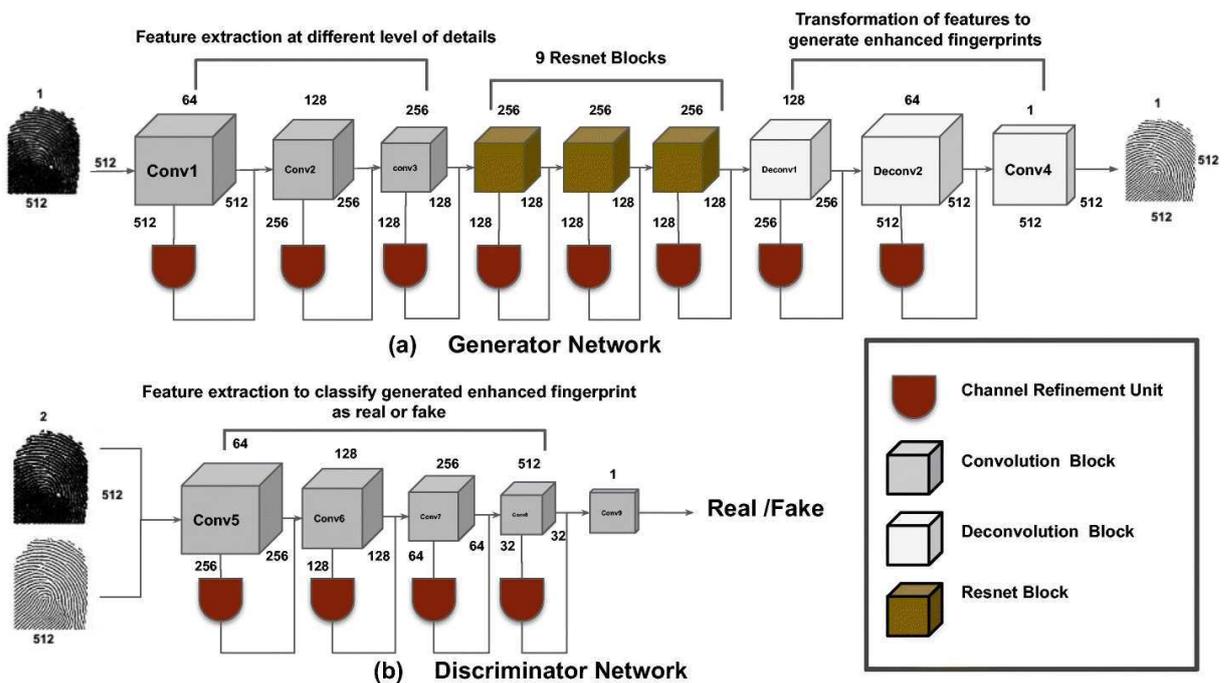


Figura 28: Arquitetura da CR-GAN desenvolvida na pesquisa de Joshi et al. (2022). Fonte: Joshi et al. (2022).

No quadro Quadro 2 contém uma visualização resumida das pesquisas, abordagens adotadas, bases de dados utilizadas e melhores resultados informados pelos estudos citados neste capítulo. Com base nas informações do quadro, é possível observar que tanto as abordagens baseadas em redes generativas adversariais quanto as abordagens baseadas em redes convolucionais apresentaram resultados promissores no campo do realce de impressões digitais latentes. Houve um período em que houve uma tendência em direção às GANs, conforme demonstrado pelas pesquisas mencionadas no quadro. No entanto, pesquisas mais recentes que utilizaram redes convolucionais e obtiveram resultados equivalentes nas principais bases de dados de testes utilizadas atualmente.

No Quadro 3 são dadas informações conceituais importantes que distinguem os trabalhos de realce de impressões digitais latentes. Além das informações conceituais, também são marcadas as funções de custo utilizadas por cada pesquisa. As funções de custo mostradas na tabela são: Erro Médio Absoluto (MAE), Erro Médio Quadrático

Quadro 2: Resumo dos trabalhos relacionados

Pesquisa	Abordagem	Melhor Acurácia	Bases de dados de teste					
			IIIT-D <i>latent</i>	NIST SD27	MOLF	RIFD	FVC	MSLFD
Svoboda et al. (2017)	CAE	79,04% no <i>rank</i> -10 para a base IIITD <i>latent</i> e 22,36% no <i>rank</i> -50 para a MOLF	X		X			
Li et al. (2017)	CNN multi-tarefa	Valores exatos não informados		X				
Dabouei et al. (2018)	GAN + redes siamesas	70,89% no <i>rank</i> -50 na base MOLF	X		X			
Joshi et al. (2019)	GAN	35,66% no <i>rank</i> -50 para a base MOLF			X			X
Liu et al. (2019)	GAN	Valores exatos não informados		X				
Qian et al. (2019)	DenseUNet	Valores exatos não informados		X				
Wong e Lai (2020)	CNN multi-tarefa	89,89% no <i>rank</i> -50 para a base MOLF. Porém, foi adotado um protocolo de avaliação diferente das demais pesquisas			X		X	
Huang et al. (2020)	PGAN	Valores exatos não informados		X				
Xu et al. (2020)	GAN	45,88% no <i>rank</i> -25 para a base MOLF	X	X	X			
Liu e Qian (2020)	U-nets aninhadas	Valores exatos não informados		X	X			
Deshpande et al. (2020)	CNN + filtros de Gabor + filtros FFT	85,5% no <i>rank</i> -1 na NIST SD27		X	X		X	
Horapong et al. (2021)	AE	43,68% no <i>rank</i> -30 para a base MOLF		X	X			
Zhu et al. (2023)	GAN	Valores exatos não informados		X	X			
Joshi et al. (2022)	GAN	35,25% no <i>rank</i> -50 para a base MOLF			X	X		

(MSE), função de custo perceptual (percep.), função de custo adversarial (Adv.), função de custo no domínio da frequência (DCT/DFT), Entropia Cruzada ou Entropia Cruzada Binária (CE/BCE) e função de custo Índice de Similaridade Estrutural (SSIM).

Com base nas informações contidas no Quadro 3, podemos inferir que esses métodos utilizam principalmente abordagens baseadas em redes convolucionais e redes generativas adversariais para enfrentar desafios específicos no processo de realce. Esses desafios envolvem evitar a reconstrução de áreas ruidosas nas imagens de saída e preservar informações cruciais para a identificação. Alguns métodos aplicam técnicas de segmentação para identificar regiões recuperáveis ou anômalas, enquanto outros utilizam filtros de Gabor ou funções de custo específicas para aprimorar a qualidade do realce.

É importante ressaltar que nem todos os métodos abordam diretamente o problema que pode resultar da reconstrução de áreas ruidosas nas imagens realçadas. Alguns concentram-se em etapas relacionadas, como a extração de informações de orientação, a detecção de minúcias ou o apenas o aprimoramento das estruturas das cristas.

3.3 Outras pesquisas relacionadas

Na pesquisa conduzida por Yadav et al. (2021) é apresentada uma abordagem eficaz para corrigir a exposição de fotografias capturadas em condições adversas de iluminação na vida real. A técnica se baseia na observação de que imagens subexpostas apresentam

Quadro 3: Resumo de informações conceituais sobre os trabalhos apresentados nesta seção.

Método	Informações conceituais	Função de custo						
		MAE	MSE	Percep.	Adv.	DCT / DFT	CE / BCE	SSIM
Svoboda et al. (2017)	<ul style="list-style-type: none"> Arquitetura autocodificadora com 5 camadas no codificador e 5 no decodificador Minimiza as diferenças entre a imagem de entrada e o <i>ground truth</i>. Não abordado o problema específico do realce de áreas ruidosas nas imagens 		X					
Li et al. (2017)	<ul style="list-style-type: none"> Utilizam uma rede convolucional com duas ramificações, uma para orientação e outra para realce Não abordado o problema específico do realce de áreas ruidosas nas imagens 		X					
Dabouei et al. (2018)	<ul style="list-style-type: none"> Método baseado em GANs. Foca em locais da imagem com informações de identificação para evitar preenchimento incorreto de áreas ruidosas durante a reconstrução Utiliza duas redes siamesas no discriminador. Além dos <i>ground truths</i>, utiliza mapas de frequência, segmentação e orientação 	X		X	X			
Joshi et al. (2019)	<ul style="list-style-type: none"> Método baseado em GAN condicional Dados de treinamento incluem diferentes ruídos e planos de fundo, focando na generalização do método Não abordado o problema específico do realce de áreas ruidosas nas imagens 	X			X			

Quadro 3: Resumo de informações conceituais sobre os trabalhos apresentados nesta seção (continuação).

Método	Informações conceituais	Função de custo						
		MAE	MSE	Percep.	Adv.	DCT / DFT	CE / BCE	SSIM
Liu et al. (2019)	<ul style="list-style-type: none"> Método baseado em GANs. Possui módulos de orientação, segmentação e estimativa de qualidade O módulo de estimativa de qualidade visa delimitar as áreas que podem ser reconstruídas Realiza o alinhamento e concatenação das impressões digitais latentes com suas versões roladas no processo de treinamento 	X			X			
Qian et al. (2019)	<ul style="list-style-type: none"> U-net com a inclusão de camadas densas entre as camadas convolucionais Filtros de Gabor para obter os <i>ground truths</i> dos dados de treinamento Não abordado o problema específico do realce de áreas ruidosas nas imagens 		X					
Wong e Lai (2020)	<ul style="list-style-type: none"> CNN multitarefa treinada para reconstruir informações da estrutura de cristas e orientação Informações de orientação guiam o realce Foram reportados experimentos com a base MOLF, porém, impressões digitais latentes não foram o foco da pesquisa 						X	
Huang et al. (2020)	<ul style="list-style-type: none"> Aplicam uma GAN com treinamento progressivo Utiliza informações de segmentação para delimitar as áreas a serem reconstruídas Ruído de impressões digitais latentes reais extraído com a técnica TVD adicionado às impressões digitais sintéticas 	X			X		X	
Xu et al. (2020)	<ul style="list-style-type: none"> Desenvolveram um esquema de aumento de dados baseado em GANs Transformação de impressões digitais "limpas" em versões latentes, juntamente com seus <i>ground truths</i> Utiliza as imagens latentes geradas para o treinamento da rede de realce 	X			X		X	

Quadro 3: Resumo de informações conceituais sobre os trabalhos apresentados nesta seção (continuação).

Método	Informações conceituais	Função de custo						
		MAE	MSE	Percep.	Adv.	DCT / DFT	CE / BCE	SSIM
Liu e Qian (2020)	<ul style="list-style-type: none"> Método baseado em redes U-net aninhadas. Uma dedicada à segmentação e a outra ao realce Objetivo da segmentação: delimitar as áreas recuperáveis da impressão digital. Ruído de impressões digitais latentes reais extraído com a técnica TVD adicionado às impressões digitais sintéticas 		X					X
Deshpande et al. (2020)	<ul style="list-style-type: none"> Aplicação de um método híbrido Combinação dos filtros de Gabor com algumas camadas da rede Fingernet Utilização da transformada de Fourier para gerar o mapa de frequência das imagens e aprimorar a extração das minúcias 							
Horapong et al. (2021)	<ul style="list-style-type: none"> Método baseado em realimentação utilizando dois autocodificadores Uma rede realiza o realce da imagem e a outra foi treinada para identificar partes anômalas do realce Regiões anômalas detectadas são reenviadas para o primeiro mecanismo para um segundo processo de realce 		X					
Zhu et al. (2023)	<ul style="list-style-type: none"> Desenvolveram um método baseado em GANs Utilização de mapas de esqueleto e orientação como referências Foco na reconstrução de áreas que contenham informações identificáveis de minúcias 	X			X			
Joshi et al. (2022)	<ul style="list-style-type: none"> Desenvolveram um método baseado em GANs Inclui um mecanismo de atenção que reduz a redundância entre os pesos dos canais dos <i>kernels</i> convolucionais Não abordado o problema específico da reconstrução de áreas ruidosas nas imagens realçadas 	X			X			
RM-net desenvolvida nesta pesquisa	<ul style="list-style-type: none"> Método baseado em uma rede neural convolucional M-Net, aprimorada por meio de blocos residuais Trata o problema de reconstrução errônea de partes muito ruidosas nas imagens de saída aplicando função de custo no domínio da frequência. 	X				X		

padrões distintos nas frequências altas devido ao ruído, enquanto imagens adequadamente expostas são dominadas por frequências baixas, que compõem o conteúdo relevante. A solução proposta utiliza uma função de custo multi-escala no domínio de frequência, combinada com a função de custo MAE tradicional, para treinar um *framework* de correção de exposição. Essa função é diferenciável, permitindo o treinamento ponta a ponta por meio de *backpropagation*. Além de lidar eficientemente com imagens JPEG subexpostas e ruidosas, a função de perda também é genérica, podendo ser integrada facilmente em outros *frameworks* de correção de exposição. Essa é a função de perda que foi adaptada para a RM-net desenvolvida nesta tese.

Cao et al. (2019) apresentaram o MSU *Latent AFIS*, um sistema completo de identificação de impressões digitais latentes, abrangendo desde a segmentação de regiões de interesse (Nguyen et al., 2018a) até a extração de características (Nguyen et al., 2018b) e comparação (Cao e Jain, 2019). Nesse sistema, destaca-se o aprimoramento das impressões digitais latentes, realizado por meio de um AE convolucional. Para treinar o AE, foram degradadas imagens de boa qualidade, selecionadas baseadas na métrica NFIQ, para criar os pares de treinamento. Esse processo de degradação envolveu a divisão das imagens em sub-regiões sobrepostas, adição de ruído gaussiano e filtragem gaussiana. É importante mencionar que o AE foi treinado principalmente em impressões roladas, mas foi adaptado com sucesso para melhorar impressões digitais latentes, considerando apenas o componente de textura, obtido por meio de TVD. Essas melhorias resultaram em um sistema eficaz para aprimoramento e busca de impressões digitais latentes, testado em duas diferentes bases de dados, NIST SD27 e WVU. O resultados obtidos foram reportados em termos de acurácia *rank-1* sendo, respectivamente 64,7% na base NIST SD27 e 75,3% na base WVU. O código do MSU *Latent AFIS* está disponível publicamente no endereço: <https://github.com/prip-lab/MSU-LatentAFIS>. Entretanto, o sistema foi desenvolvido com bibliotecas que, atualmente, já foram descontinuadas e o código já não está mais sendo atualizado ou mantido. Além disso, o código não está estruturado de modo a permitir a utilização direta das etapas de pré-processamento e aprimoramento da impressão digital de forma independente. Para isso, é necessário isolar e organizar as funções e estruturas em um módulo a parte. A documentação existente não oferece detalhes suficientes para identificar essas funcionalidades, exigindo uma análise cuidadosa para atualização e adaptação do código. Portanto, a comparação das funcionalidades de melhoria da impressão digital do MSU *Latent AFIS* com a RM-net proposta nesta tese é considerada em trabalhos futuros.

3.4 Considerações finais sobre o capítulo

Neste capítulo foram apresentadas pesquisas relevantes da literatura que estão relacionadas ao tema desta tese de doutorado. Dentre os trabalhos citados, é possível perceber

que existe em comum a necessidade da criação de uma base de dados sintética de treinamento para os modelos propostos. A metodologia de criação varia de pesquisa para pesquisa, mas sempre buscando, ao final, simular características encontradas nas impressões digitais latentes reais. Dentre os métodos apresentados, existem os que se baseiam em arquiteturas de redes neurais convolucionais que adotam o modelo de codificador-decodificar e aqueles que adotam a utilização de redes generativas. As principais bases de dados de impressões digitais latentes utilizados para avaliar os métodos são o MOLF DB4 e o NIST SD27. Os resultados são comumente reportados em termos de acurácia *rank-50* e, nesse contexto, o método proposto por Dabouei et al. (2018) é o que apresenta os maiores valores de acurácia.

No capítulo 4, apresentado a seguir, são relatados os materiais e ferramentas utilizados nesta pesquisa, incluindo as bases de dados de treinamento e teste, bem como os programas de geração de impressões digitais sintéticas. Além disso, são fornecidos detalhes sobre o método desenvolvido, a metodologia de treinamento e a rede neural convolucional RM-net.

4 Materiais e Método

Neste capítulo são explicados os materiais empregados e o método desenvolvido na presente pesquisa. Inicialmente, são apresentados os conjuntos de dados utilizados para validar o método proposto. Em seguida, descreve-se o processo empregado para criar uma base de dados usada no treinamento da rede neural convolucional. A arquitetura de rede neural que serviu como base para o modelo desenvolvido também é apresentada. Por fim, são apresentados o método para realçar impressões digitais latentes, juntamente com a arquitetura da RM-net e seu processo de treinamento.

4.1 Bases de dados

Nesta seção, são descritas as bases de dados utilizadas na pesquisa. Inicialmente, são apresentadas as bases de dados IIIT-D *Multisensor Optical And Latent Fingerprint* (MOLF) e a IIIT-D *latent fingerprint database* que foram utilizadas como conjuntos de testes. Uma base de dados auxiliar, utilizada para fins de aumento de dados, a NIST SD300 também é apresentada. Em seguida, é abordado o processo de criação de uma base de dados sintética que foi utilizada para o treinamento da RM-net desenvolvida nesta pesquisa.

Existem outras bases de dados de impressões digitais latentes reconhecidas, mas algumas delas não foram abordadas neste estudo. Especificamente, a NIST SD27 foi retirada pelos mantenedores para aprimoramento da documentação e não está mais disponível até o presente momento desta pesquisa. Além disso, a base MSP não está disponível publicamente (Chugh et al., 2018) e não foram obtidas informações para possibilitar requisitar o acesso à base WVU⁶. A base de dados NIST SD302 (N2N)⁷ será objeto de investigação em trabalhos futuros.

4.1.1 *Multisensor Optical And Latent Fingerprint database*

A IIIT-D MOLF (Sankaran et al., 2015b) é uma base de dados de impressões digitais desenvolvida e disponibilizada pelo *Indraprastha Institute of Information Technology, Delhi*⁸. Essa base é composta de 19.200 impressões digitais coletadas de 100 indivíduos usando cinco formas de captura diferentes: (i) sensor Lumidigm Venus IP65 Shell, (ii) sensor Secugen Hamster-IV, (iii) sensor CrossMatch L-Scan Patrol, (iv) impressões digitais latentes e (v) impressões digitais latentes simultâneas.

As capturas (i), (ii) e (iii) fizeram uso de três sensores ópticos. A resolução das imagens capturadas nos sensores Lumidigm, Secugen e CrossMatch é de 500 ppi. As

⁶<https://biic.wvu.edu/data-sets/3-d-face-databset>

⁷<https://www.nist.gov/itl/iad/image-group/nist-special-database-302>

⁸<https://www.iiitd.ac.in/>

dimensões das imagens capturadas por estes três sensores são 352×544 , 258×336 e 1600×1500 pixels, respectivamente. Para cada um dos 100 indivíduos, cada uma das 10 impressões digitais foi capturada em duas sessões e, em cada sessão, duas instâncias independentes são capturadas. Para cada sensor existem 4000 imagens, formando os conjuntos DB1, DB2 e DB3_A, com 1000 classes de impressão digital. Exemplos dessas impressões digitais são mostrados na Figura 29.

As impressões digitais latentes simultâneas, que formam o subconjunto de dados DB5, foram capturadas usando um processo químico de pó preto e digitalizadas por meio de câmera fotográfica. Os detalhes dessa configuração podem ser vistos na pesquisa de Sankaran et al. (2015b). O subconjunto DB4 foi criado a partir do recorte das imagens do DB5. Amostras da captura das impressões digitais latentes simultâneas são ilustradas na Figura 30.

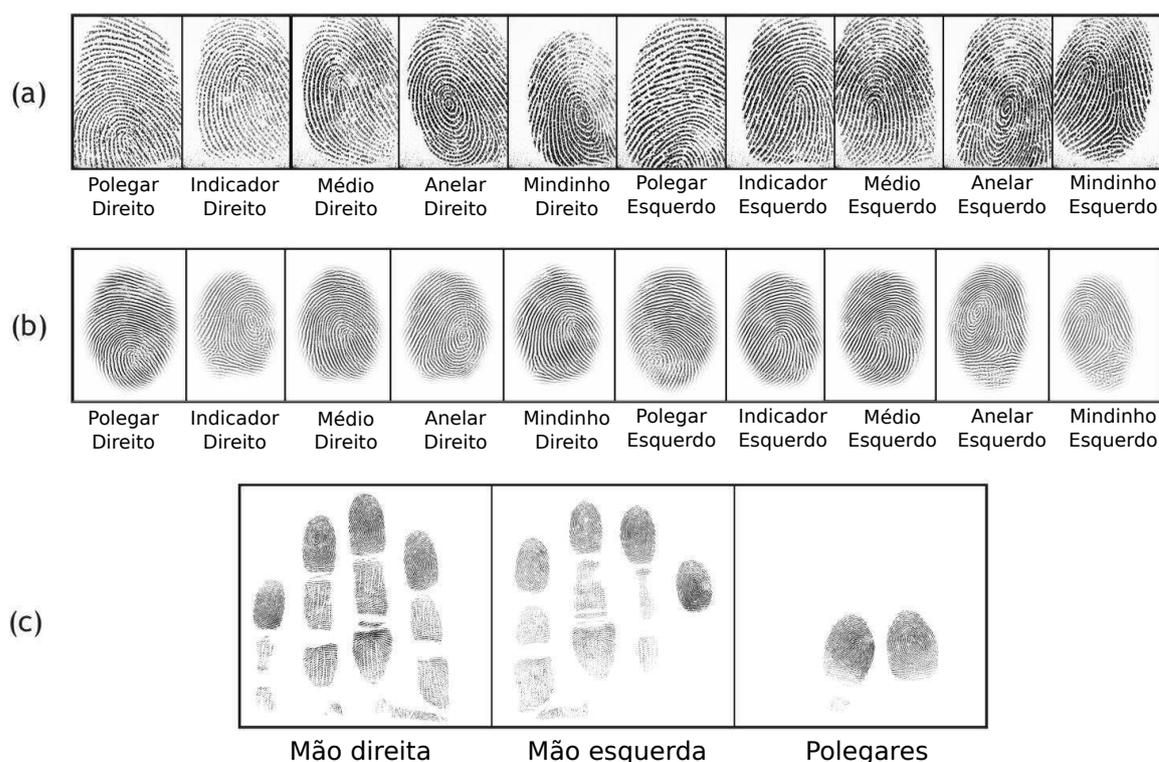


Figura 29: Amostras de impressões digitais capturadas por sensor existentes no banco de dados MOLF: (a), (b) e (c) são impressões digitais de 500ppi capturadas pelos sensores Secugen, Lumidigm, e CrossMatch L-Scan Patrol, respectivamente. Fonte: adaptado de Sankaran et al. (2015b).

O DB5 consiste de um total de 1.600 impressões digitais latentes simultâneas. Essas impressões digitais latentes foram recortadas manualmente para obtenção das impressões digitais latentes individuais que formam o DB4. Isso resultou em um total de 4.400 impressões digitais latentes de 100 indivíduos com 1000 classes. Assim, o DB4 contém duas ocorrências de impressão latente de cada polegar e dedo mínimo, quatro ocorrências de dedo anular, seis ocorrências de dedo médio e oito ocorrências de dedo indicador.

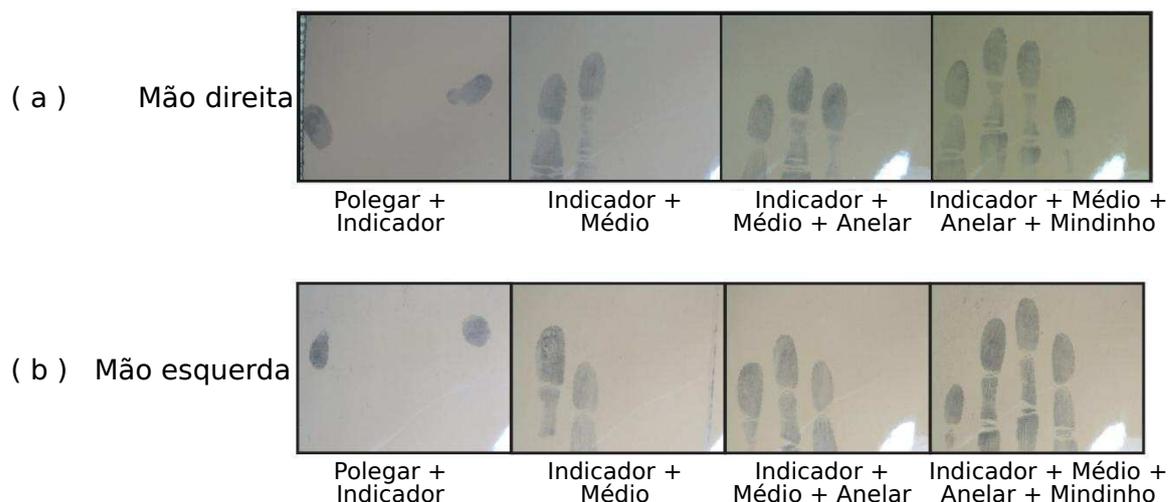


Figura 30: Amostras de impressões digitais latentes simultâneas da base de dados MOLF: (a) e (b) são conjuntos de impressões digitais latentes simultâneas da mão direita e mão esquerda de um indivíduo, respectivamente. Fonte: adaptado de Sankaran et al. (2015b).

No Quadro 4 é exibido um sumário da base de dados MOLF, mostrando os subconjuntos, os tipos de impressões digitais, a quantidade de imagens, a resolução das imagens e o protocolo de captura adotado para cada um deles.

4.1.2 IIIT-D *latent fingerprint database*

A base de dados IIIT-D *latent fingerprint database* (Sankaran et al., 2011) consiste em 1046 impressões digitais latentes, abrangendo todos os dez dedos de quinze indivíduos distintos. Essa base de dados inclui múltiplas instâncias para cada impressão digital, permitindo comparações entre as impressões latentes. As impressões digitais que compõem a base foram coletadas de dois fundos diferentes, cartão e azulejo, com o objetivo de simular diferentes ambientes de coleta. A forma de coleta das amostras, com diferentes níveis de umidade das cristas da pele dos dedos, resultou em múltiplas impressões de qualidade e informações variáveis. As imagens foram capturadas com sensores de impressão digital de 500ppi e 1000ppi, sendo que as primeiras foram obtidas por meio do dispositivo Crossmatch L1 scan e as últimas pelo dispositivo SecuGen Hamster IV. A Figura 31 exibe amostras de imagens que compõem a base de dados IIIT-D *latent fingerprint database*.

4.1.3 NIST *Special database 300*

A base de dados SD300 (Fiumara et al., 2018) resulta de um trabalho do NIST em conjunto com o FBI em que foi realizada a digitalização de 888 cartões de prisão com impressões digitais capturadas utilizando tinta, que estavam em várias condições

Quadro 4: Diferentes subconjuntos da base de dados MOLF, juntamente com o tipo de impressão digital, protocolo de captura e suas propriedades

Subconjunto	Tipo	Qtd. de imagens	Resolução das imagens	Protocolo de captura
DB1	Multiespectral, capturados via sensor Lumidigm Venus IP65 Shell	4000	352 × 544	100 indivíduos × 10 dedos × 2 sessões × 2 instâncias
DB2	Capturados via sensor Secugen Hamster-IV	4000	258 × 336	100 indivíduos × 10 dedos × 2 sessões × 2 instâncias
DB3	Capturados via sensor CrossMatch L-Scan Patrol	1200	1600 × 1500	100 indivíduos × 3 capturas (mão esquerda, mão direita e polegares) × 2 sessões × 2 instâncias
DB3_A	Impressões recortadas do DB3	4000	variado	100 indivíduos × 10 dedos × 2 sessões × 2 instâncias
DB4	Latentes, recortadas do DB5	4400	variado	100 indivíduos × 2 mãos × 2 sessões × 11 instâncias
DB5	Latentes simultâneos	1600	1924 × 1232	100 indivíduos × 2 mãos × 2 sessões × 4 instâncias

físicas, desde intactas até muito danificadas e desbotadas, coletadas durante o exercício das funções dos profissionais da lei. A base de dados NIST *Special Database 300* contém imagens de 10 impressões digitais roladas, coletas de impressões planas dos quatro dedos (indicador, médio, anelar e mindinho) de cada mão, duas impressões planas referentes aos polegares e as impressões recortadas individualmente das imagens planas. Além disso, a base de dados inclui as coordenadas usadas para gerar o recorte das impressões digitais.

Os cartões foram digitalizados utilizando os escâneres Epson Perfection 4990 ou Epson V700 e as capturas se deram em três resoluções diferentes: 500, 1.000 e 2.000 *pixels* por polegada (PPI). As três resoluções foram digitalizadas em escala de cinza a uma profundidade de 8 *bits* por *pixel*. A Figura 32 mostra algumas imagens pertencentes à base de dados NIST SD300.

Nesta pesquisa, a base de dados NIST SD300 foi utilizada com o propósito de aumento de dados. Algumas imagens dessa base foram selecionadas para compor o conjunto



Figura 31: Amostras de impressões digitais latentes existentes na base de dados IIIT-D *latent fingerprint database*. Fonte: De autoria própria.



Figura 32: Amostras de impressões digitais existentes na base de dados NIST SD300. Fonte: De autoria própria.

de dados de treinamento da rede neural desenvolvida.

4.1.4 Base de dados sintética *Ad hoc*

Um dos desafios para treinar métodos baseados em deep learning (DL) para realizar a tarefa de realce de imagens de impressões digitais latentes é a falta de bases de dados publicamente disponíveis que contenham as impressões digitais latentes e suas imagens de referência correspondentes. Normalmente, esses modelos demandam uma grande quantidade de dados. Por conta disso, foi necessário gerar impressões digitais latentes sintéticas que simulam características de ruído semelhantes às observadas em impressões digitais latentes reais.

Para criar o banco de dados de treinamento utilizado nesta pesquisa, foi utilizada a versão de demonstração do programa *Synthetic Fingerprint Generator (SFinGe)*. SFinGe é um software que implementa um método para a geração de imagens sintéticas de impressões digitais. Com ele, é possível gerar impressões digitais, com seus respectivos *ground truths*, configurando várias opções que simulam diferentes características existentes em

uma impressão digital real (Maltoni et al., 2009b). Um exemplo de par, impressão digital e *ground truth*, gerado pelo SFinGe, é mostrado na Figura 33.



Figura 33: Amostra de impressão digital e seu respectivo *ground truth*, da esquerda para direita, gerados pelo programa SFinGe. Fonte: De autoria própria.

A versão do SFinGe utilizada possui algumas limitações. Por isso, foi necessário programar uma macro para automatizar o avanço das telas, a exploração das diferentes opções disponíveis para variar os tipos de impressões digitais criadas e a captura das impressões digitais com seus *ground truths*. A Figura 34 mostra uma visão geral resumida das etapas de criação dos pares impressão digital e *ground truth*. Nela, cada passo refere-se a uma tela do SFinGe. As etapas destacadas em verde são as que envolvem a captura das telas. Todo o passo a passo e os detalhes do procedimento de captura das imagens, exemplificado com imagens das telas do programa, são explicados no Apêndice A.

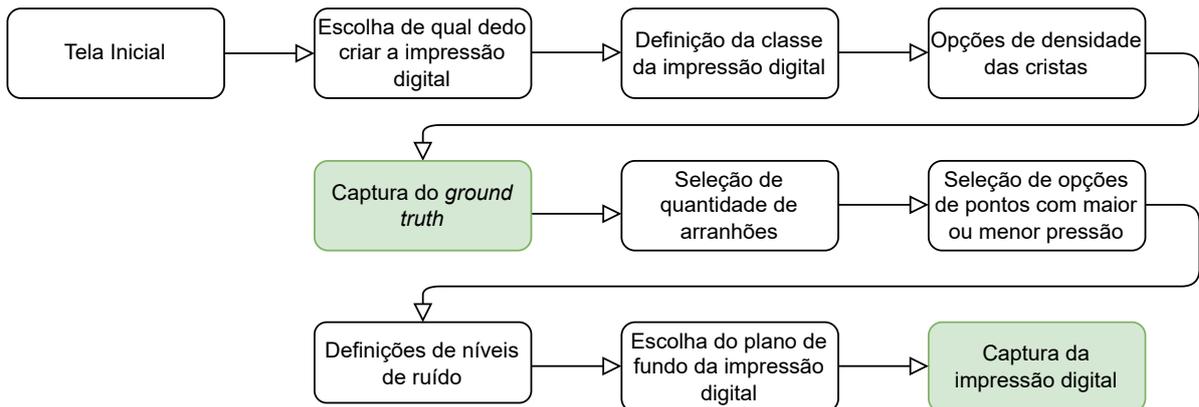


Figura 34: Etapas do processo para gerar o par impressão digital e *ground truth*. Fonte: De autoria própria.

4.1.4.1 Ruídos e deteriorações adicionados

Conforme explicado no Capítulo 1, impressões digitais latentes costumam ser borradas, apresentam ruídos variados e, algumas vezes, elementos sobrepostos como textos,

linhas e até mesmo outras impressões digitais. Para criar o banco de dados de treinamento utilizado nesta pesquisa, foram adicionados ruídos e simulações de deteriorações às imagens do banco de dados sintético gerado. Para aplicação dos diferentes tipos de ruídos e deteriorações foram utilizadas as bibliotecas de manipulação de imagens PIL⁹ e OpenCV¹⁰. Todo o processo é descrito a seguir.

- ***Gaussian blur***: é comum que impressões digitais latentes apresentem padrões de manchas e borrados não uniformes. No intuito de simular essa característica, 5 variações de *Gaussian blur* foram aplicadas. As duas primeiras variações compreendem a imagem da impressão digital por completo, utilizando os valores do parâmetro *radius* = 1 e *radius* = 2. Este parâmetro define o desvio padrão do *kernel* gaussiano.

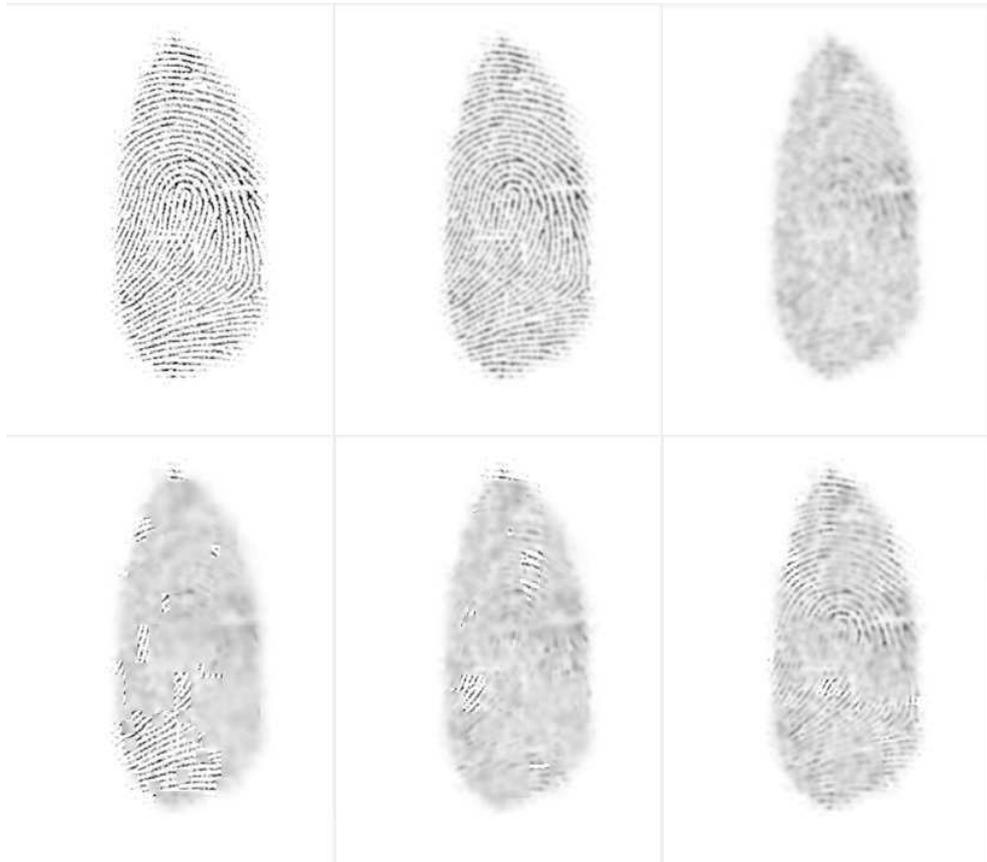


Figura 35: Amostra das diferentes aplicações de *gaussian blur* nas imagens do conjunto de treinamento. A primeira imagem refere-se a impressão digital sintética original gerada pelo SFInGe. As demais imagens são a impressão digital original após a aplicação dos efeitos de borrado. Fonte: De autoria própria.

Para as outras três variações, a estratégia adotada consiste em selecionar pequenas partes da imagem, com tamanho 10×10 ou 40×40 , permitindo sobreposição, e

⁹<https://pillow.readthedocs.io/en/stable/>

¹⁰<https://opencv.org/>

borrá-las. Os valores do parâmetro *radius* aplicados nesse caso foram de 1, 1,5, e 2,0. Na Figura 35 são exemplificados os borrados aplicados nas imagens da base de dados de treinamento.

- ***Gaussian noise***: outra característica encontrada nas imagens de impressão latente é um ruído resultante do processo de coleta por meio da utilização de um pó químico. Com o objetivo de simular essa característica, *Gaussian noise*, com parâmetros $\mu = 0$ e $\sigma^2 = 0,01$, foi adicionado às imagens, de forma global, conforme pode ser visto no exemplo da Figura 36.



Figura 36: Amostra da aplicação de *gaussian noise* nas imagens do conjunto de treinamento. A primeira imagem refere-se a impressão digital sintética original gerada pelo SFinGe. A segunda imagem é a impressão digital original após a aplicação do *gaussian noise*. Fonte: De autoria própria.

- ***Salt and Pepper***: outro tipo de ruído adicionado às imagens foi o *Salt and Pepper*, usando $x = 30\%$ dos pixels da imagem substituídos por ruído, variando-se, desta forma, os tipos de ruídos presentes nos dados de treinamento da rede, com foco em obter uma maior generalização. Amostras das imagens resultantes após a adição do ruído tipo *Salt and Pepper* são mostradas na Figura 37.
- **Contraste**: objetivando simular diferentes situações de luminância que podem ocorrer em relação aos ambientes em que as impressões digitais latentes são encontradas, foi aplicada uma redução aleatória de contraste, variando entre 10% e 60%. Um exemplo dessa redução é exibido na Figura 38.
- **Brilho**: com o mesmo objetivo da aplicação de redução de contraste nas imagens, uma redução aleatória de brilho também foi aplicada, também variando entre 10% e 60%. Essa redução de brilho é exemplificada na Figura 39;
- **Deteriorações combinadas**: outros dois tipos de deteriorações estão presentes no banco de dados de treinamento sintético criado. Essas deteriorações são resultantes

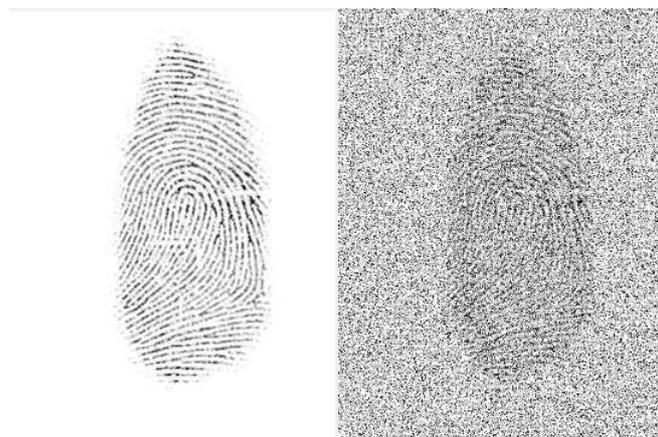


Figura 37: Amostra da aplicação de ruído *Salt and Pepper* nas imagens do conjunto de treinamento. A primeira imagem refere-se a impressão digital sintética original gerada pelo SFinGe. A segunda imagem é a impressão digital original após a aplicação do ruído *Salt and Pepper*. Fonte: De autoria própria.



Figura 38: Amostra da aplicação de redução de contraste nas imagens do conjunto de treinamento. A primeira imagem refere-se a impressão digital sintética original gerada pelo SFinGe. A segunda imagem é a impressão digital original após a redução do contraste. Fonte: De autoria própria.

da combinação da aplicação do *Gaussian blur*, com o parâmetro $radius = 1$, com os ruídos *Gaussian noise* e *Salt and Pepper*. O resultado dessa combinação de ruídos e borrados pode ser visto na Figura 40.

As mesmos ruídos e deteriorações aplicados às imagens geradas pelo SFinGe também foram aplicadas em imagens de impressões digitais reais da base de dados NIST SD300. Dessa base foram selecionadas as imagens de melhor qualidade, após uma análise considerando a pontuação dada pelo NFIQ. Dessas imagens foram extraídos as saídas esperadas utilizando o *NIST Biometric Image Software* (NBIS). Amostras de imagens da base de dados NIST SD300, após a adição de ruídos, são exibidas na Figura 41.

Atualmente, a base de dados de treinamento não contempla elementos sobrepostos, como textos e linhas, e nem diferentes planos de fundo que simulem diferentes superfícies

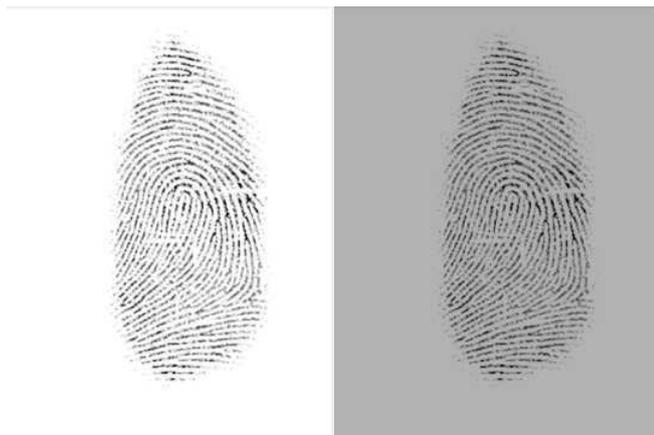


Figura 39: Amostra da aplicação de redução de brilho nas imagens do conjunto de treinamento. A primeira imagem refere-se a impressão digital sintética original gerada pelo SFinGe. A segunda imagem é a impressão digital original após a redução do brilho. Fonte: De autoria própria.

em que as impressões digitais latentes podem ser encontradas. No entanto, esse aprimoramento da base de dados é uma proposta para trabalhos futuros que darão continuidade à pesquisa.

O número total de imagens geradas e usadas para o treinamento da rede neural desenvolvida foi de 69.472, sendo 59.280 impressões digitais sintéticas geradas com base nas imagens SFinGe e 10.192 com base em imagens selecionadas do NIST SD300.

4.2 Arquitetura da rede neural

Esta subseção, detalha a arquitetura da rede neural desenvolvida nesta pesquisa. A M-net, que serviu de base para a RM-net, é apresentada inicialmente. Em seguida, são relatados os aprimoramentos da M-net que resultaram na RM-net.

4.2.1 M-net

A M-net é uma arquitetura de rede neural convolucional seguindo a estrutura codificador-decodificador, ou arquitetura *hourglass*, proposta por Mehta e Sivaswamy (2017), que foi originalmente aplicada na tarefa de segmentação de estruturas profundas do cérebro humano em imagens de ressonância magnética. Inspirada na U-net (Ronneberger et al., 2015; Çiçek et al., 2016), a M-net diferencia-se pela introdução de dois novos caminhos laterais de saltos ou atalhos de conexões (*skip connections*) no codificador e no decodificador, sendo esses novos caminhos chamados de perna esquerda (*left leg*) e perna direita (*right leg*), respectivamente. A arquitetura da M-net é ilustrada na Figura 42.

Apesar ter sido originalmente aplicada a um problema de classificação, a M-net recentemente tem sido aplicada com sucesso a problemas de regressão, como tarefas de



Figura 40: Amostra da combinação de ruídos e borrado nas imagens do conjunto de treinamento. A primeira imagem refere-se a impressão digital sintética original gerada pelo SFinGe. A segunda imagem é a impressão digital original após a combinação de *gaussian noise* e *gaussian blur* e a terceira imagem é após a combinação de ruído *Salt and Pepper* e *gaussian blur*. Fonte: De autoria própria.

aprimoramento (Fan et al., 2022a), e de redução de ruído de imagens (Adiga e Sivaswamy, 2019), (Liu et al., 2020) e (Fan et al., 2022b).

4.2.2 RM-net

O modelo RM-net desenvolvido nesta pesquisa é baseado em autcodificadores convolucionais, mais especificamente, é uma adaptação da arquitetura da M-net descrita na subseção anterior. A arquitetura geral da rede neural pode ser vista na Figura 43. Os agrupamentos de blocos principais consistem em dois blocos repetidos de uma camada convolucional (Conv), uma camada de *Batch Normalisation* (BN) e a função de ativação ReLU. As dimensões dos *kernels* para os quatro grupos de blocos do codificador foram ajustados para os valores 9×9 , 7×7 , 5×5 e 3×3 . Entre cada agrupamento de blocos uma camada *dropout*, com probabilidade 0,2, é incluída e a saída desses agrupamentos é concatenada e reduzida com uma operação de *pooling* máximo.

Blocos *resnet* foram adicionados à rede para aprimorar a reconstrução de estruturas locais das impressões digitais e, conseqüentemente, melhorar a qualidade das imagens realçadas. Cada bloco *resnet* consiste em um agrupamento de dois blocos individuais que contém uma camada convolucional com tamanho do *kernel* 3×3 , uma camada de BN e função de ativação ReLU, cada. A arquitetura dos blocos *resnet* pode ser visualizada na Figura 44. Experimentos preliminares, que auxiliaram na definição da quantidade de blocos residuais adicionados são descritos no Apêndice C.



Figura 41: Amostras de impressões digitais latentes sintéticas geradas a partir de imagens da base de dados NIST SD300. Fonte: De autoria própria.

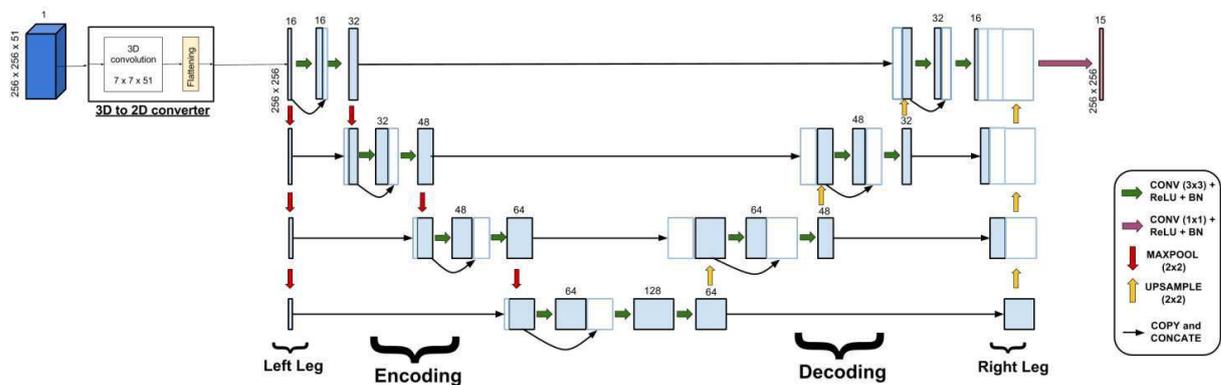


Figura 42: Arquitetura da M-net. Fonte: Mehta e Sivaswamy (2017).

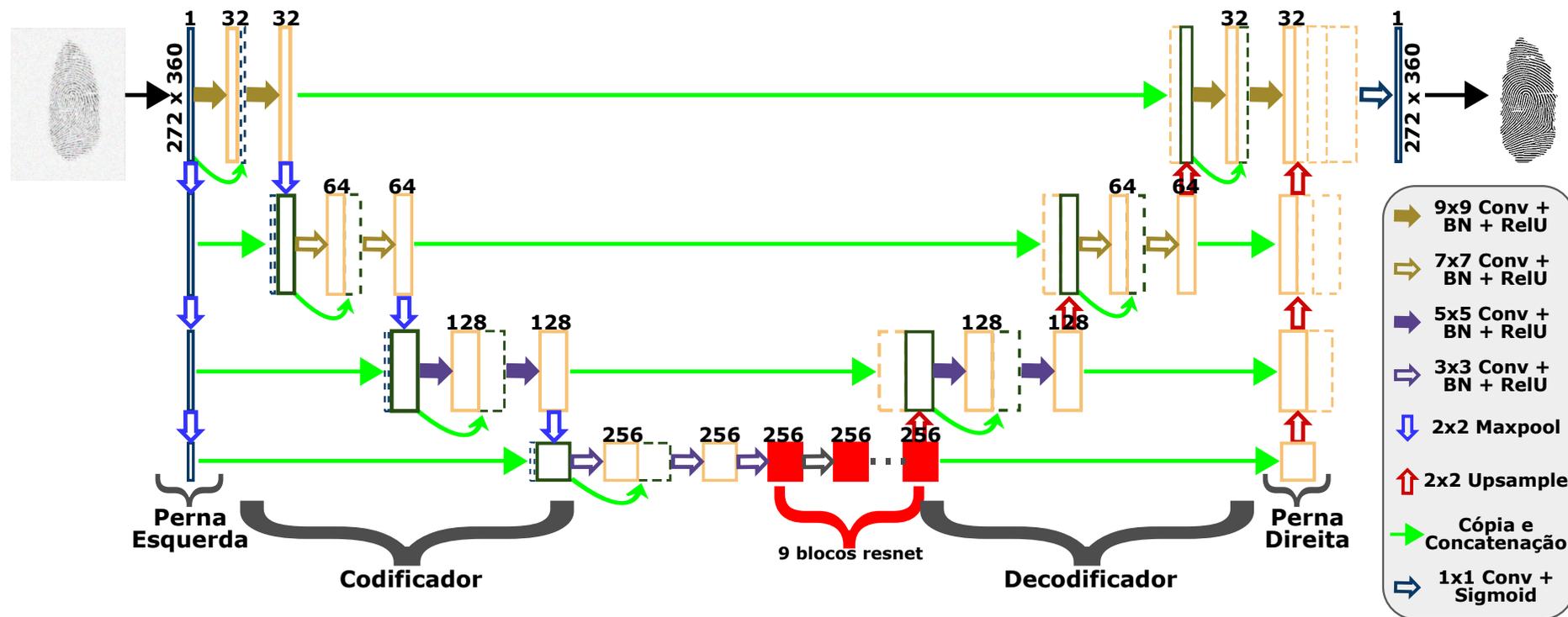


Figura 43: Arquitetura da rede neural RM-Net proposta, destacando em vermelho os blocos resnet adicionados como principal alteração na arquitetura da M-net original. Fonte: De autoria própria.

Os agrupamentos de blocos do decodificador são semelhantes aos do codificador, sendo as camadas de *pooling* máximo substituídas por camadas de *upsampling* para aumentar a resolução do mapa de características de modo que a imagem de saída do decodificador possua a mesma resolução da imagem de entrada da rede. Por fim, uma camada convolucional com *kernel* de tamanho 1×1 e com função de ativação sigmóide, fornece resultado de saída da rede.

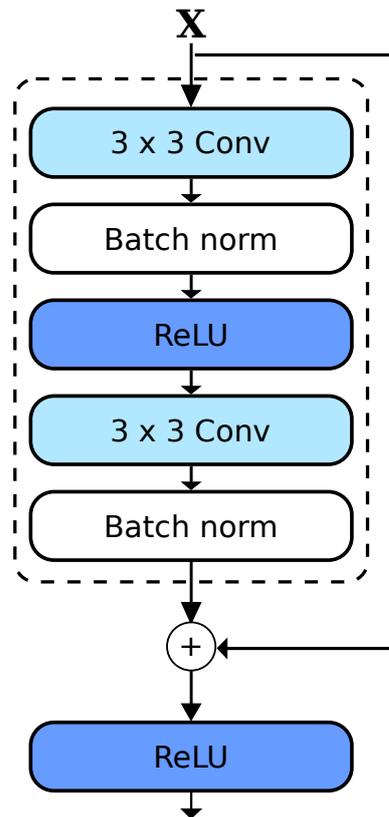


Figura 44: Configuração dos blocos resnet adicionados à arquitetura da M-net. Fonte: De autoria própria.

4.3 Função de custo

A função de custo adotada para o treinamento da RM-net é uma adaptação das funções de custo baseadas no domínio da frequência apresentadas por Yadav et al. (2021), combinadas com a MAE (Qi et al., 2020). Essas funções são construídas a partir da utilização da DCT e da DFT. Experimentos preliminares, que auxiliaram na definição da função de custo da RM-net são descritos no Apêndice C. Especificamente, a função de custo MAE é apresentada na Equação (19).

$$MAE = \frac{1}{n} \sum_{i=1}^n \|y_i - \hat{y}_i\| \quad (19)$$

em que n representa o número de pixels da imagem e y_i e \hat{y}_i representam o *ground-truth* e a imagem realçada reconstruída, respectivamente.

A função de custo baseada na DCT $L_{DCT}^{\frac{M}{K} \times \frac{N}{K}}$ entre as imagens I_1 e I_2 de dimensões $M \times N$ é definida na Equação (20).

$$L_{DCT}^{\frac{M}{K} \times \frac{N}{K}} = \frac{K^2}{M \times N} \|DCT(I_1) - DCT(I_2)\|_{\frac{M}{K} \times \frac{N}{K}} \quad (20)$$

em que K é um fator de escalonamento e $DCT(I)$ refere-se à transformada discreta de cosseno da imagem I . Como proposto em Yadav et al. (2021), calcula-se esse custo em três escalas diferentes das imagens e obtém-se (21):

$$DCT_{Final}(I_1, I_2) = L_{DCT}^{\frac{M}{1} \times \frac{N}{1}} + L_{DCT}^{\frac{M}{2} \times \frac{N}{2}} + L_{DCT}^{\frac{M}{4} \times \frac{N}{4}} \quad (21)$$

De forma similar, também é definida $DFT_{Final}(I_1, I_2)$, como mostrado em (22):

$$DFT_{Final}(I_1, I_2) = L_{DFT}^{\frac{M}{1} \times \frac{N}{1}} + L_{DFT}^{\frac{M}{2} \times \frac{N}{2}} + L_{DFT}^{\frac{M}{4} \times \frac{N}{4}} \quad (22)$$

Durante o processo de treinamento da rede, são calculados DCT e DFT dos *ground truth* e das imagens preditas e é calculada a diferença média absoluta entre eles. Esse processo é repetido em duas resoluções menores, metade e um quarto do tamanho.

Foram realizados experimentos iniciais combinando a MAE com a DFT e a MAE com a DCT, respectivamente. A combinação MAE-DFT obteve melhores resultados, considerando os testes de casamento, do que a combinação MAE-DCT. Em um experimento final, as três funções de custo foram combinadas, resultando nos melhores resultados dentre todos os experimentos conduzidos. A função de custo final utilizada na rede neural é apresentada na Equação (23).

$$\begin{aligned} L_{Final}(I_1, I_2) &= \lambda_{MAE} \cdot MAE \\ &+ \lambda_{DCT} \cdot DCT_{Final}(I_1, I_2) \\ &+ \lambda_{DFT} \cdot DFT_{Final}(I_1, I_2) \end{aligned} \quad (23)$$

em que λ é um parâmetro de ponderação. Após testes exploratórios experimentais, seus valores foram ajustados para $\lambda_{MAE} = 0,9$, $\lambda_{DCT} = 0,05$ e $\lambda_{DFT} = 0,05$. Os testes para encontrar os valores de λ se deram da seguinte forma. Inicialmente, λ_{MAE} foi definida como 0,1 e as outras duas funções de custo foram ponderadas igualmente, com $\lambda_{DCT} = \lambda_{DFT} = 1 - \frac{\lambda}{2}$. Após ciclo de treinamento, testes de casamento e análise dos resultados, λ_{MAE} era acrescida em 0,1. Esse processo foi repetido até λ_{MAE} atingir o valor de 0,9, o

qual forneceu os melhores resultados.

4.3.1 Metodologia de treinamento

Para o treinamento da RM-net, a base de dados sintética foi dividida em dois conjuntos, treinamento e validação. As bases de dados MOLF DB4 e IIIT-D *Latent fingerprint database* são os conjuntos de dados utilizados para testes. Os resultados são reportados em termos de acurácia *rank-n*. Do total de imagens da base de dados sintética criada, 20% foi aplicado como conjunto de validação e o restante como conjunto de treinamento da rede neural.

Com o intuito de evitar um superajuste dos dados e também manter uma melhor generalização, as técnicas de regularização aplicadas durante o treinamento foram a utilização de BN e camadas de *dropouts*, conforme explicado na Subseção 4.2. Detalhes de implementação da RM-net são fornecidos no Apêndice B.

4.3.2 Parâmetros e hiperparâmetros

A arquitetura da RM-net é composta por um total de 18.596.897 parâmetros, dos quais 18.584.353 são parâmetros treináveis, e os 12.544 restantes são parâmetros não treináveis relacionados à média e variância calculadas por camadas de *Batch Normalization*. Os melhores valores de acurácia dos testes de casamento foram obtidos pelo modelo treinado até a época 29. O tamanho do *batch* utilizado foi 8, e o otimizador empregado foi o *Adaptive Momentum Estimation* (Adam), com os seguintes valores padrão de taxa de aprendizagem: $\alpha = 0,0002$, $\beta_1 = 0,5$, $\beta_2 = 0,999$, e $\epsilon = 10^{-7}$.

4.4 Considerações finais sobre o capítulo

Neste capítulo foram apresentados detalhes dos materiais utilizados nesta pesquisa, incluindo a descrição do processo de criação do conjunto sintético de testes e informações sobre as bases de dados utilizadas para os testes finais de realce e casamento das impressões digitais latentes realçadas pela RM-net. Além disso, foram descritas a arquitetura e a metodologia de treinamento da rede neural, bem como as linguagens de programação, *frameworks*, ambiente de desenvolvimento, programas e demais ferramentas utilizadas neste estudo.

O próximo capítulo abordará uma análise dos resultados obtidos e também os critérios de avaliação baseados nas métricas adotadas, seguidos de uma comparação com outros métodos presentes na literatura.

5 Experimentos e Resultados

Neste capítulo, são apresentados os resultados alcançados pelo método desenvolvido. Para avaliação do desempenho, foram realizados dois experimentos distintos. O primeiro, latente-para-sensor, consistiu em comparar as imagens de impressões digitais latentes realçadas pelo método com as imagens capturadas por escâner, a fim de avaliar a qualidade do realce e a capacidade de preservação de características únicas da impressão digital. O segundo experimento, denominado latente-para-latente, consistiu em um teste de identificação entre impressões digitais latentes realçadas. Além disso, os resultados obtidos foram comparados com os métodos de realce de impressões digitais latentes encontrados na literatura.

5.1 Treinamento da rede neural

Na Figura 45 é apresentada a curva de aprendizado da RM-net durante o processo de treinamento da rede neural. Observa-se que a função de custo apresentou uma diminuição consistente tanto para o conjunto de treinamento quanto para o conjunto de validação, ao longo do aumento das épocas. Nota-se também que, em geral, o valor do

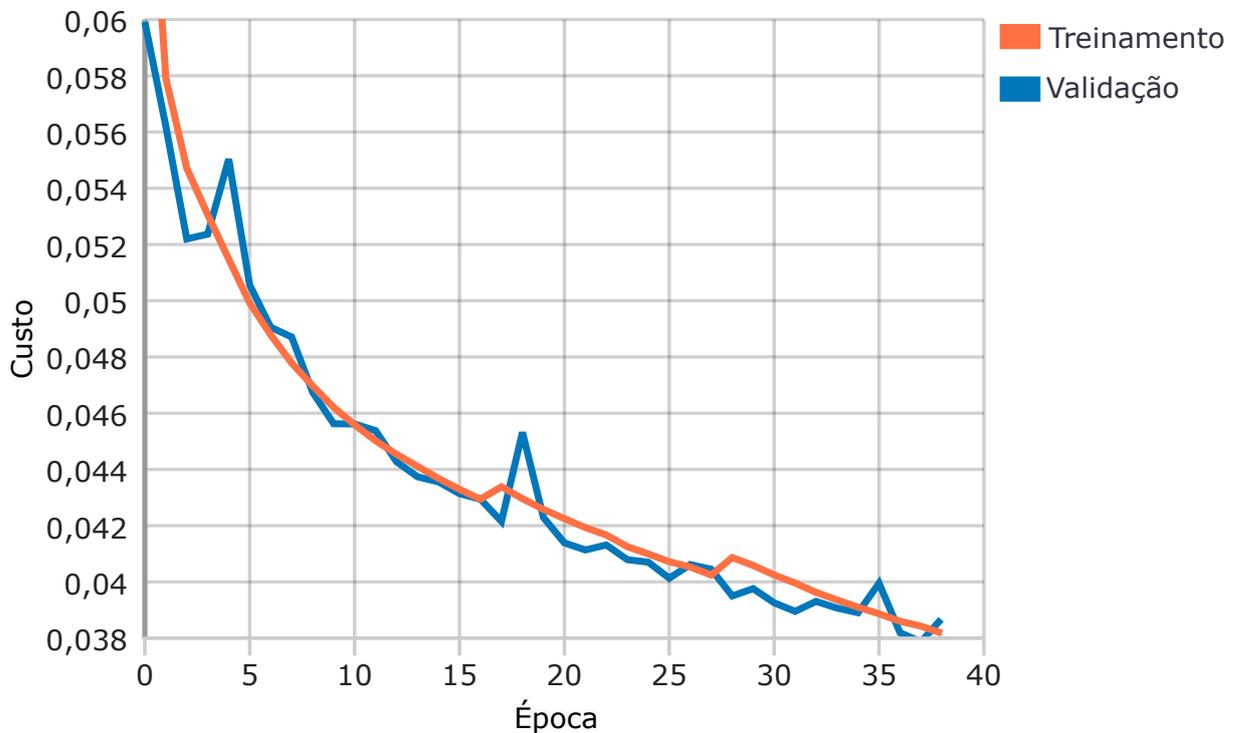


Figura 45: Curvas da evolução da função de custo durante o treinamento da RM-net. Fonte: De autoria própria.

custo no conjunto de validação foi menor que no conjunto de treinamento. Esse fenômeno é atribuído à técnica de regularização aplicada na rede por meio do *dropout*. Métodos

de regularização geralmente reduzem um pouco a precisão do treinamento para melhorar a precisão da validação e teste. Em alguns casos, isso pode fazer com que o custo no conjunto de validação seja menor que no conjunto de treinamento. Vale ressaltar que a regularização, incluindo o *dropout*, não é aplicada durante as etapas de validação e teste.

A preservação da estrutura de cristas é um aspecto fundamental em algoritmos de realce de impressões digitais, garantindo a fidelidade das características presentes na impressão digital original na versão realçada. Nesse contexto, a métrica de similaridade SSIM é uma ferramenta auxiliar utilizada para avaliar essa preservação. Quanto maior o valor do SSIM, maior a indicação de que a semelhança estrutural entre a saída esperada (*ground truth*) e a reconstrução é mantida, o que implica a preservação da estrutura de cristas. Na Figura 46 é mostrada a evolução do valor do SSIM em relação às épocas durante o treinamento da RM-net.

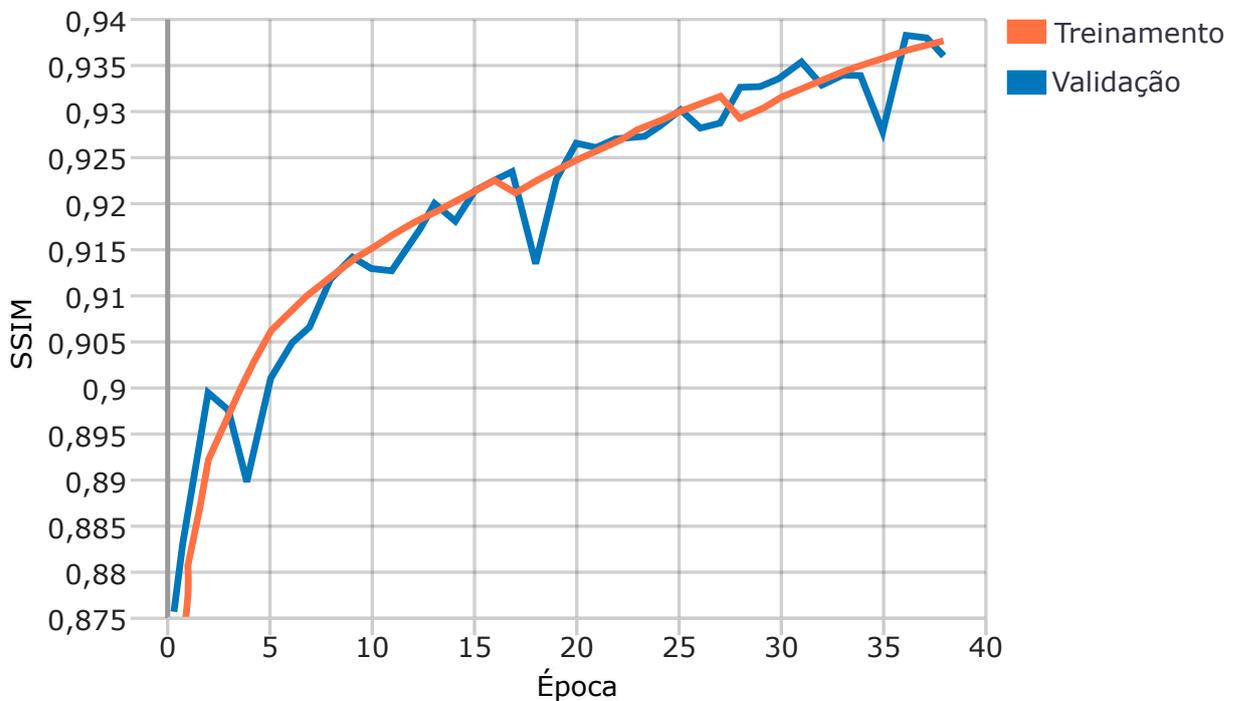


Figura 46: Curvas da evolução do SSIM durante o treinamento da RM-net. Fonte: De autoria própria.

5.2 Avaliação em um conjunto sintético de teste

Uma avaliação inicial do desempenho do método desenvolvido para o realce de impressões digitais foi conduzida em um banco de dados sintético de testes. Esse conjunto de dados foi criado seguindo os mesmos procedimentos descritos na Seção 4.1.4. O conjunto sintético de testes consiste em um total de 96 imagens, cada uma acompanhada de seu respectivo *ground truth*. Essa configuração permite a comparação entre a imagem realçada pela rede desenvolvida e o *ground truth*, utilizando a métrica de similaridade

estrutural SSIM, explicada na Subseção 2.4.2, que quantifica a semelhança entre as duas imagens.

Na Figura 47 é visto uma comparação entre as impressões digitais sintéticas do conjunto de testes e suas respectivas versões realçadas pela RM-net. Na primeira coluna da

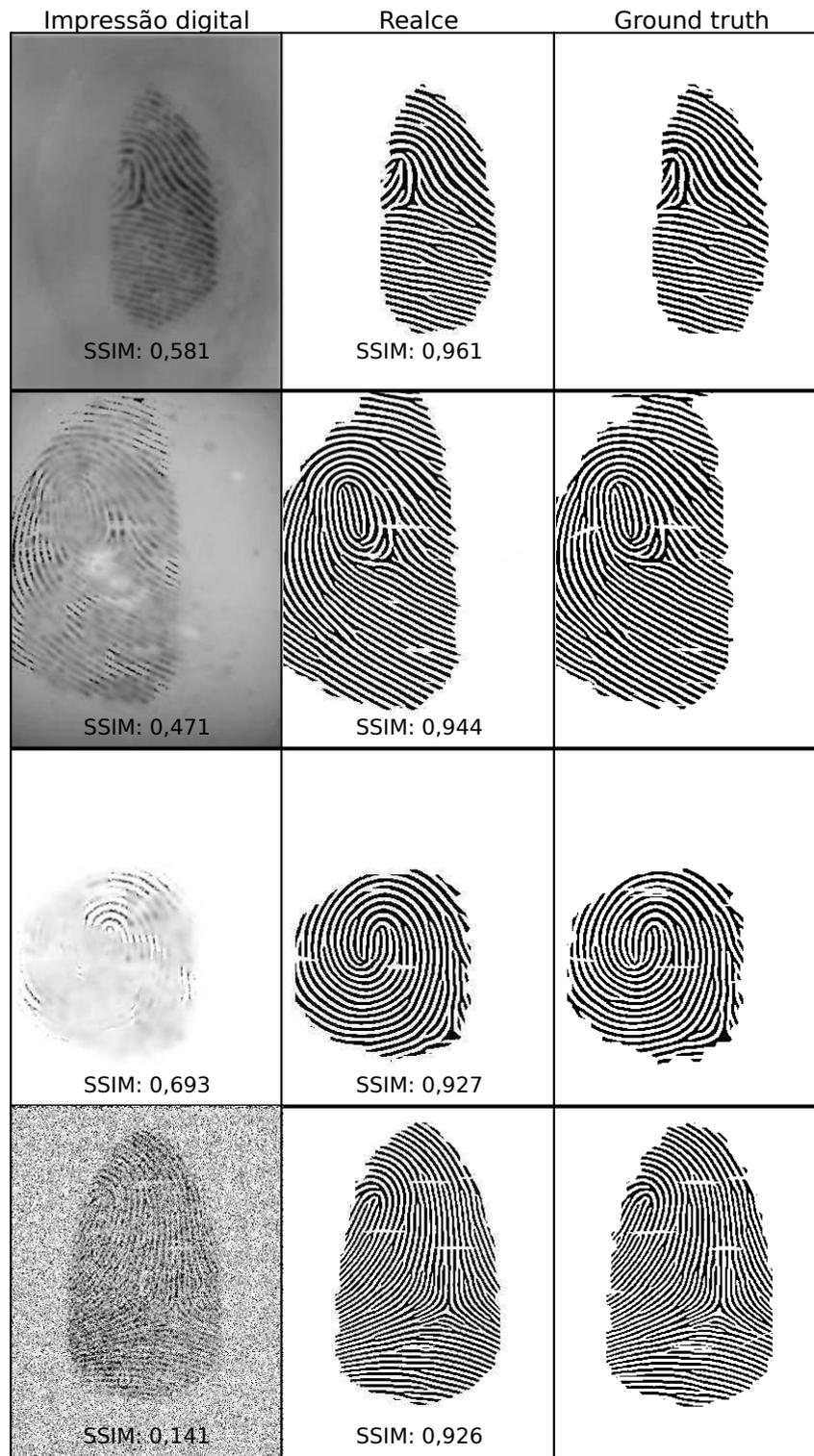


Figura 47: Amostras de reconstruções do conjunto sintético de testes e valores do SSIM obtidos em comparação aos *ground truths*. Fonte: De autoria própria.

imagem estão amostras das impressões digitais sintéticas originais com os seus respectivos valores de SSIM em relação aos seus *ground truths*. A segunda coluna exhibe o resultado do processo de realce das imagens correspondentes da primeira coluna, também com os seus respectivos valores obtidos da métrica SSIM em relação aos seus *ground truths*. A média do SSIMs do conjunto sintético de testes foi de 0,909. É importante ressaltar que o SSIM atribui um valor entre 0 e 1 ao comparar duas imagens em que quanto mais próximo de 1, mais similares são as imagens.

5.3 Experimento latente-para-sensor

Neste experimento, a base de dados MOLF é utilizada como conjunto de testes. O protocolo de teste adotado segue o estabelecido por Sankaran et al. (2015b), que consiste em utilizar as imagens de impressões digitais latentes realçadas do conjunto de dados DB4 como *probes* e compará-las com as imagens dos DB1, DB2 e DB3_A, utilizados como galerias.

Os resultados são apresentados em termos de acurácia, expressa em porcentagem, nos *ranks* 1, 25, 30 e 50. Além disso, os resultados do método desenvolvido são comparados com os de outros métodos estado-da-arte encontrados durante a pesquisa dos trabalhos relacionados. No entanto, o método proposto por Wong e Lai (2020), mencionado no Capítulo 3, não foi incluído na comparação devido à utilização de um protocolo de teste diferente para avaliar o desempenho na identificação das impressões digitais latentes realçadas.

Os resultados do experimento são resumidos na Tabela 1, que apresenta os valores de acurácia obtidos nos testes de correspondência. Esses resultados demonstram que a RM-net alcançou os melhores valores de acurácia para os *ranks* de 1 a 30 em todas as galerias analisadas.

Para avaliar o desempenho na identificação de impressões digitais latentes aprimoradas, foi gerada a Curva de Correspondência Cumulativa (CMC, do inglês *Cumulative Match Characteristic*), ilustrada na Figura 48. A CMC mostra a proporção de identidades corretamente correspondidas em função do *rank*, sendo que o *rank* 1 representa a maior precisão (ou seja, identificar corretamente o indivíduo na primeira tentativa). O método desenvolvido nesta pesquisa obteve taxas de acurácia mais elevadas em comparação com os outros métodos nos *ranks* 1 a 30, com uma diferença considerável observada no *rank* 1, em que a sua acurácia foi aproximadamente quatro vezes maior do que a do método PIDI (Dabouei et al., 2018), que demonstrou o melhor desempenho entre os métodos da literatura nos *ranks* mais altos.

Com o intuito de demonstrar a contribuição da função de custo no domínio da frequência no realce das imagens de impressões digitais latentes, foi conduzida uma análise

Tabela 1: Acurácias obtidas no experimento latente-para-sensor utilizando o VeriFinger como extrator de minúcias e algoritmo de casamento.

Galeria	Realce	Acurácia no <i>rank</i>			
		1	25	30	50
DB1	Raw	-	3,13	-	6,80
	Svoboda et al. (2017)	-	19,51	-	26,24
	Dabouei et al. (2018)	-	42,04	-	70,89
	RM-net	39,20	48,95	-	52,25
DB2	Raw	-	2,33	-	6,37
	Svoboda et al. (2017)	-	15,23	-	21,81
	Dabouei et al. (2018)	-	37,14	-	66,11
	RM-net	34,31	45,00	-	48,39
DB3	Raw	-	3,17	-	6,51
	Svoboda et al. (2017)	-	18,34	-	24,78
	Dabouei et al. (2018)	-	41,27	-	68,61
	Xu et al. (2020)	36,82	45,88	-	-
	Horapong et al. (2021)	35,48	-	43,68	-
	RM-net	39,77	49,25	49,89	52,43

comparativa do modelo RM-net treinado utilizando a função de custo no domínio da frequência (referenciado na tabela como *f. loss*) em comparação com duas outras versões: uma versão do modelo treinada apenas com a função de custo MAE, que é uma das funções mais utilizada em tarefas de reconstrução de imagens, e uma versão padrão da M-net, sem a inclusão de blocos residuais. A avaliação dos resultados foi realizada considerando os *ranks* 1, 25 e 50, resumidos na Tabela 2.

Tabela 2: Acurácias obtidas no experimento latente-para-sensor. Comparação dos resultados obtidos pelos modelos M-net padrão, RM-net treinada com a função de custo MAE e a RM-net treinada com a função de custo no domínio da frequência.

Galeria	Modelo	Acurácia		
		<i>Rank-1</i>	<i>Rank-25</i>	<i>Rank-50</i>
DB1	M-net	37,04	42,38	42,75
	RM-net (MAE)	35,70	45,45	48,52
	RM-net (<i>f. loss</i>)	39,20	48,95	52,25
DB2	M-net	31,34	36,48	36,70
	RM-net (MAE)	31,66	41,34	44,29
	RM-net (<i>f. loss</i>)	34,31	45,00	48,39
DB3	M-net	36,95	42,84	43,39
	RM-net (MAE)	36,16	44,95	47,93
	RM-net (<i>f. loss</i>)	39,77	49,25	52,43

Sobre os resultados exibidos na Tabela 2, a RM-net (MAE) obteve resultados ligeiramente melhores que a M-net, indicando que a adição de blocos residuais trouxe

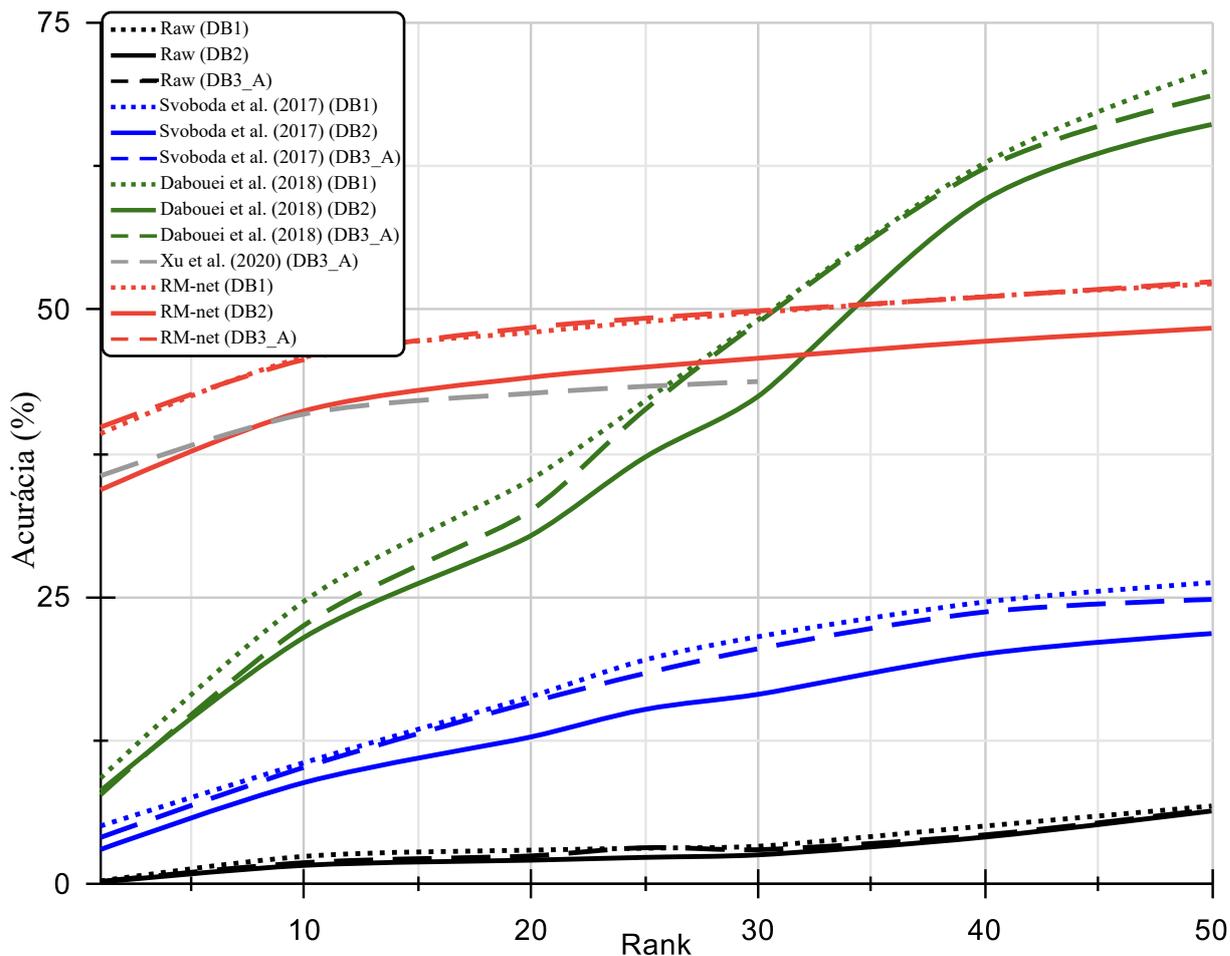


Figura 48: Curvas CMC comparando-se os métodos de realce. Fonte: De autoria própria.

melhorias ao realce e, conseqüentemente, aos testes de identificação. Esses resultados se confirmam como resposta à Questão **Q1**, que pergunta como melhorar o aprendizado da estrutura de crista em modelos de aprendizagem profunda para o realce de imagens de impressões digitais latentes. Os blocos residuais permitem que a rede neural aprenda a identificar padrões mais complexos, o que pode ser benéfico para tarefas de tradução de imagem para imagem no contexto de impressões digitais latentes.

Na Figura 49, são apresentados alguns exemplos de realces de imagens de impressões digitais latentes gerados usando o modelo RM-net treinado apenas com a função de custo MAE, e aqueles gerados usando a rede treinada com a função de custo no domínio da frequência. Os círculos vermelhos denotam as minúcias identificadas pelo algoritmo de extração nas imagens de saída.

Como resposta à Questão **Q3**, é possível observar que a função de custo no domínio da frequência é uma estratégia viável para reduzir a criação de padrões de cristas errôneos e artificiais em áreas em que existem distorções severas nas imagens originais. Isso ocorre porque a função de custo no domínio da frequência se baseia na média das frequências

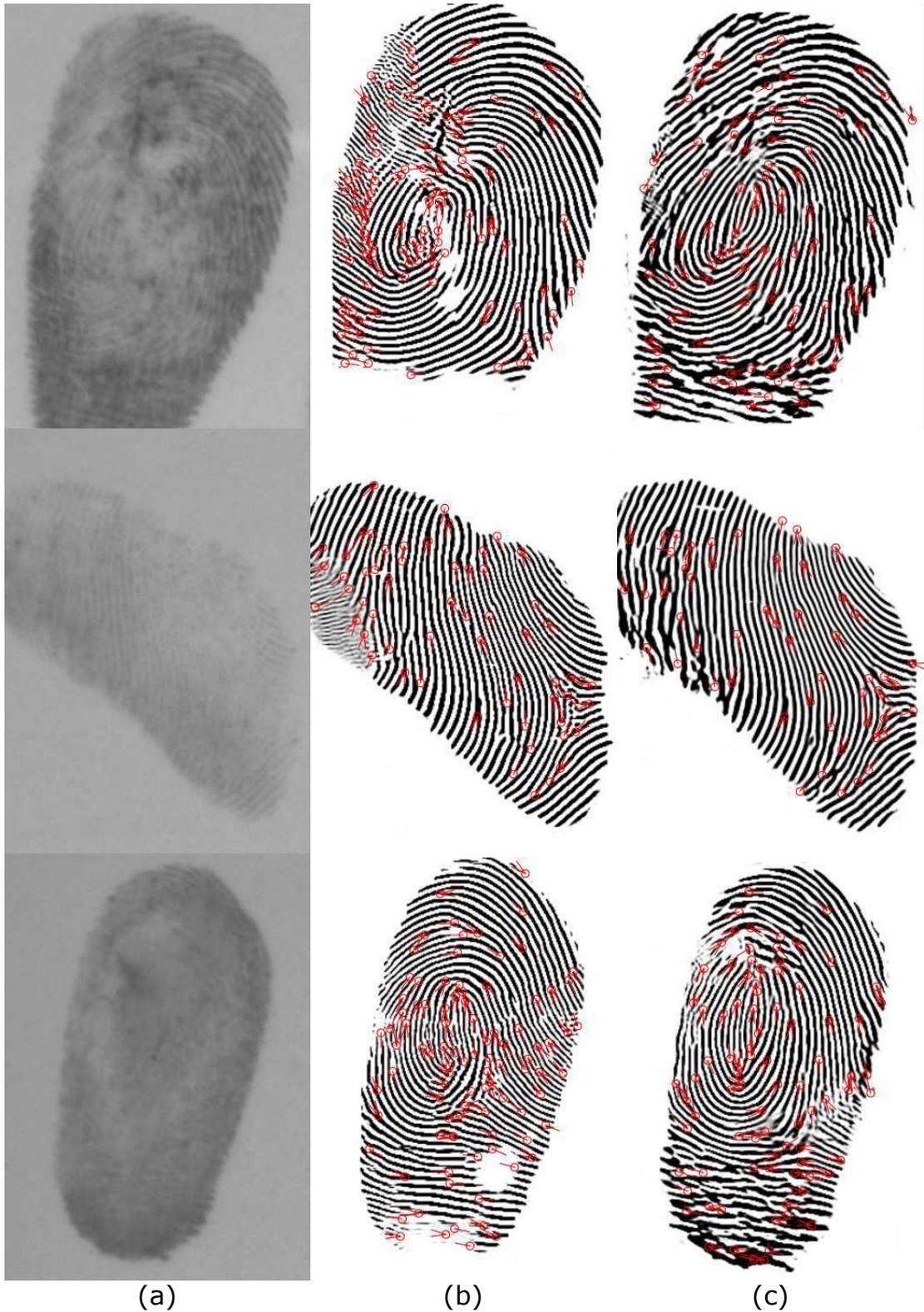


Figura 49: Comparação entre as imagens de saída da RM-net treinada sob diferentes condições: (b) utilizando a função de custo MAE e (c) utilizando a função de custo no domínio da frequência. (a) é a impressão digital latente original. Fonte: De autoria própria.

das saídas esperadas (*ground truths*) para realizar essas correções nos realces gerados.

Os resultados do realce revelam que o modelo RM-net treinado com a função de custo no domínio da frequência reduz o número de minúcias falsas identificadas nas imagens de saída, melhorando os resultados dos subsequentes testes de correspondência.

5.3.1 Qualidade e visualização do realce das imagens da base MOLF DB4

A qualidade das impressões digitais latentes reconstruídas pode ser avaliada por meio das pontuações atribuídas pela métrica NFIQ. Com o auxílio dessa métrica é possível observar o desempenho das reconstruções realizadas pelo método desenvolvido em comparação com as imagens originais. Conforme discutido na Seção 2.4, o NFIQ atribui pontuações mais baixas a imagens de melhor qualidade. Na Figura 50 é mostrado um gráfico com a diferença entre as qualidades das imagens antes e após o realce, conforme avaliado pelo NFIQ.

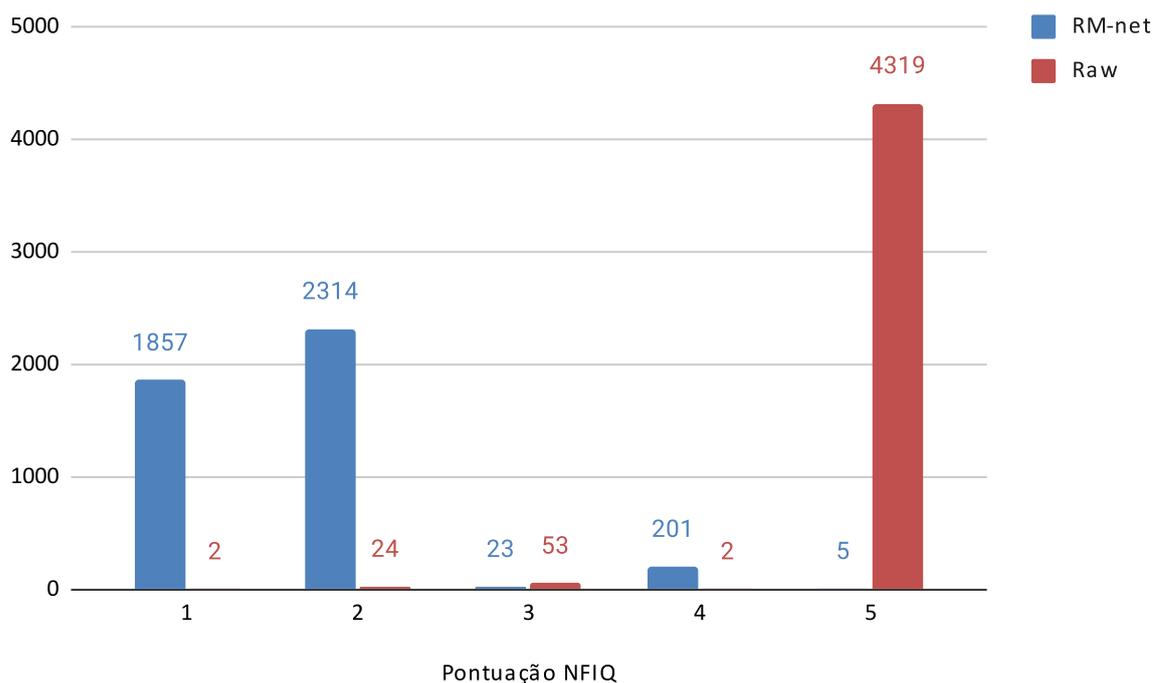


Figura 50: Comparação das pontuações de qualidade das imagens com e sem a aplicação do realce, de acordo com o NFIQ. Fonte: De autoria própria.

Conforme ilustrado no gráfico apresentado na Figura 50, observa-se que cinco imagens realçadas pelo método desenvolvido receberam a pontuação 5 no NFIQ. Essas imagens são exibidas na Figura 51. A atribuição dessa pontuação deve-se a duas situações distintas. Em alguns casos, a imagem realçada contém pouca informação de uma impressão digital, resultando em uma baixa quantidade de características identificáveis. Em



Figura 51: Imagens do realce que receberam pontuação 5, de acordo com o NFIQ. Fonte: De autoria própria.

outras situações, a atribuição da pontuação máxima ocorre devido à presença de ruído excessivo em relação à quantidade correta de estrutura de cristas reconstruída.

Na Figura 52 são mostrados alguns casos de sucesso da aplicação do realce das impressões digitais latentes pelo método desenvolvido nesta pesquisa. As imagens (a), (c), (e), (g), (i), (k), (m) e (o) da Figura 52 são impressões digitais latentes existentes no conjunto de dados MOLF DB4 para as quais o NFIQ atribuiu a pontuação de valor 5, isto é, são consideradas imagens de impressão digital de qualidade baixa. As imagens (b), (d), (f), (h), (j), (l), (n) e (p) da mesma figura são os respectivos realces.

Em contrapartida, existem casos em que alguns desafios podem ser observados. Métodos de realce com foco na reconstrução de cristas, ao tentar reconstruir partes da impressão digital que possuem distorções muito severas, costumam preencher essas áreas com padrões incorretos de cristas, portanto, terminam por introduzir minúcias que atrapalham a identificação da amostra (Dabouei et al., 2018). A aplicação da função de custo no domínio da frequência reduz esse problema, na maioria dos casos, mas não o resolve em definitivo. Um exemplo desse comportamento é mostrado na Figura 53, na qual as imagens (a), (c), (e), (g), (i), (k), (m) e (o) são impressões digitais latentes do conjunto de testes. As imagens (b), (d), (f), (h), (j), (l), (n) e (p) são os respectivos realces realizados pelo método desenvolvido nesta pesquisa. As áreas destacadas em vermelho são partes muito deterioradas da impressão digital original que foram preenchidas com uma estrutura de cristas imprecisa após o processo de realce.

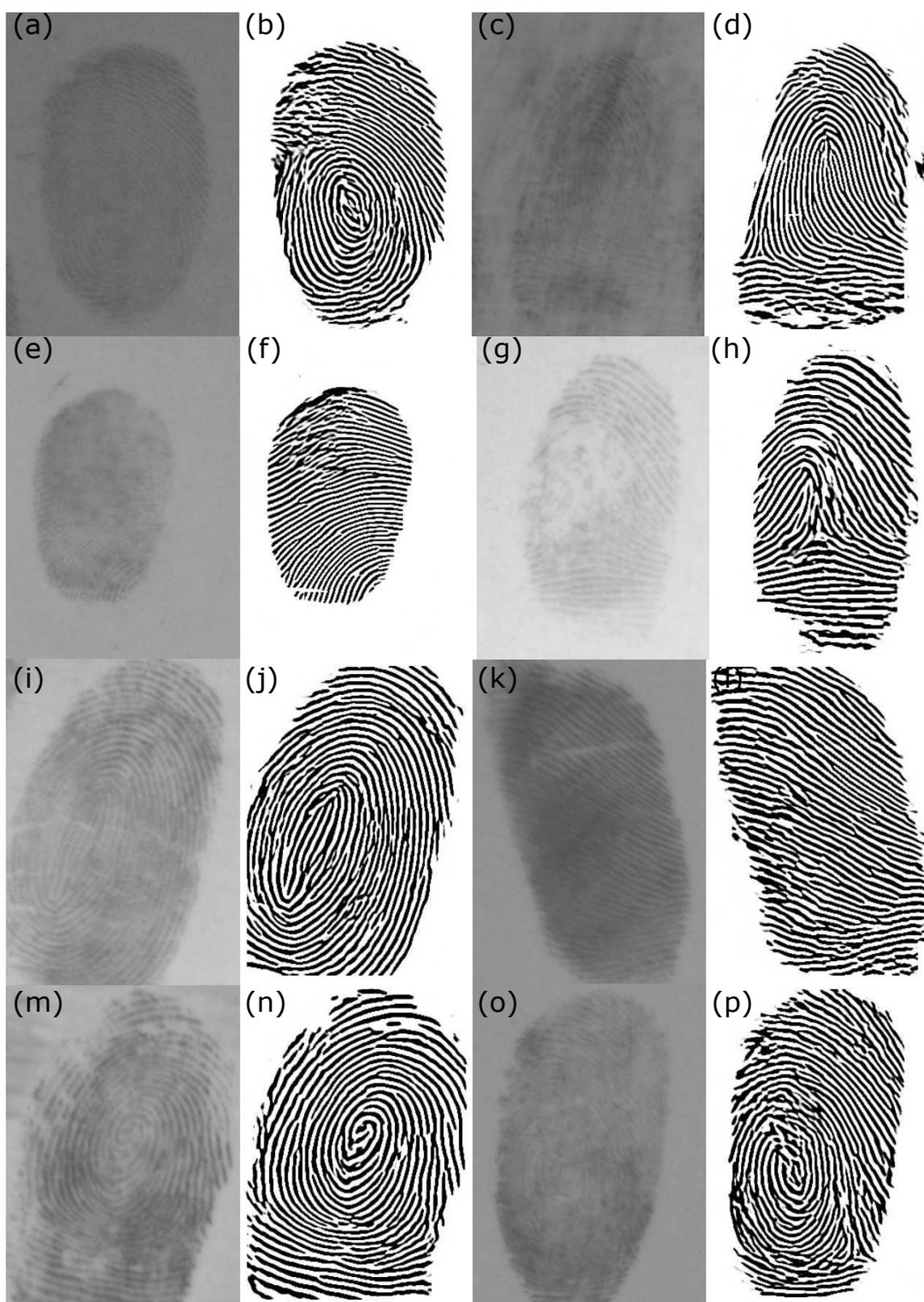


Figura 52: Casos em que o realce das imagens obteve uma identificação precisa no teste de casamento. Isto é, a impressão digital latente realçada foi corretamente associada à sua correspondente na galeria de testes DB1. Fonte: De autoria própria.

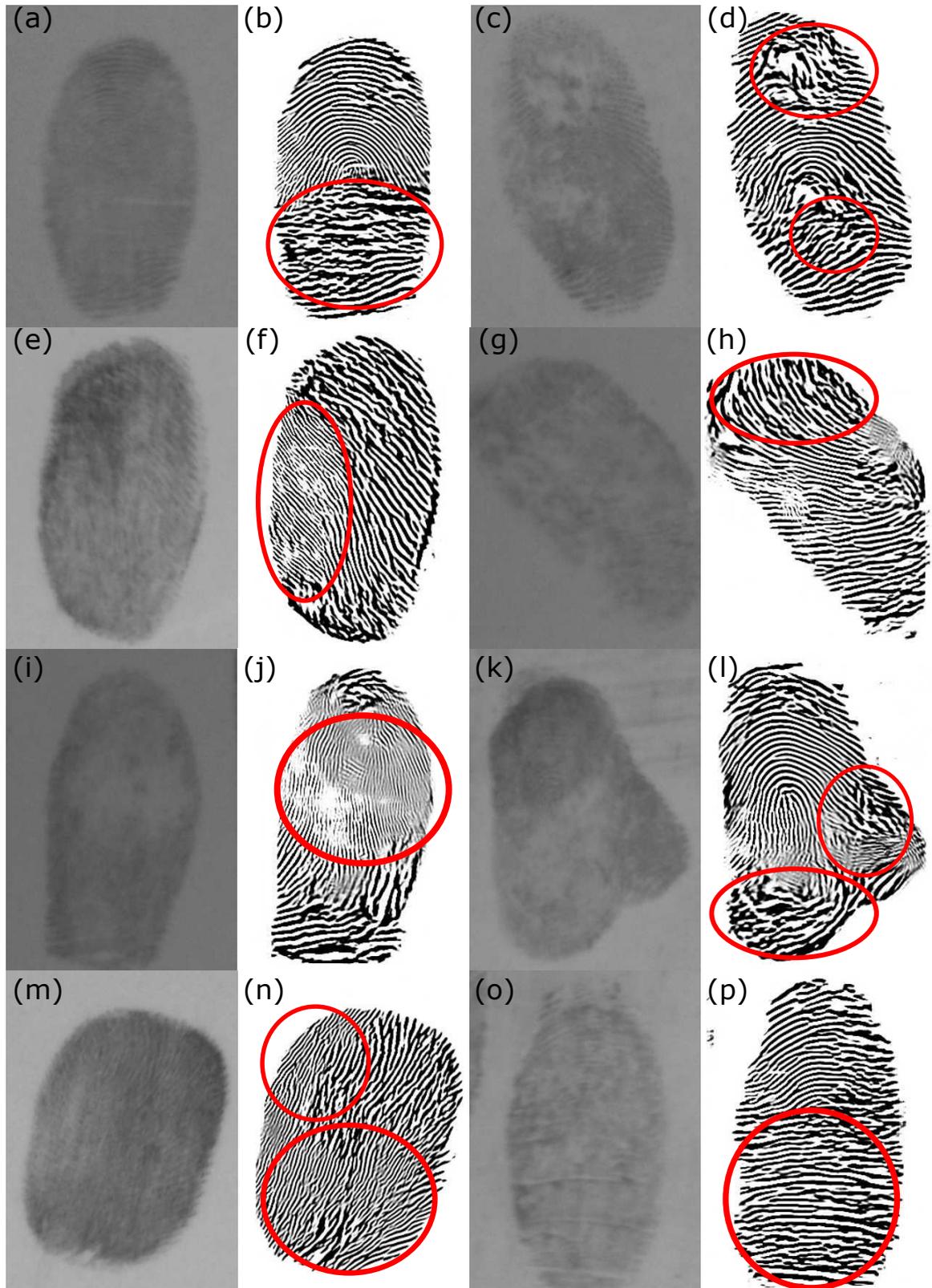


Figura 53: Casos em que o modelo preenche áreas muito deterioradas da impressão digital original com estrutura de cristas imprecisa na reconstrução. Fonte: De autoria própria.

5.4 Experimento latente-para-latente

Este experimento compreende a comparação entre impressões digitais latentes e utiliza a base de dados III-D *latent fingerprint database* como conjunto de testes. O experimento segue o mesmo protocolo estabelecido em Sankaran et al. (2011), o qual consiste na escolha aleatória de 395 imagens como galeria e 520 amostras como *probes*, garantindo que cada classe contenha pelo menos uma amostra na galeria. Os resultados são relatados em termos de acurácia (em porcentagem) nos *ranks* 1, 10 e 25. A Tabela 3 resume a comparação do resultado dos testes de casamento das imagens realçadas pelo método desenvolvido nesta pesquisa em relação às imagens sem processamento e também em comparação com os resultados dos métodos estado da arte Svoboda et al. (2017), Dabouei et al. (2018) e (Xu et al., 2020) publicados na literatura. Na Figura 54 é possível visualizar exemplos do realce realizado pelo método desenvolvido nesta pesquisa em imagens da base de dados III-D *latent fingerprint database*.

Tabela 3: Acurácias do experimento latente-para-latente usando o VeriFinger como extrator de minúcias e algoritmo de casamento.

Realce	Acurácia no <i>rank</i>		
	1	10	25
<i>Raw</i>	61,02	74,00	77,44
Svoboda et al. (2017)	71,04	82,56	88,28
Dabouei et al. (2018)	79,23	88,02	94,67
Xu et al. (2020)	86,86	92,91	95,12
RM-net	91,54	95,19	96,73

Também foram traçadas as curvas CMC para este experimento de correspondência de impressões digitais latentes, conforme apresentado na Figura 55, sendo observado que o método desenvolvido nesta pesquisa obteve valores de acurácia superiores em comparação com os demais métodos nos *ranks* 1 a 25. Destaca-se que os realces realizados pela RM-net resultaram em uma maior acurácia no *rank* 1, com um valor de 91,54%. Esses resultados sugerem que o método desenvolvido apresenta um desempenho superior na identificação correta de indivíduos nos *ranks* mais baixos, representando, assim, uma abordagem promissora para a tarefa de realce de impressões digitais latentes.

Assim como no experimento latente-para-sensor, também foi realizada uma análise comparativa da RM-net treinada com a função de custo no domínio da frequência (f. *loss*) em comparação com a RM-net treinada apenas com a função de custo MAE e com uma versão padrão da M-net, sem a inclusão de blocos residuais. Os resultados são apresentados na Tabela 4, destacando-se a contribuição da função de custo no domínio da frequência para aumentar as acurácias obtidas no cenário em que é testada a correspondência entre as impressões digitais latentes. Isso evidencia a importância da abordagem do



Figura 54: Amostras de impressões digitais latentes da base de dados III-D *latent fingerprint database* na primeira linha, juntamente com seus respectivos realces realizados pela RM-net na segunda linha.

método desenvolvido, que incorpora informações no domínio da frequência para melhorar a precisão da identificação e correspondência entre impressões digitais latentes.

Tabela 4: Acurácias *rank-n* obtidos no experimento latente-para-latente, comparando os resultados da rede neural treinada utilizando apenas a função de perda MAE e a função de perda no domínio da frequência.

Função de custo	Acurácia		
	<i>Rank-1</i>	<i>Rank-10</i>	<i>Rank-25</i>
M-net	90,38	92,69	93,85
RM-net MAE	87,11	92,31	93,46
RM-net (f. loss)	91,54	95,19	96,73

Os resultados apresentados nas Tabelas 2 e 4 sugerem que as técnicas de processamento de imagens no domínio da frequência, normalmente aplicadas em problemas de tradução de imagem para imagem de contexto geral, também podem apresentar contribuição no contexto do realce de impressões digitais latentes, em resposta à Questão Q2. Especificamente, os testes de casamento que utilizaram as imagens realçadas pela RM-net (f. loss), que incorpora uma função de perda no domínio da frequência, demonstraram consistentemente um desempenho superior em comparação com outras redes neurais em

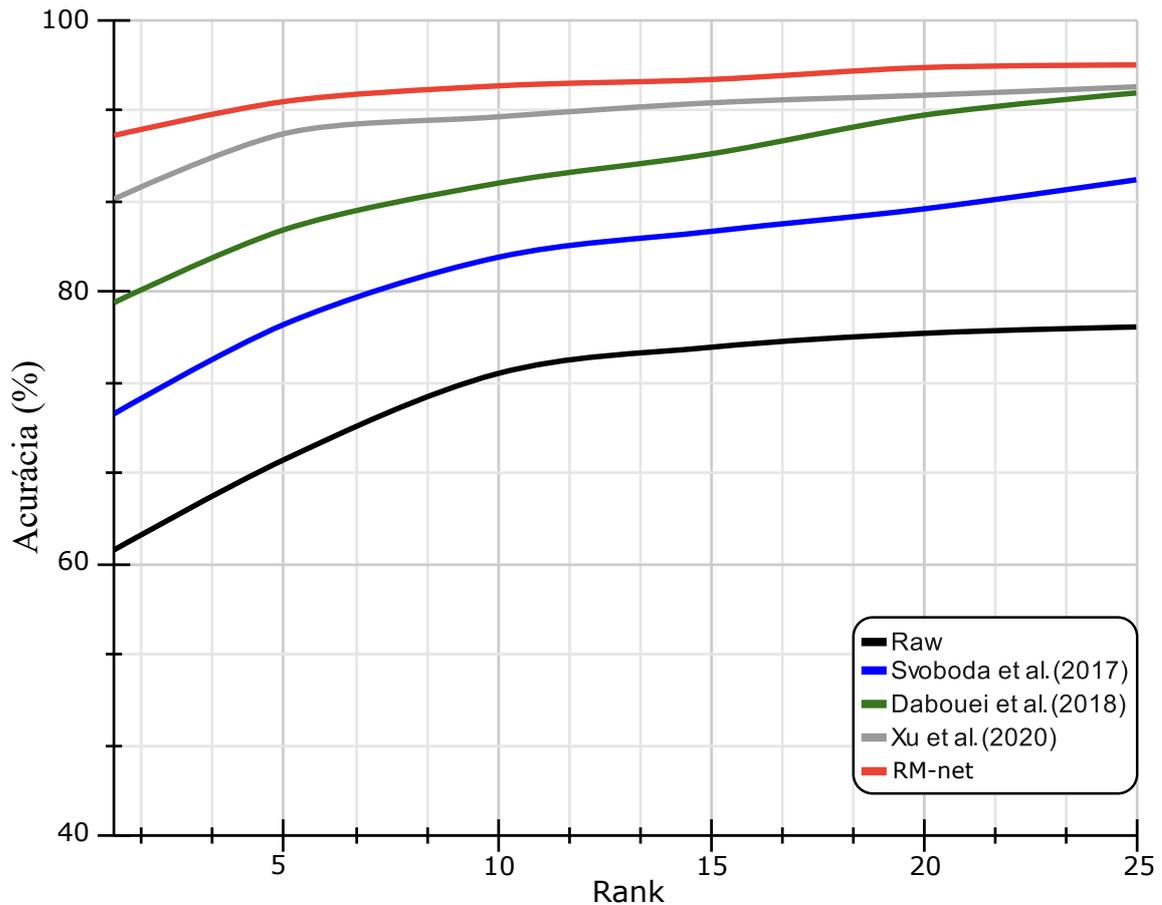


Figura 55: Curvas CMC referentes aos *ranks* do experimento latente-para-latente, comparando as imagens raw e também os métodos estado-da-arte publicados na literatura.

todos os *ranks* comparados.

5.5 Considerações finais do capítulo

Neste capítulo foi apresentada uma análise dos resultados obtidos pelo método desenvolvido nesta pesquisa. Observa-se que o realce realizado pela RM-net contribuiu para a obtenção de melhores valores de acurácias nos resultados dos testes de correspondência das impressões digitais latentes no conjunto de dados de teste. Por exemplo, na galeria DB1 a acurácia no *rank*-50 antes da aplicação do realce é de 6,06% e, após o realce, esse valor foi elevado para 52,25%. Em comparação com os principais métodos identificados na literatura, o método desenvolvido demonstrou resultados competitivos, principalmente em *ranks* mais baixos, com destaque para seu desempenho no *rank*-1.

Foram apresentadas ilustrações e exemplos de casos em que o realce das imagens das impressões digitais latentes realizado pelo método desenvolvido produziu melhorias nos resultados da identificação, bem como casos em que as reconstruções resultaram em

preenchimento de áreas das imagens com informações incorretas ou imprecisas.

No próximo capítulo deste documento de tese são apresentadas algumas discussões e considerações finais sobre esta pesquisa realizada e sobre o método desenvolvido, bem como algumas possibilidades de melhorias e propostas de trabalhos futuros.

6 Considerações Finais

Neste capítulo são realizadas algumas considerações sobre a pesquisa conduzida nesta tese de doutorado. Inicialmente, são realizadas algumas discussões com base nos resultados obtidos em comparação aos objetivos e perguntas definidos no Capítulo 1. Por fim, são tecidas algumas conclusões e apresentadas propostas de trabalhos futuros.

6.1 Considerações sobre a pesquisa proposta

Nesta tese, foi desenvolvido um método de aprendizagem profunda com o objetivo de realizar o realce de imagens de impressões digitais latentes. O modelo desenvolvido é fruto de uma adaptação de uma arquitetura de autocodificador conhecida como M-net, denominada neste trabalho como RM-net. A rede desenvolvida recebe como entrada uma imagem de impressão digital latente, sem a necessidade de aplicação de pré-processamento, e entrega na saída uma versão realçada da estrutura de cristas dessa imagem.

A arquitetura do modelo consiste em três componentes principais: (i) o codificador, responsável pela redução da dimensionalidade e criação da representação latente da imagem de entrada; (ii) o decodificador, responsável pela reconstrução da versão realçada da imagem de impressão digital latente fornecida como entrada e (iii) os blocos residuais. O aperfeiçoamento da aprendizagem do modelo de rede neural escolhido foi conseguido por meio da modificação da arquitetura da rede M-net, incorporando blocos residuais após a última camada do codificador.

Durante o desenvolvimento da arquitetura e treinamento da RM-net, foram incorporadas à rede técnicas de processamento de imagens no domínio da frequência para lidar com o problema de reconstrução não natural e errônea em áreas onde a imagem original apresenta muito ruído ou pouca informação. Essas técnicas, como a DCT e a DFT, foram adicionadas como complemento à função de custo, permitindo a comparação das representações das saídas originais e das saídas previstas no domínio da frequência. Esse fato se apresenta como resposta à Questão **Q4**, mostrando que a incorporação da DCT e da DFT impactou positivamente nas imagens geradas pela rede, resultando em uma redução na ocorrência de falsas minúcias e na melhoria dos resultados de identificação por parte dos algoritmos de extração e casamento.

Devido à indisponibilidade de uma base de dados de impressões digitais latentes reais juntamente com seus respectivos *ground truths*, uma etapa crucial desta pesquisa foi a construção de uma base de dados sintética para o treinamento da rede desenvolvida. Para criar essa base de dados, foram geradas imagens sintéticas de impressões digitais utilizando uma versão de demonstração do *software* SFinGe que é capaz de gerar impressões digitais e seus respectivos *ground truths*. Devido a limitações presentes na versão de demonstração

do programa, foi necessário desenvolver um método automático para gerar variações e capturar essas imagens.

Após a criação da base de dados sintética, procedeu-se à incorporação de distorções nas imagens geradas, visando simular características presentes em impressões digitais latentes reais. Essas distorções foram selecionadas e aplicadas na tentativa de representar algumas peculiaridades e nuances observadas em conjuntos de dados reais de impressões digitais latentes. O objetivo principal dessa abordagem foi obter uma base de dados diversificada, abrangendo uma ampla variedade de cenários e condições, a fim de aprimorar a capacidade de generalização da rede neural desenvolvida.

O objetivo principal da realização do realce em impressões digitais é aprimorar a qualidade das imagens de entrada, visando auxiliar na melhora dos resultados obtidos por algoritmos de identificação e correspondência. Nesse contexto, o desempenho do método desenvolvido foi avaliado por meio de testes de casamento de impressões digitais, empregando um dos algoritmos mais amplamente utilizado nas pesquisas da área, o VeriFinger. Os experimentos foram conduzidos em dois cenários distintos, a correspondência entre impressões digitais latentes e aquelas obtidas por meio de sensores, bem como a correspondência entre impressões digitais latentes com outras impressões digitais latentes.

6.2 Considerações sobre os resultados obtidos

No capítulo 4 foi explicado que os melhores resultados dos testes de correspondência foram obtidos ao se utilizar o realce realizado pela rede que foi treinada com uma função de custo que combina a função MAE com a DCT e a DFT. Sobre isso, é válido tecer algumas considerações, conforme discutido a seguir.

Embora ambas as técnicas sejam utilizadas para análise de sinais no domínio da frequência, existem algumas diferenças entre elas. A DCT é uma transformada real, operando apenas em números reais e produzindo coeficientes reais como resultado. A DCT utiliza cossenos para representar as diferentes frequências em cada segmento e é conhecida por sua eficiência na concentração de energia em coeficientes de baixa frequência para dados de imagem, de acordo com Roy et al. (2012). Por outro lado, a DFT é uma transformada complexa, que opera em números complexos e produz coeficientes reais e imaginários como resultado (Cariolaro et al., 2002; Asmara et al., 2017). Essas diferenças resultam em representações distintas das imagens no domínio da frequência. Ao incorporar tanto a DCT quanto a DFT na função de custo durante o treinamento da rede neural, torna-se possível explorar as características específicas de cada técnica.

Sobre os valores de acurácias obtidos nos testes de correspondência. A análise dos resultados indica que o método desenvolvido se apresenta como uma abordagem viável para o realce de imagens de impressões digitais latentes. Inicialmente, ao comparar as

imagens sem realce com aquelas submetidas ao método proposto, verificou-se uma melhoria na qualidade das imagens, resultando em uma taxa de identificação aprimorada pelos algoritmos de casamento. Além disso, em uma análise comparativa com os métodos identificados durante a revisão da literatura, constatou-se que o método desenvolvido alcançou os melhores valores de acurácia nos *ranks*-1 ao 30 no experimento latente-para-sensor e também demonstrou um desempenho consistente no experimento latente-para-latente, obtendo os melhores valores de acurácia do *rank*-1 ao 25.

Com relação à comparação dos resultados e diferenças entre os métodos RM-net e PIDI, proposto por Dabouei et al. (2018), que apresenta os melhores resultados no *rank*-50 para a base de dados MOLF, destacam-se algumas considerações. A RM-net, como mencionado anteriormente, é um método que se baseia na reconstrução das cristas fazendo a comparação em relação às saídas reais esperadas. A adição de blocos residuais à rede contribuiu para o aprendizado de estruturas mais complexas, permitindo que a rede busque reconstruir a maior quantidade possível de informações presentes nas impressões digitais. No entanto, essa abordagem apresenta uma limitação já mencionada, pois a rede tende a reconstruir áreas nas quais há informações pouco claras, resultando na geração de ruído e, conseqüentemente, falsas minúcias. Esse problema foi mitigado pela função de custo no domínio da frequência adotada, que realiza pequenas correções nas frequências das cristas, reduzindo assim a ocorrência dessas falsas minúcias. A hipótese subjacente é que, mesmo com a presença de falsas minúcias, as informações corretas de identificação das impressões digitais estão contidas nos realces gerados pelo método, permitindo que algoritmos de casamento utilizem as informações verdadeiras e sendo capazes de descartar parte das informações falsas. Isso resulta em melhores resultados de identificação em *ranks* mais baixos ao utilizar as imagens realçadas pelo método RM-net.

Por outro lado, o método PIDI adota uma abordagem baseada em GANs, visando realçar apenas as áreas em que existem informações consistentes de identificação. Para isso, são utilizados mapas de frequência, orientação, região de interesse e uma rede siamesa adicional que foi treinada para extrair informações perceptuais desses mapas. A rede generativa reconstrói apenas as áreas que são consideradas relevantes com base nessas informações. Essa abordagem sugere que o método descarta informações que a rede RM-net é capaz de recuperar. A estratégia adotada por esse método possui a vantagem de realizar um realce focado em regiões que contêm informações mais confiáveis, resultando em melhores resultados em *ranks* mais altos. No entanto, essa abordagem parece comprometer os resultados em *ranks* mais baixos.

Os experimentos com o conjunto sintético de testes, em que são observados os resultados obtidos da métrica SSIM, mostram a capacidade do método de manter a similaridade entre as imagens realçadas e as saídas esperadas (*ground truths*). Dentro do contexto de realce de imagens de impressões digitais latentes, esta é uma propriedade

relevante, que indica que o método é capaz de preservar a estrutura de cristas existente e, conseqüentemente, as características necessárias para a utilização da imagem realçada no processo de identificação. No que diz respeito à métrica de qualidade NFIQ, observou-se que as imagens sem realce apresentaram um valor médio de 4,96, enquanto após a aplicação do realce, esse valor foi reduzido para 1,63 (sendo que, para o NFIQ, valores menores indicam melhor qualidade). Esse resultado demonstra que o método desenvolvido foi capaz de efetivamente melhorar a qualidade das imagens.

6.3 Propostas de trabalhos futuros

Como propostas de trabalhos futuros, visando o aprimoramento do método e outras contribuições com o campo de pesquisa relacionado, são destacadas as descritas a seguir:

- **Aprimoramento do método:** uma direção promissora seria a adaptação da RM-net para uma rede generativa especializada em remoção de ruído e realce. Isso envolveria a exploração de arquiteturas mais avançadas, como redes adversárias generativas ou redes autocodificadoras variacionais, que são capazes de aprender características mais complexas beneficiando-se do treinamento *adversarial* com uma rede discriminador.
- **Utilização de informações de orientação da imagem:** uma outra direção de pesquisa seria a incorporação de informações de orientação das imagens de impressão digital durante o processo de treinamento da rede. Tal abordagem visa preservar as regiões das impressões digitais latentes mais recuperáveis, e também evitar a reconstrução inadequada de áreas problemáticas.
- **Aprimoramento da base de dados sintética:** a fim de tornar a base de dados sintética ainda mais representativa das características das impressões digitais latentes reais, seria válido aprimorar o método de criação dessa base. Isso envolveria a pesquisa de como realizar a incorporação de variações mais realistas, considerando diferentes tipos de impressões digitais latentes, tais como impressões obtidas em diferentes superfícies e condições ambientais. Além disso, a inclusão de variados graus de degradação e ruído nas imagens sintéticas poderia proporcionar um conjunto de dados capaz de aprimorar a generalização do método desenvolvido nesta pesquisa.
- **Treinamento com bases de dados sintéticas geradas por métodos generativos:** uma abordagem interessante seria explorar bases de dados sintéticas geradas por métodos generativos publicados na literatura. Essas bases de dados poderiam ser utilizadas para treinar e avaliar o método desenvolvido, comparando seu desempenho com outros métodos existentes. Isso proporcionaria uma avaliação mais abrangente e

uma comparação direta com as abordagens mais recentes, demonstrando o potencial e as vantagens do método desenvolvido em relação a essas técnicas.

- Adaptação e implementação do método de realce do MSU *Latent AFIS*: para possibilitar uma comparação com os resultados obtidos utilizando o realce da RM-net, propõe-se a adaptação e implementação do método de realce do MSU *Latent AFIS* isolando-o das demais funcionalidades do sistema. Desse modo, é possível realizar o treinamento do método com a base de dados desenvolvida nesta pesquisa e obter-se os valores de acurácia para a base de dados MOLF.

REFERÊNCIAS

- Adiga, S. V. and Sivaswamy, J. (2019). Fpd-m-net: Fingerprint image denoising and inpainting using m-net based convolutional neural networks. In *Inpainting and Denoising Challenges*, pages 51–61. Springer.
- Ahmed, N., Natarajan, T., and Rao, K. R. (1974). Discrete cosine transform. *IEEE Transactions on Computers*, C-23(1):90–93.
- Aithal, S. and Karani, K. P. (2017). Fingerprint image segmentation: A review of state of the art techniques. *International Journal of Management, Technology, and Social Sciences (IJMTS)*, 2:28–39.
- Ansari, A. H. (2011). Generation and storage of large synthetic fingerprint database. Master’s thesis, Indian Institute of Science, Bangalore – 560 012.
- Asmara, R. A., Agustina, R., and Hidayatulloh (2017). Comparison of discrete cosine transforms (dct), discrete fourier transforms (dft), and discrete wavelet transforms (dwt) in digital image watermarking. *International Journal of Advanced Computer Science and Applications*, 8.
- Azzouni, A., Boutaba, R., and Pujolle, G. (2017). Neuroute: Predictive dynamic routing for software-defined networks. In *2017 13th International conference on network and service management (CNSM)*, pages 1–6. IEEE.
- Bank, D., Koenigstein, N., and Giryes, R. (2020). Autoencoders. *arXiv preprint arXiv:2003.05991*.
- Baur, C., Wiestler, B., Albarqouni, S., and Navab, N. (2018). Deep autoencoding models for unsupervised anomaly segmentation in brain MR images. In *International MICCAI Brainlesion Workshop*, pages 161–169. Springer.
- Bjorck, J., Gomes, C., Selman, B., and Weinberger, K. Q. (2018). Understanding batch normalization. *arXiv preprint arXiv:1806.02375*.
- Bracewell, R. N. (2000). *The Fourier transform and its applications*. McGraw-Hill, 3rd edition.
- Cao, K. and Jain, A. K. (2019). Automated latent fingerprint recognition. *IEEE transactions on pattern analysis and machine intelligence*, 41(4):788–800.
- Cao, K., Liu, E., and Jain, A. K. (2014). Segmentation and enhancement of latent fingerprints: A coarse to fine ridgestructure dictionary. *IEEE transactions on pattern analysis and machine intelligence*, 36(9):1847–1859.
- Cao, K., Nguyen, D.-L., Tymoszek, C., and Jain, A. K. (2019). End-to-end latent fingerprint search. *IEEE Transactions on Information Forensics and Security*, 15:880–894.

- Cappelli, R., Ferrara, M., and Maltoni, D. (2010). Minutia cylinder-code: A new representation and matching technique for fingerprint recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(12):2128–2141.
- Cappelli, R., Maio, D., and Maltoni, D. (2002). Synthetic fingerprint-database generation. In *Object recognition supported by user interaction for service robots*, volume 3, pages 744–747. IEEE.
- Cariolaro, G., Erseghe, T., and Kraniuskas, P. (2002). The fractional discrete cosine transform. *IEEE Transactions on Signal Processing*, 50(4):902–911.
- Champod, C. and Chamberlain, P. (2009). *Handbook of Forensic Science*, chapter 3. Routledge.
- Champod, C., Lennard, C. J., Margot, P., and Stoilovic, M. (2004). *Fingerprints and other ridge skin impressions*. CRC press.
- Chavan, S., Mundada, P., and Pal, D. (2015). Fingerprint authentication using gabor filter based matching algorithm. In *2015 International Conference on Technologies for Sustainable Development (ICTSD)*, pages 1–6. IEEE.
- Chugh, T., Cao, K., Zhou, J., Tabassi, E., and Jain, A. K. (2018). Latent fingerprint value prediction: Crowd-based learning. *IEEE Transactions on Information Forensics and Security*, 13(1):20–34.
- Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., and Ronneberger, O. (2016). 3d u-net: learning dense volumetric segmentation from sparse annotation. In *International conference on medical image computing and computer-assisted intervention*, pages 424–432. Springer.
- Dabouei, A., Kazemi, H., Iranmanesh, S. M., Dawson, J., Nasrabadi, N. M., et al. (2018). Id preserving generative adversarial network for partial latent fingerprint reconstruction. In *2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pages 1–10. IEEE.
- Deshpande, U. U., Malemath, V., Patil, S. M., and Chaugule, S. V. (2020). End-to-end automated latent fingerprint identification with improved dcnn-fft enhancement. *Frontiers in Robotics and AI*, 7:594412.
- Dessimoz, D. and Champod, C. (2008). Linkages between biometrics and forensic science. In *Handbook of biometrics*, pages 425–459. Springer.
- Dror, I. E., Wertheim, K., Fraser-Mackenzie, P., and Walajtys, J. (2012). The impact of human-technology cooperation and distributed cognition in forensic science: biasing effects of afis contextual information on human experts. *Journal of forensic sciences*, 57(2):343–352.
- Dvornychenko, V. and Garris, M. D. (2006). *Summary of NIST latent fingerprint testing workshop*. US Department of Commerce, Technology Administration, National Institute of . . .

- Fan, C.-M., Liu, T.-J., and Liu, K.-H. (2022a). Half wavelet attention on m-net+ for low-light image enhancement.
- Fan, C.-M., Liu, T.-J., Liu, K.-H., and Chiu, C.-H. (2022b). Selective residual m-net for real image denoising. In *2022 30th European Signal Processing Conference (EUSIPCO)*, pages 469–473.
- Feng, J., Zhou, J., and Jain, A. K. (2013). Orientation field estimation for latent fingerprint enhancement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(4):925–940.
- Fierrez, J., Ortega-Garcia, J., Toledano, D. T., and Gonzalez-Rodriguez, J. (2007). Biosec baseline corpus: A multimodal biometric database. *Pattern Recognition*, 40(4):1389–1392.
- Fiumara, G., Flanagan, P., Grantham, J., Bandini, B., Ko, K., and Libert, J. (2018). National institute of standards and technology special database 300:uncompressed plain and rolled images from fingerprint cards. Technical Note 1993, National Institute of Standards and Technology.
- Fraser, J. and Williams, R. (2013). *Handbook of forensic science*. Routledge.
- Ghosh, A., Sufian, A., Sultana, F., Chakrabarti, A., and De, D. (2020). Fundamental concepts of convolutional neural network. In *Recent Trends and Advances in Artificial Intelligence and Internet of Things*, pages 519–567. Springer.
- Gonzales, R. and Woods, R. (2018). Digital image processing 4th edition.
- Gonzalez, R. C. and Woods, R. C. (2009). *Processamento digital de imagens*. Pearson Educación.
- Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press. <http://www.deeplearningbook.org>.
- Guan, H., Dienstfrey, A. M., and Theofanos, M. F. (2013). A new metric for latent fingerprint image preprocessing. In *2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 84–91.
- Guissois, A. E. (2019). Skin Lesion Classification Using Deep Neural Network. Master’s thesis, University of Mohamed El-Bachir El-Ibrahimi.
- Haber, L. and Haber, R. N. (2007). Scientific validation of fingerprint evidence under Daubert. *Law, Probability and Risk*, 7(2):87–109.
- Hawthorne, M. (2017). *Fingerprints: analysis and understanding*. CRC Press.
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.

- Hong, L., Wan, Y., and Jain, A. (1998). Fingerprint image enhancement: algorithm and performance evaluation. *IEEE transactions on pattern analysis and machine intelligence*, 20(8):777–789.
- Horapong, K., Srisutheenon, K., and Areekul, V. (2021). Progressive and corrective feedback for latent fingerprint enhancement using boosted spectral filtering and spectral autoencoder. *IEEE Access*, 9:96288–96308.
- Hore, A. and Ziou, D. (2010). Image quality metrics: Psnr vs. ssim. In *2010 20th international conference on pattern recognition*, pages 2366–2369. IEEE.
- Houck, M. M. (2006). Csi: reality. *Scientific American*, 295(1):84–89.
- Huang, X., Qian, P., and Liu, M. (2020). Latent fingerprint image enhancement based on progressive generative adversarial network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 800–801.
- Indovina, M., Dvornychenko, V., Hicklin, R., Kiebusinski, G., et al. (2012). Elft-efs evaluation of latent fingerprint technologies: extended feature sets [evaluation# 2]. *National Institute of Standards and Technology, US Department of Commerce NISTIR*, 7859:535.
- Ioffe, S. and Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, pages 448–456. PMLR.
- Jain, A., Bolle, R., and Pankanti, S. (1999). *Biometrics: personal identification in networked society*, volume 479. Springer Science & Business Media.
- Jain, A. K. (1989). *Fundamentals of digital image processing*. Prentice-Hall, Inc.
- Jain, A. K., Chen, Y., and Demirkus, M. (2006). Pores and ridges: High-resolution fingerprint matching using level 3 features. *IEEE transactions on pattern analysis and machine intelligence*, 29(1):15–27.
- Jain, A. K., Deb, D., and Engelsma, J. J. (2022). Biometrics: Trust, but verify. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 4(3):303–323.
- Jain, A. K. and Feng, J. (2010). Latent fingerprint matching. *IEEE Transactions on pattern analysis and machine intelligence*, 33(1):88–100.
- Jain, A. K., Flynn, P., and Ross, A. A. (2007). *Handbook of biometrics*. Springer Science & Business Media.
- Jain, A. K., Prabhakar, S., Hong, L., and Pankanti, S. (2000). Filterbank-based fingerprint matching. *IEEE transactions on Image Processing*, 9(5):846–859.
- Jain, A. K. and Ross, A. (2015). Bridging the gap: from biometrics to forensics. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1674):20140254.

- Jain, A. K., Ross, A. A., and Nandakumar, K. (2011). *Introduction to biometrics*. Springer Science & Business Media.
- Joshi, I., Anand, A., Vatsa, M., Singh, R., Roy, S. D., and Kalra, P. (2019). Latent fingerprint enhancement using generative adversarial networks. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 895–903. IEEE.
- Joshi, I., Utkarsh, A., Singh, P., Dantcheva, A., Roy, S. D., and Kalra, P. K. (2022). On restoration of degraded fingerprints. *Multimedia Tools and Applications*, 81(24):35349–35377.
- Kaplan, A. and Haenlein, M. (2019). Siri, siri, in my hand: Who’s the fairest in the land? on the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons*, 62(1):15–25.
- Karimpouli, S. and Tahmasebi, P. (2019). Segmentation of digital rock images using deep convolutional autoencoder networks. *Computers & Geosciences*, 126:142–150.
- LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521(7553):436–444.
- Li, J., Feng, J., and Kuo, C.-C. J. (2017). Deep convolutional neural network for latent fingerprint enhancement. *Signal Processing: Image Communication*, 60:52–63.
- Liu, C., Tang, C., Xu, M., and Lei, Z. (2020). Binarization of espi fringe patterns based on an m-net convolutional neural network. *Applied Optics*, 59(30):9598–9606.
- Liu, M. and Qian, P. (2020). Automatic segmentation and enhancement of latent fingerprints using deep nested unets. *IEEE Transactions on Information Forensics and Security*, 16:1709–1719.
- Liu, Y., Tang, Y., Li, R., and Feng, J. (2019). Cooperative orientation generative adversarial network for latent fingerprint enhancement. In *2019 International Conference on Biometrics (ICB)*, pages 1–8. IEEE.
- Maio, D., Maltoni, D., Cappelli, R., Wayman, J. L., and Jain, A. K. (2002). Fvc2002: Second fingerprint verification competition. In *Object recognition supported by user interaction for service robots*, volume 3, pages 811–814. IEEE.
- Maio, D., Maltoni, D., Cappelli, R., Wayman, J. L., and Jain, A. K. (2004). Fvc2004: Third fingerprint verification competition. In *International conference on biometric authentication*, pages 1–7. Springer.
- Maltoni, D. (2004). Generation of synthetic fingerprint image databases. In *Automatic Fingerprint Recognition Systems*, pages 361–384. Springer.
- Maltoni, D., Maio, D., Jain, A. K., and Prabhakar, S. (2009a). *Handbook of fingerprint recognition*. Springer Science & Business Media.

- Maltoni, D., Maio, D., Jain, A. K., and Prabhakar, S. (2009b). Synthetic fingerprint generation. *Handbook of fingerprint recognition*, pages 271–302.
- Mehta, R. and Sivaswamy, J. (2017). M-net: A convolutional neural network for deep brain structure segmentation. In *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, pages 437–440. IEEE.
- Michalski, R. S., Carbonell, J. G., and Mitchell, T. M. (2013). *Machine learning: An artificial intelligence approach*. Springer Science & Business Media.
- Msiza, I. S., Mistry, J., Leke-Betechuoh, B., Nelwamondo, F. V., and Marwala, T. (2011). On the introduction of secondary fingerprint classification. In *State of the art in Biometrics*, page 314. IntechOpen.
- Nair, V. and Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th International Conference on International Conference on Machine Learning, ICML’10*, page 807–814, Madison, WI, USA. Omnipress.
- Nguyen, D.-L., Cao, K., and Jain, A. K. (2018a). Automatic latent fingerprint segmentation. In *IEEE International Conference on BTAS*.
- Nguyen, D.-L., Cao, K., and Jain, A. K. (2018b). Robust minutiae extractor: Integrating deep networks and fingerprint domain knowledge. In *The 11th International Conference on Biometrics, 2018*.
- Nussbaumer, H. J. and Nussbaumer, H. J. (1981). *The fast Fourier transform*. Springer.
- Oppenheim, A. and Schaffer, R. (2010). Discrete-time signal processing (vol. xxviii, p. 1108).
- Orczyk, T. and Wieclaw, L. (2011). Fingerprint ridges frequency. In *2011 Third World Congress on Nature and Biologically Inspired Computing*, pages 558–561.
- Petscharnig, S., Lux, M., and Chatzichristofis, S. (2017). Dimensionality reduction for image features using deep learning and autoencoders. In *Proceedings of the 15th international workshop on content-based multimedia indexing*, pages 1–6.
- Pfeiffer, J., Broscheit, S., Gemulla, R., and Göschl, M. (2018). A neural autoencoder approach for document ranking and query refinement in pharmacogenomic information retrieval. In *Proceedings of the BioNLP 2018 workshop*, pages 87–97.
- Poh, N., Chan, C., Kittler, J., Fierrez, J., and Galbally, J. (2012). D3. 3: Description of metrics for the evaluation of biometric performance. *Biometrics Evaluation and Testing*, 1.
- Pratt, W. K. (2007). *Digital Image Processing: PIKS Inside*. Wiley, 4th edition.
- Qi, J., Du, J., Siniscalchi, S. M., Ma, X., and Lee, C.-H. (2020). On mean absolute error for deep neural network based vector-to-vector regression. *IEEE Signal Processing Letters*, 27:1485–1489.

- Qian, P., Li, A., and Liu, M. (2019). Latent fingerprint enhancement based on denseunet. In *2019 International Conference on Biometrics (ICB)*, pages 1–6. IEEE.
- Rasamoelina, A. D., Adjailia, F., and Sinčák, P. (2020). A review of activation function for artificial neural network. In *2020 IEEE 18th World Symposium on Applied Machine Intelligence and Informatics (SAMII)*, pages 281–286. IEEE.
- Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer.
- Roy, A. B., Dey, D., Mohanty, B., and Banerjee, D. (2012). Comparison of fft, dct, dwt, wht compression techniques on electrocardiogram and photoplethysmography signals. In *IJCA Special Issue on International Conference on Computing, Communication and Sensor Network CCSN*, pages 6–11.
- Sabhanayagam, T., Venkatesan, V. P., and Senthamaraiannan, K. (2018). A comprehensive survey on various biometric systems. *International Journal of Applied Engineering Research*, 13(5):2276–2297.
- Sakurada, M. and Yairi, T. (2014). Anomaly detection using autoencoders with nonlinear dimensionality reduction. In *Proceedings of the MLSDA 2014 2nd Workshop on Machine Learning for Sensory Data Analysis*, pages 4–11.
- Sankaran, A., Agarwal, A., Keshari, R., Ghosh, S., Sharma, A., Vatsa, M., and Singh, R. (2015a). Latent fingerprint from multiple surfaces: Database and quality analysis. In *2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pages 1–6. IEEE.
- Sankaran, A., Dhamecha, T. I., Vatsa, M., and Singh, R. (2011). On matching latent to latent fingerprints. In *2011 international joint conference on biometrics (IJCB)*, pages 1–6. IEEE.
- Sankaran, A., Vatsa, M., and Singh, R. (2014). Latent fingerprint matching: A survey. *IEEE Access*, 2:982–1004.
- Sankaran, A., Vatsa, M., and Singh, R. (2015b). Multisensor optical and latent fingerprint database. *IEEE access*, 3:653–665.
- Santurkar, S., Tsipras, D., Ilyas, A., and Madry, A. (2018). How does batch normalization help optimization? *arXiv preprint arXiv:1805.11604*.
- Singh, H. P. and Dimri, P. (2020). A survey of latent fingerprint indexing and segmentation based matching. In *Advances in Data and Information Sciences: Proceedings of ICDIS 2019*, pages 677–685. Springer.
- Singla, N., Kaur, M., and Sofat, S. (2020). Automated latent fingerprint identification system: A review. *Forensic science international*, 309:110187.

- Solomon, C. and Breckon, T. (2011). *Fundamentals of Digital Image Processing: A practical approach with examples in Matlab*. John Wiley & Sons.
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958.
- Svoboda, J., Monti, F., and Bronstein, M. M. (2017). Generative convolutional networks for latent fingerprint reconstruction. In *2017 IEEE International Joint Conference on Biometrics (IJCB)*, pages 429–436. IEEE.
- Tabassi, E., Wilson, C., and Schlenoff, C. (2004). Fingerprint image quality. <https://doi.org/10.6028/NIST.IR.7151>.
- Tang, Y., Gao, F., Feng, J., and Liu, Y. (2017). Fingernet: An unified deep network for fingerprint minutiae extraction. In *2017 IEEE International Joint Conference on Biometrics (IJCB)*, pages 108–116. IEEE.
- Tistarelli, M. and Champod, C. (2017). *Handbook of biometrics for forensic science*. Springer.
- Tschannen, M., Bachem, O., and Lucic, M. (2018). Recent advances in autoencoder-based representation learning. *arXiv preprint arXiv:1812.05069*.
- Ulery, B. T., Hicklin, R. A., Buscaglia, J., and Roberts, M. A. (2011). Accuracy and reliability of forensic latent fingerprint decisions. *Proceedings of the National Academy of Sciences*, 108(19):7733–7738.
- Ulery, B. T., Hicklin, R. A., Buscaglia, J., and Roberts, M. A. (2012). Repeatability and reproducibility of decisions by latent fingerprint examiners. *PloS one*, 7(3):e32800.
- Vasanth, K. et al. (2019). Latent fingerprint image enhancement using gabor functions via multi-scale patch based sparse representation and matching based on neural networks. In *2019 International Conference on Communication and Signal Processing (ICCSP)*, pages 0365–0369. IEEE.
- Wang, Y., Hu, J., and Phillips, D. (2007). A fingerprint orientation model based on 2d fourier expansion (fomfe) and its application to singular-point detection and fingerprint indexing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(4):573–585.
- Wang, Y., Yao, H., Zhao, S., and Zheng, Y. (2015). Dimensionality reduction strategy based on auto-encoder. In *Proceedings of the 7th International Conference on Internet Multimedia Computing and Service*, pages 1–4.
- Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612.

- Watson, C., Wilson, C., Indovina, M., Snelick, R., and Marshall, K. (2004). Studies of one-to-one matching with vendor sdk matchers. In *Technical Report, Technical Report NISTIR 7119*. U.S. Department of Commerce, National Institute of Standards and Technology.
- Wong, W. J. and Lai, S.-H. (2020). Multi-task cnn for restoring corrupted fingerprint images. *Pattern Recognition*, 101:107203.
- Xu, Y., Wang, Y., Liang, J., and Jiang, Y. (2020). Augmentation data synthesis via gans: Boosting latent fingerprint reconstruction. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2932–2936. IEEE.
- Yadav, O., Ghosal, K., Lutz, S., and Smolic, A. (2021). Frequency-domain loss function for deep exposure correction of dark images. *Signal, Image and Video Processing*, 15(8):1829–1836.
- Yegnanarayana, B. (2009). *Artificial neural networks*. PHI Learning Pvt. Ltd.
- Yin, W., Goldfarb, D., and Osher, S. (2007). The total variation regularized l1 model for multiscale decomposition. *Multiscale Modeling & Simulation*, 6(1):190–211.
- Zhang, J., Lai, R., and Kuo, C.-C. J. (2013). Adaptive directional total-variation model for latent fingerprint segmentation. *IEEE Transactions on Information Forensics and Security*, 8(8):1261–1273.
- Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., and Liang, J. (2018). Unet++: A nested u-net architecture for medical image segmentation. In *Deep learning in medical image analysis and multimodal learning for clinical decision support*, pages 3–11. Springer.
- Zhu, Y., Yin, X., and Hu, J. (2023). Fingergan: a constrained fingerprint generation scheme for latent fingerprint enhancement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Apêndice A Captura das Imagens de Impressões Digitais Geradas pelo SFinGe

A versão do programa SFinGe utilizada possui como limitação a geração de apenas uma única imagem de impressão digital por vez, sem possibilidade de salvar ou exportar a imagem gerada. Diante disso, para automatizar o processo de criação de uma base de dados, a captura das imagens foi realizada por meio da implementação de uma macro, programada utilizando a ferramenta Pullover's Macro Creator¹¹. A seguir, são descritas as etapas realizadas nesse processo, bem como são apresentadas as principais telas utilizadas para explorar as variações de opções disponíveis no SFinGe.

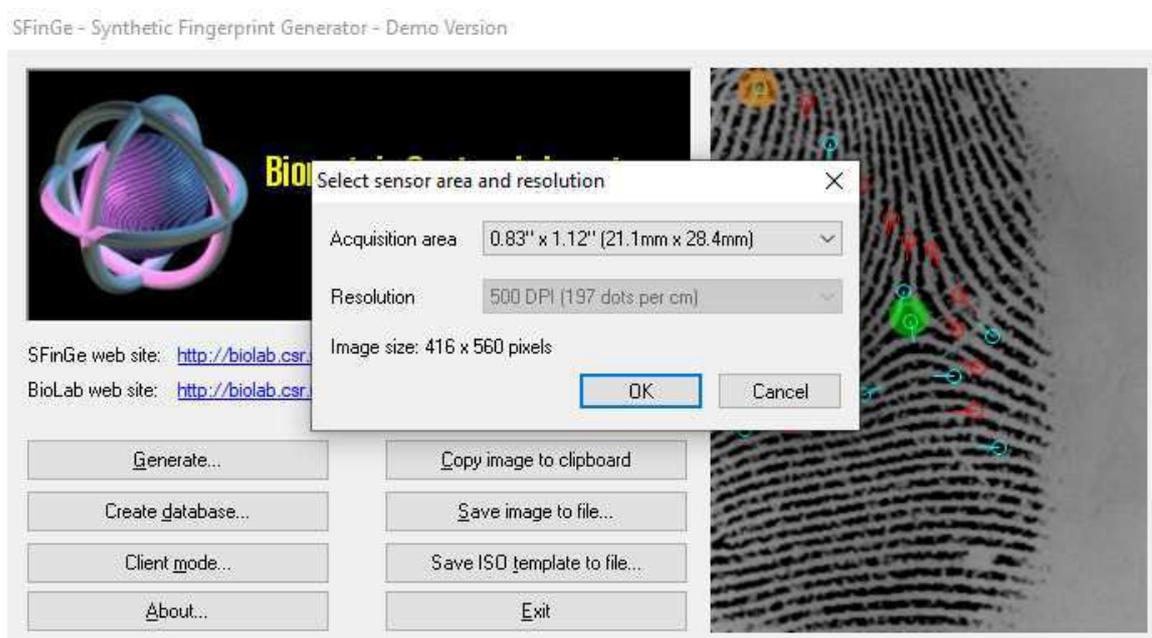
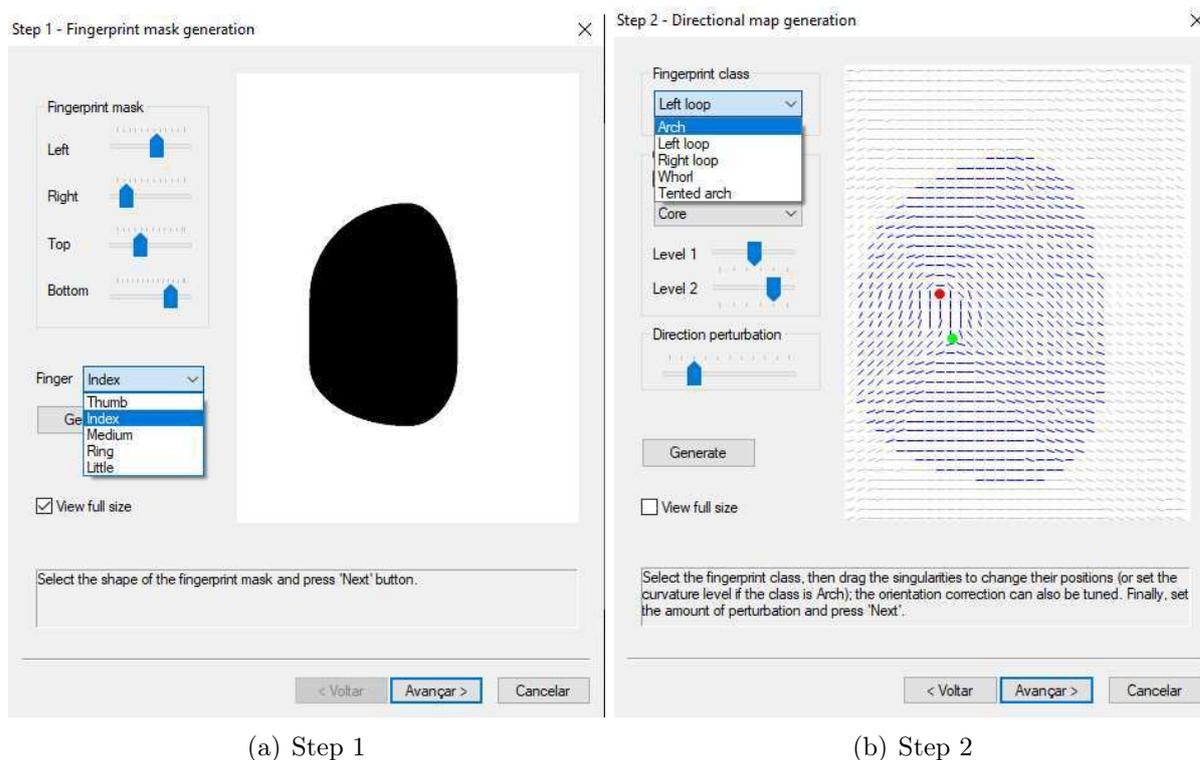


Figura 56: Tela inicial do SFinGe. Fonte: De autoria própria.

Na Figura 56 é mostrada a janela inicial do programa e ilustra a primeira etapa do procedimento de captura. Nesta etapa é verificado se a primeira janela do sistema foi inicializada e, caso positivo, seleciona-se a opção “Generate” e, em seguida, a opção “Ok” para avançar.

Na Figura 57(a) é exibida a janela *Step 1* do sistema. Nesta etapa, é feita a seleção referente ao dedo que a imagem da impressão digital gerada representará. A cada vez que um item da caixa de opções “Finger” é selecionada, é gerada uma nova máscara da impressão digital com posição e tamanho aleatórios. Este foi o recurso utilizado para evitar a necessidade de uma programação específica para essas opções. Na tela *Step 2*, mostrada na Figura 57(b), é feita a seleção da classe (arcos, *loops* ou verticilos) da impressão digital. A diversificação dessa opção é importante para que o banco de dados gerado possua amostras dos diferentes tipos de estruturas de cristas das impressões digitais. Cada seleção de um item da caixa de opções “Fingerprint class” resulta em uma nova estrutura gerada de forma aleatória.

¹¹<https://www.macrocreator.com/download/>



(a) Step 1

(b) Step 2

Figura 57: SFinGe, telas *Step 1* (a), em que é feita a escolha referente ao dedo que a imagem da impressão digital gerada representará, e *Step 2* (b), em que é realizada a escolha da classe da impressão digital. Fonte: De autoria própria.

O próximo passo, tela *Step 3*, mostrada na Figura 58, é a etapa em que é realizada a captura e modificação da posição da barra de rolagem para selecionar uma opção de densidade das cristas. Em seguida, seleciona-se o botão “Start ridge generation”, o qual produz a imagem da impressão digital binarizada, utilizada como *ground truth*. Após a definição das opções, é ativada a captura de tela do sistema operacional, onde é realizada a captura da região da imagem do *ground truth* da impressão digital, através do rastreamento do movimento do ponteiro do *mouse*. Um *script*, escrito em Python, externo ao programa gera e armazena um nome para o arquivo capturado que será salvo em disco.

Na janela seguinte, *Step 4*, mostrada na Figura 59(a), é possível selecionar, via barra de rolagem, a quantidade de arranhões e/ou cicatrizes na impressão digital. O posicionamento e a forma em que essas características aparecem são definidos de forma aleatória pelo próprio SFinGe. Na Figura 59(b) é exibida a tela em que é realizada uma escolha entre diferentes níveis de simulação de pressionamento do dedo em alguns pontos da impressão digital. É possível escolher variações entre níveis baixos de pressão, conforme mostrado na Figura 60(a), e também um nível mais alto de pressão em alguns pontos, Figura 60(b). Para a base de dados criada, foram trabalhadas de forma aleatória variações dessas opções por considerar que esta é uma característica muito presente nas impressões digitais latentes.

Na janela *Step 8*, exibida na Figura 61(a), o SFinGe converte a versão binarizada da imagem para a impressão digital sintética. Nesta etapa, são adicionadas diferentes combinações

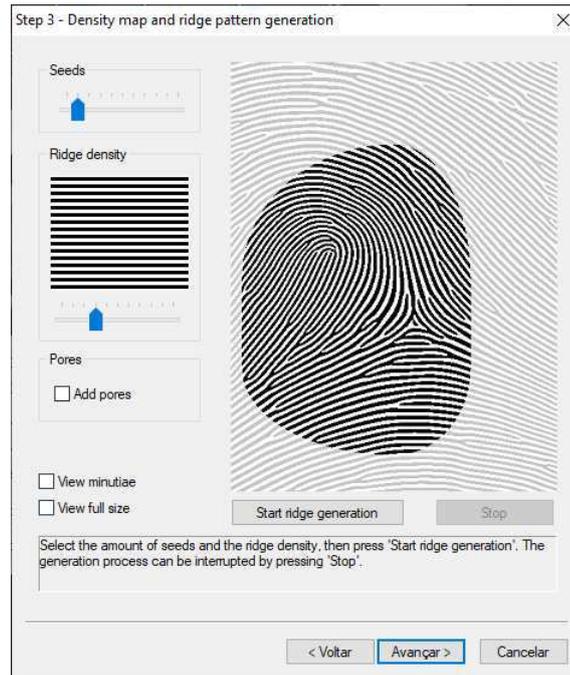
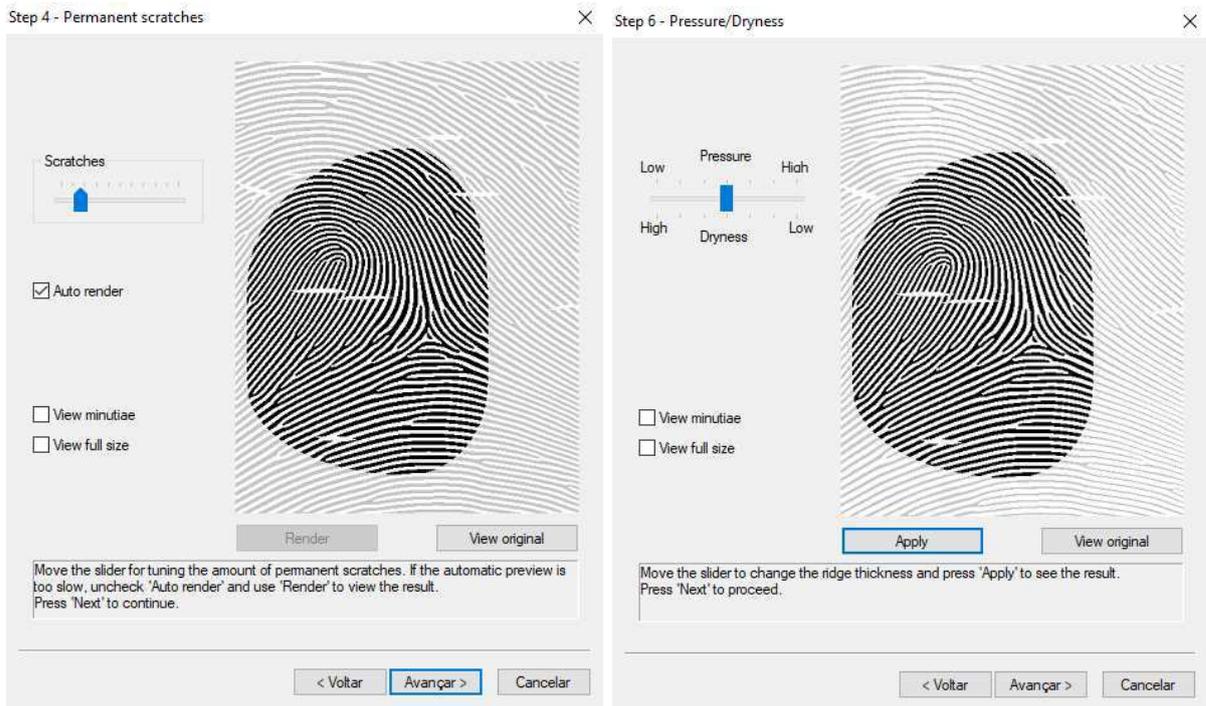


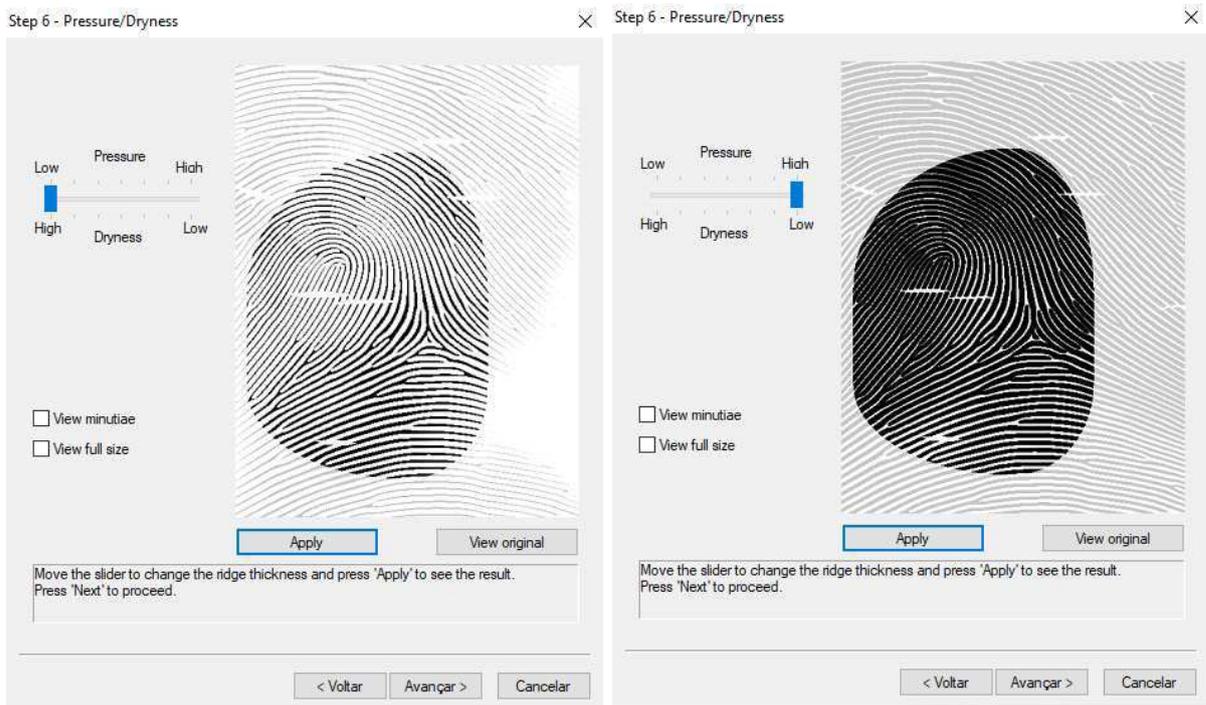
Figura 58: SFinGe, tela *Step 3*, onde é possível definir a densidade das cristas da impressão digital. Fonte: De autoria própria.



(a) Step 4

(b) Step 6

Figura 59: SFinGe, telas *Step 4* (a), em que é definida a existência de arranhões ou cicatrizes na impressão digital, e *Step 6* (b), onde pode ser definido se existem pontos de diferentes níveis de pressionamento do dedo. Fonte: De autoria própria.



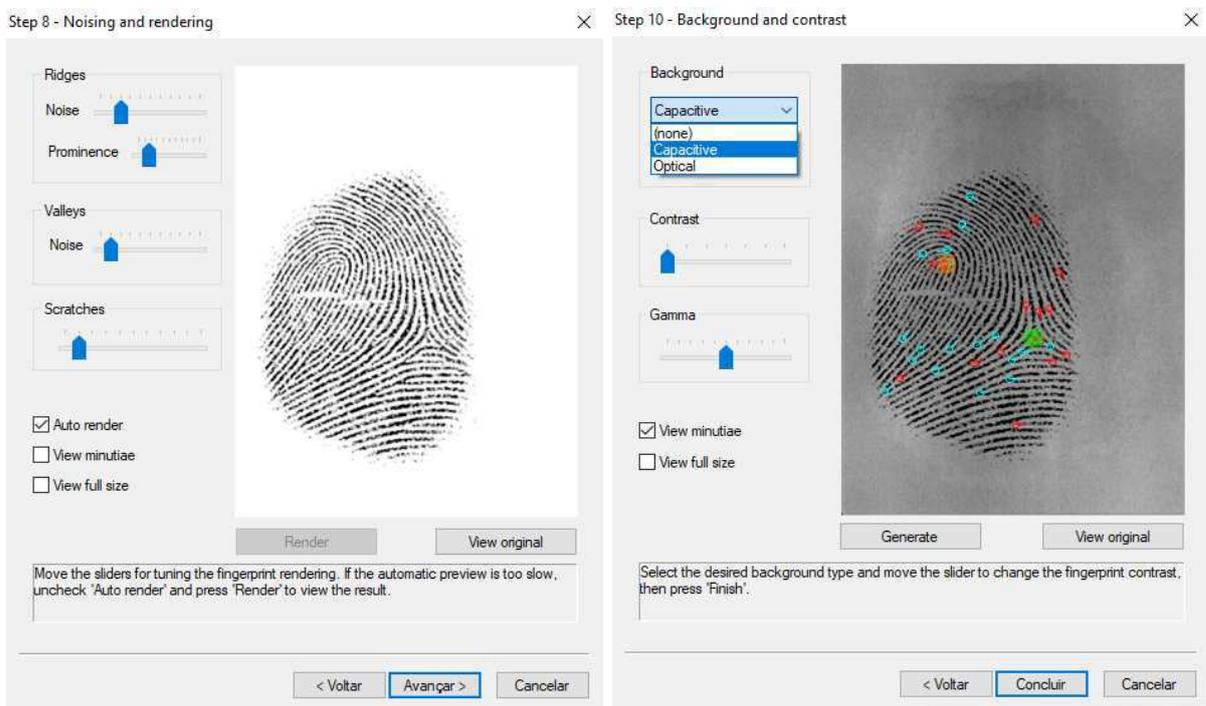
(a) Step 6 - Pontos com baixa pressão

(b) Step 6 - Pontos com alta pressão

Figura 60: SFInGe, telas *Step 6*. Opções de pressão do dedo em alguns pontos da impressão digital. Fonte: De autoria própria.

das opções de ruído nas cristas e vales da impressão digital. As simulações de ruídos realizadas pelo SFInGe estão associadas a ruídos que podem surgir durante a coleta das impressões digitais por meio de sensores, e não aos ruídos presentes nas impressões digitais latentes. A tela *Step 10* é a última e nela é realizada a seleção entre três possíveis fundos para a impressão digital: fundo totalmente branco, fundo simulando a captura por sensor capacitivo e fundo simulando a captura por sensor óptico.

Como última etapa, é realizado novamente o procedimento de execução do *script* externo para recuperar o nome criado na captura da imagem do arquivo do *ground truth*. É executada a captura de tela do sistema operacional, rastreando a posição do ponteiro do mouse e, após isso, feito o salvamento do arquivo em disco, finalizando assim a criação de um par formado pela impressão digital sintética e pelo *ground truth*.



(a) Step 8

(b) Step 10

Figura 61: SFinGe, telas *Step 8* (a), em que é possível selecionar diferentes níveis de ruídos nas cristas e nos vales da impressão digital, e *Step 10* (b), onde é realizada a escolha do tipo de plano de fundo da impressão digital. Fonte: De autoria própria.

Apêndice B Detalhes de Implementação da RM-net

O código da RM-net foi escrito na linguagem Python versão 3.7 utilizando o *framework* Keras¹² na versão v2.3.1 com o TensorFlow versão v2.4.1 como *backend*. Para acompanhamento e registro das informações da evolução do treinamento, a ferramenta Tensorboard, na versão v2.1.0, foi utilizada. Os experimentos foram conduzidos no Google Colab Pro¹³ em instâncias com 25 GB de memória RAM e, na maioria das vezes, em uma GPU NVIDIA[®] Tesla P100 com 16 GB de memória HBM2.

A plataforma Google Colab Pro possui características que devem ser consideradas. Recentemente, foi adotado um modelo de pontos computacionais para os usuários, que oferece 100 pontos por mês na assinatura Pro. A quantidade de pontos utilizada por hora varia de acordo com a GPU selecionada para o processamento. O tempo máximo de execução contínua é de 24 horas. A memória RAM disponível é de 25 GB e o armazenamento temporário de arquivos é de 100 GB.

¹²<https://keras.io>

¹³<https://colab.research.google.com/>

Apêndice C Experimentos Preliminares Auxiliares

Este apêndice apresenta os resultados de experimentos preliminares realizados para auxiliar na definição da quantidade de blocos residuais que foram adicionados à arquitetura da RM-net desenvolvida nesta tese. Também são descritos o *setup* de configuração, os dados e ferramentas utilizados nestes experimentos.

Para todos os testes, foram utilizadas uma base de dados de treinamento de 23 mil imagens e as ferramentas do NBIS, com o algoritmo MINDTCT para extração e o Bozorth para correspondência. Além disso, são apresentados resultados que guiaram a escolha de utilizar a combinação de MAE, DFT e DCT como função de custo.

Devido às limitações de hardware e do ambiente de testes (Google Colab), esses experimentos foram usados para acelerar as tomadas de decisão. Isso permitiu a avaliação de diferentes configurações de rede e funções de custo em um tempo menor do que o que seria necessário para o treinamento da rede utilizando todos os dados, enquanto se avaliava cada modificação de configuração. As avaliações foram realizadas com base no casamento dos realces com as imagens da base MOLF DB1.

C.1 Experimentos para definição dos blocos residuais

Em uma abordagem experimental para otimizar a arquitetura da rede M-net, inicialmente, testes foram conduzidos utilizando a M-net sem a presença de blocos residuais. A fim de determinar o impacto desses blocos na eficácia do sistema, blocos residuais foram adicionados gradualmente em diferentes configurações, ou seja, 3, 6 e 9 blocos residuais. Os resultados obtidos dessas avaliações, exibidos na Tabela 5, indicam que a adição de blocos residuais resulta em melhorias progressivas no desempenho da rede.

Tabela 5: Acurácias dos experimentos conduzidos considerando a adição de blocos residuais

<u>Blocos Resnet</u>	<u>Rank-50</u>
0	27,23%
3	30,23%
6	32,52%
9	33,16%

Observando os resultados mostrados na Tabela, ao aumentar o número de blocos residuais de 3 para 6, houve um aumento de aproximadamente 2,29 pontos percentuais, considerando o *rank-50*. No entanto, ao estender a configuração para 9 blocos residuais, a melhoria no desempenho foi mais modesta, com um aumento de cerca de 0,64 pontos percentuais. Esse fato sugere que a inclusão de blocos residuais oferece benefícios, dentro do contexto de realce de impressões digitais latentes, particularmente em configurações iniciais, mas o ganho de desempenho tende a diminuir à medida que mais blocos residuais são adicionados. Devido a essa tendência de diminuição nos ganhos de desempenho, juntamente com o aumento no tempo de treinamento que

acompanha o aumento da profundidade da rede e as limitações de hardware, foi decidido manter a configuração de 9 blocos residuais para a rede.

C.2 Experimentos para definição da função de custo

Para determinar a função de custo que foi aplicada na rede RM-net, foram conduzidos experimentos que envolveram uma análise explorando diferentes combinações de funções de custo com foco nas suas contribuições de melhorias nos testes de correspondência. Foram investigadas quatro abordagens de custo: “MAE”, que considera apenas a diferença absoluta; “MAE & FFT” e “MAE & DCT”, que incorporam informações de frequência usando a Transformada Discreta de Fourier (DFT) e a Transformada Cosseno Discreta (DCT); e “MAE & DFT & DCT”, que combina ambas as informações de frequência. Os resultados do experimento são mostrados na Tabela 6.

Tabela 6: Acurácias dos experimentos conduzidos considerando diferentes combinações das funções de custo

Rank	Função de custo			
	MAE	MAE & DFT	MAE & DCT	MAE & DFT & DCT
1	21,09	22,89	19,66	22,70
10	29,25	31,07	26,79	31,14
20	32,25	34,14	30,00	34,59
25	34,02	35,36	31,34	35,82
30	35,02	36,11	32,29	36,82
40	36,57	37,86	33,93	38,87
50	38,20	39,57	35,61	40,48

Com base nos resultados apresentados na tabela, observa-se que a inclusão de informações de frequência, representadas pelas informações presentes nas colunas “MAE & FFT” e “MAE & DCT”, geralmente resulta em um desempenho superior em comparação com os valores obtidos pelo experimento representado na coluna “MAE” que não incorpora essas informações. Notavelmente, a métrica “MAE & FFT & DCT” demonstra o melhor desempenho na maioria dos casos, destacando que a combinação da MAE com as funções de custo no domínio da frequência contribuiu para melhorar os resultados dos testes de correspondência das impressões digitais latentes realçadas. Com base nesses resultados foi decidido utilizar a combinação das três funções de custo como função de custo final da RM-net.