AUTOENCODER VARIACIONAL E REDES ADVERSÁRIAS GENERATIVAS PARA O RECONHECIMENTO DE FACES A PARTIR DE UMA ÚNICA IMAGEM POR PESSOA.

Adson Diego Dionisio da Silva

Tese de Doutorado apresentado ao Programa de Pós-graduação em Engenharia de Sistemas e Computação, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Doutor em Engenharia de Sistemas e Computação.

Orientador: Ricardo Cordeiro de Farias

Rio de Janeiro

Julho de 2025

AUTOENCODER VARIACIONAL E REDES ADVERSÁRIAS GENERATIVAS PARA O RECONHECIMENTO DE FACES A PARTIR DE UMA ÚNICA IMAGEM POR PESSOA.

Adson Diego Dionisio da Silva

TESE SUBMETIDO AO CORPO DOCENTE DO INSTITUTO ALBERTO LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA (COPPE) DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE DOUTOR EM CIÊNCIAS EM ENGENHARIA DE SISTEMAS E COMPUTAÇÃO.

Examinado por:

Prof. Ricardo Cordeiro de Farias, PhD

Prof. Luiz Arthur Silva de Faria, PhD

Prof. Claudio Micelle de Farias, PhD

Prof. Wilfrido Gómez Flores, PhD

Prof. Wagner Coelho de Albuquerque Pereira, PhD

Prof. Anselmo Antunes Montenegro, PhD

RIO DE JANEIRO, RJ – BRASIL

JULHO DE 2025

Resumo do Tese apresentado à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Doutor em Ciências (D.Sc.)

AUTOENCODER VARIACIONAL E REDES ADVERSÁRIAS GENERATIVAS PARA O RECONHECIMENTO DE FACES A PARTIR DE UMA ÚNICA IMAGEM POR PESSOA.

Adson Diego Dionisio da Silva

Julho/2025

Orientador: Ricardo Cordeiro de Farias

Programa: Engenharia de Sistemas e Computação

Aplicações reais de sistemas de reconhecimento facial, principalmente na área de segurança, frequentemente dispõem de apenas uma imagem por indivíduo para identificação em ambientes não controlados, onde fatores como iluminação, pose, expressão e oclusão variam significativamente. Esse cenário, conhecido como "amostra única por pessoa" (SSPP - Single Sample Per Person), ainda representa um desafio significativo para os métodos atuais de reconhecimento facial. Com o intuito de abordar esse problema, propõe-se o arcabouço AD-VAE (*Adversarial Disentangling Variational Autoencoder*) , que combina técnicas de *autoencoder*s variacionais (VAE - *Variational Autoencoder*) com redes adversariais generativas (GAN - Generative Adversarial Network) para a geração de protótipos faciais que preservam a identidade do indivíduo. A arquitetura AD-VAE é composta por quatro redes neurais: codificador, decodificador, gerador e discriminador multitarefa. O método proposto foi avaliado em bases de dados amplamente utilizadas na literatura, incluindo AR, E-YaleB, CAS-PEAL, FERET e LFW, obtendo resultados superiores aos métodos do estado da arte, com taxas de reconhecimento que variam de 84,9% a 99,6%. Os resultados obtidos demonstram a eficácia e robustez da abordagem, bem como seu potencial para aplicações práticas e futuras pesquisas na área.

Abstract of Thesis presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Doctor of Science (D.Sc.)

VARIATIONAL AUTOENCODER AND GENERATIVE ADVERSARIAL NETWORKS FOR FACE RECOGNITION FROM A SINGLE SAMPLE PER PERSON.

Adson Diego Dionisio da Silva

July/2025

Advisor: Ricardo Cordeiro de Farias

Department: Systems Engineering and Computer Science

Real-world facial recognition systems, especially in the security domain, often rely on only a single image per individual for identification in uncontrolled environments, where factors such as lighting, pose, expression, and occlusion vary significantly. This scenario, known as Single Sample Per Person (SSPP), still poses a significant challenge for current facial recognition methods. To address this problem, we propose the AD-VAE (Adversarial Disentangling Variational Autoencoder) framework, which combines Variational Autoencoder (VAE) techniques with Generative Adversarial Networks (GAN) to generate identity-preserving facial prototypes. The AD-VAE architecture is composed of four neural networks: encoder, decoder, generator, and a multitask discriminator. The proposed method was evaluated on widely used benchmark datasets, including AR, E-YaleB, CAS-PEAL, FERET, and LFW, achieving superior performance compared to state-of-the-art methods, with recognition rates ranging from 84.9% to 99.6%. The obtained results demonstrate the effectiveness and robustness of the approach, as well as its potential for practical applications and future research in the field.

Sumário

Lista de Abreviaturas

AAM *Active Appearance Model*.

ACF *Adaptive Convolution Feature*.

ADA *Adaptive Discriminant Analysis*.

AD-VAE *Adversarial Disentangled Variational Autoencoder*.

AQI *Adaptive Quotient Image*.

AVAE *Adversarial Variational Autoencoder*.

BoW *Bag-of-Word*.

CAPES Coordenação de Aperfeiçoamento de Pessoal de Nível Superior.

CCM-CNN *Cross-Correlation Matching Convolutional Neural Network*.

CNH Carteira Nacional de Habilitação.

CNN *Convolutional Neural Network*.

CoMax-KCFA *Coupled Max-Pooling Kernel Class-Dependence Feature Analysis*.

CRC *Collaborative Representation Classification*.

DA *Domain Adaptation*.

DMMA *Discriminative Multimanifold Analysis*.

DpLSA *Discriminat Probabilistic Latent Semantic Analysis*.

DR-GAN *Disentangled Representation learning-Generative Adversarial Network*.

DTL *Discriminative Transfer Learning*.

DTLSR *Discriminative Transfer Learning with Sparsity Reguarization*.

ELBO *Evidence Lower Bound*.

FA *Factor Analysis*.

FR *Face Recognition*.

GAN *Generative Adversarial Network*.

GEM *Generic Elastic Models*.

HOG *Histogram of Oriented Gradients*.

IEEE Instituto de Engenheiros Eletricistas e Eletrônicos.

KDA *Kernel Discriminant Analysis*.

KNN *K-Nearest Neighbors*.

LBP *Local Binary Patterns*.

LDA *Linear Discriminant Analysis*.

LFW Labeled Faces in the Wild.

LGM *Logarithm Gradient Magnitude*.

LGO *Logarithm Gradient Orientation*.

PCA *Principal Component Analysis*.

pLSA *Probabilistic Latent Semantic Analysis*.

RDA *Regularized Discriminant Analysis*.

RF Reconhecimento de Faces.

RG Registro Geral.

SDAE *Stacked Denoising Auto-Encoder*.

SDM *Supervised Descent Method*.

SGD *Stochastic Gradient Descent*.

SiBI Sistema de Bibliotecas e Informação.

SIFT *Scale-Invariant Feature Transform*.

SRC *Sparse Representation Classifier* .

SRF Sistema de Reconhecimento de Faces.

SSFR *Single Sample Face Recognition*.

SSPP *Single Sample Per Person*.

SSRC *Superposed Sparse Representation based Classification*.

SVM *Support Vector Machine*.

TER *Total Error Rate*.

TIC Superintendência de Tecnologia da Informação e Comunicação.

UFRJ Universidade Federal do Rio de Janeiro.

VAE *Variational Autoencoder*.

VD-GAN *Variation Disentangled Generative Adversarial Autoencoder Network*.

WPCA *Whitened Principal Component Analysis*.

Lista de Figuras

Lista de Tabelas

Capítulo 1

Introdução

Ativo desde 1960 [1], o campo de reconhecimento automático de faces tem apresentado um crescimento vertiginoso nos últimos anos. O sucesso de aplicações de análise de vídeo tem trazido muita atenção ao tema, evidenciado pelo surgimento de conferências especializadas e protocolos padronizados de avaliação, bem como na aplicação das diversas técnicas no mundo tecnológico contemporâneo (i.e. reconhecimento de face em celulares, sistemas de segurança, etc.). O crescimento é guiado pelos avanços de Sistema de Reconhecimento de Faces (SRF) e das novas tecnologias para processamento de informações durante os últimos anos, o que desperta o interesse dos pesquisadores de processamento de imagem, redes neurais, visão computacional e computação gráfica [2].

A principal vantagem do reconhecimento de faces sobre os outros métodos de identificação biométrica é a falta da necessidade de cooperação do usuário. Por esse motivo, ela se torna o mais acurado e um dos métodos biométricos menos intrusivos [3], podendo ser utilizado em massa para identificação em multidões [4], tornando-se um dos métodos biométricos mais populares, com uma ampla gama de aplicações [5].

Segundo TAN *et al.* [3], SRF tem recebido muita atenção nos últimos 20 anos, tanto pela academia quanto pela indústria, tendo como objetivo identificar ou verificar uma ou mais pessoas em uma imagem ou vídeo. Muitos trabalhos têm como foco aumentar a acurácia do reconhecimento, porém, ignoram o problema de quando se tem apenas uma imagem por pessoa, condições em que muitas técnicas encontradas na literatura falham. Esse problema é conhecido como "amostra única por

pessoa" ou em inglês *Single Sample Per Person* (SSPP) e *Single Sample Face Recognition* (SSFR), que é descrito pelos autores como: a partir de uma base de dados que contém apenas uma única imagem por pessoa, reconhecer se a pessoa presente na imagem de consulta está entre a relação de pessoas presentes na base de dados, considerando que as etapas de detecção e segmentação da face já foram previamente realizadas.

Este cenário é encontrado usualmente em identificação de pessoas por meio de documentos, tais como Registro Geral (RG), Carteira Nacional de Habilitação (CNH) e passaporte. Além dos problemas citados, os SSFR também enfrentam os problemas inerentes ao reconhecimento de faces, como o chamado "ambiente não controlado" ou em inglês "in the wild", em que não se tem controle sobre a pose, expressão, oclusão e iluminação das imagens que precisam ser reconhecidas ou verificadas. Dessa forma, o maior desafio do SSPP está em obter o maior número possível de informações para inferir as variações encontradas no mundo real [6].

Os primeiros métodos para Reconhecimento de Faces (RF) tinham como base a geometria, que utilizava a imagem para extrair métricas e formar um modelo para o reconhecimento.Com o surgimento de bases de dados mais abrangentes, os métodos geométricos tornaram-se obsoletos, sendo substituídos por abordagens baseadas na aparência. Técnicas que utilizavam ferramentas inteligentes de diversas disciplinas, como estatística, inteligência artificial e reconhecimento de padrões [2]. Porém, boa parte dessas técnicas têm como base a aprendizagem, o que faz com que elas necessitem de um grande volume de dados para aprender as variações que uma face pode apresentar, o que não era possível em SRF que trabalhavam com passaporte ou CNH. Algumas dessas limitações em volumes de dados foram abordadas por meio de características mais robustas e discriminantes, novas técnicas de sintetização de imagens e de aprendizado de variações dentro da mesma face por meio de bases de dados auxiliares (base de dados genérica).

Inspirados pelo trabalho de TRAN *et al.* [7], que explorou a aprendizagem de representações desentrelaçadas para reconstrução de imagens faciais com diversas variações, e considerando o *Variation Disentangled Generative Adversarial Autoencoder Network* (VD-GAN) de PANG *et al.* [8], que adaptou essa base para o problema de SSPP *Face Recognition* (FR), além do *Adversarial Variational Autoencoder* (AVAE)

de PLUMERAULT *et al.* [9], que empregou VAE para a síntese de imagens, surge a hipótese de pesquisa: a combinação de um codificador de Variational Autoencoder (que utiliza distribuições de probabilidade para geração de imagens) com uma *Generative Adversarial Network* (GAN) que incorpora o desentrelaçamento de características (similar ao VD-GAN) pode gerar imagens de protótipos em condições neutras (pose frontal, expressão séria, iluminação regular e ausência de oclusões) para o reconhecimento facial com uma única amostra por pessoa

Para mitigar essas limitações, em consonância com a hipótese de pesquisa levantada, propomos uma nova arquitetura chamada AD-VAE, que combina os pontos fortes do VAE para representar identidades de forma *disentangled* e do GAN com o objetivo de realizar o reconhecimento de faces a partir de uma única imagem por pessoa, mesmo em cenários não controlados. A arquitetura é composta por quatro redes neurais profundas: um codificador (encoder), um decodificador (decoder), um gerador (generator) e um discriminador multitarefa (discriminator).

O diferencial do AD-VAE está na divisão do treinamento em duas fases: a primeira, baseada no VAE, permite a extração de um vetor latente que representa de forma *disentangled* (desembaraçada) os traços de identidade do indivíduo; e a segunda, baseada em GAN, permite a geração de uma nova imagem protótipo a partir do vetor latente e de um vetor de ruído, buscando preservar a identidade da imagem original e eliminar as variações de iluminação, pose e expressão.

Dessa forma, o método proposto consegue gerar protótipos de alta qualidade e robustez, mesmo quando treinado com apenas uma imagem por indivíduo, obtendo resultados superiores a métodos previamente propostos na literatura, inclusive em bases de dados desafiadoras como a LFW.

1.1 Motivação

Segundo OH *et al.* [10], durante as últimas décadas, o reconhecimento de faces tem sido um tópico ativo em reconhecimento de padrões e visão computacional, devido ao grande potencial para aplicações forenses, aplicação da lei, segurança e entretenimento. Essas tecnologias avançaram significativamente, porém ainda continua um desafio desenvolver SRF para serem usados em ambientes não controlados, que

são ambientes em que não se tem controle sobre a posição, expressão, iluminação ou oclusão das faces. Nas aplicações como identificação de pessoas em aeroportos, listas de procurados, vídeos de segurança, entre outras, o desafio é ainda maior quando apenas uma imagem por pessoa está disponível para o treinamento do SRF.

Também temos o problema de que criar bases de dados com várias imagens de um mesmo indivíduo em aplicações reais (e.g. e-passaporte, identificação de motorista no trânsito, aplicação da lei, reconhecimento de foragido, reconhecimento de desaparecidos, etc.) é extremamente difícil, o que torna necessária a criação e aperfeiçoamento de sistemas que consigam abordar o problema.

Com base nesse cenário, esta pesquisa propõe o desenvolvimento de um novo arcabouço denominado AD-VAE, que visa resolver o problema do reconhecimento facial com uma única imagem por indivíduo, mesmo em ambientes não controlados. A proposta se justifica pela necessidade de superar as limitações impostas pela ausência de múltiplas imagens por pessoa, comum em bases como registros civis, documentos oficiais e dados coletados em operações de segurança.

Diante disso, os sistemas para SSPP têm como vantagens [3]:

- Simplicidade de montar o banco de dados, uma vez que é necessário apenas uma imagem por indivíduo.
- Baixo custo de armazenamento, visto que é necessário apenas uma imagem por indivíduo.
- Baixo custo computacional para o treinamento e processamento da base de dados.

1.2 Objetivos

O objetivo principal do trabalho proposto é desenvolver um método capaz de realizar o reconhecimento de faces em ambientes não controlados, a partir de apenas uma única imagem da face por pessoa, preferencialmente em condições próximas à neutralidade (pose frontal, iluminação regular, expressão neutra e ausência de oclusões). A abordagem desconsidera as etapas de detecção e segmentação facial, assumindo que a imagem de entrada já contém a face previamente recortada.

1.3 Objetivos específicos

O trabalho tem os seguintes objetivos específicos utilizados para alcançar o objetivo geral:

- Utilizando de métodos avançados de sintetização de imagens, conseguir sintetizar uma imagem, chamada de protótipo, que preserve a identidade do indivíduo e que seja o mais próximo possível da condição de neutralidade. De forma que a partir do protótipo gerado possa conseguir reconhecer o individuo em uma imagem de consulta com pose, iluminação, oclusão e expressão diferentes.

- Empregar redes neurais com transferência de aprendizado, treinadas em bases genéricas auxiliares (que não contêm os indivíduos a serem reconhecidos), para extrair características discriminativas e gerar protótipos que preservem a identidade na galeria, utilizando VAE e GAN.

- Propor, implementar e validar uma arquitetura baseada em aprendizado profundo denominada AD-VAE, que combina os princípios de representação latente disentangled do VAE com a capacidade de síntese realista do GAN, gerando protótipos identitários robustos a variações de pose, iluminação, expressão e oclusão.

1.4 Contribuições

O trabalho realizado tem como principais contribuições:

- Desenvolvimento de um arcabouço capaz de realizar o reconhecimento de faces com apenas uma amostra de face por pessoa, e que seja robusto às variações na pose, iluminação, expressão e oclusão.

- Desenvolver um método que combine da representação desembaraçada do VAE com a capacidade de sintetização das GAN, de forma a gerar protótipos em condições próximas da neutralidade que preserve a identidade dos indivíduos com a robustez às variações na face das técnicas de transferência de aprendizado.

- Até onde temos conhecimento, este é o primeiro trabalho a integrar as abordagens VAE e GAN para a tarefa de SSFR

1.5 Organização do Documento

No segundo capítulo, *Revisão Bibliográfica*, são apresentadas as pesquisas de Qualis A1 e A2 sobre o reconhecimento de faces com apenas uma imagem por pessoa, sendo apresentadas as pesquisas mais relevantes sobre o tema proposto.

No terceiro capítulo, *Metodologia*, são apresentadas a metodologia utilizada para a realização do trabalho, assim como os detalhes de implementação do método proposto.

No quarto capítulo, *Resultados e Discussões* será apresentado o detalhamento dos experimentos, das bases de dados e protocolos utilizados, dos resultados obtidos e da comparação com os métodos existentes na literatura.

No quinto capítulo, *Conclusões*, são apresentadas as considerações finais sobre o trabalho realizado, suas principais contribuições, limitações identificadas e possíveis direções para pesquisas futuras.

Capítulo 2

Revisão Bibliográfica

Neste capítulo, são apresentados os trabalhos relacionados ao tema proposto, os quais abrangem as publicações relevantes sobre o tema nos últimos 7 anos. Publicações obtidas por meio de critérios que são definidos ao decorrer desse capítulo. O objetivo da pesquisa é definir as principais técnicas e abordagens utilizadas para detecção de faces com apenas uma imagem de referência por pessoa.

2.1 Critérios de pesquisa

Inicialmente, foram utilizadas as bibliotecas digitais de informação científicas disponibilizadas pelo o Acesso Remoto Integrado da Universidade Federal do Rio de Janeiro (UFRJ), que é uma parceria entre o Sistema de Bibliotecas e Informação (SiBI) e a Superintendência de Tecnologia da Informação e Comunicação (TIC), cujo objetivo é prover acesso em tempo integral à informação científica e tecnológica internacional a toda comunidade com vínculo ativo na UFRJ. Assim, é possível acessar: Livros eletrônicos Cambridge; Livros eletrônicos Atheneu; Livros eletrônicos Ebrary; Livros eletrônicos Wiley; Portal de Periódicos da CAPES; Livros eletrônicos do Instituto de Engenheiros Eletricistas e Eletrônicos (IEEE); Livros eletrônicos Taylor & Francis; Livros eletrônicos Springer; Periódicos eletrônicos Duke University Press e Periódicos da Royal Society. Em adicional, foram utilizados os sistemas de busca de material acadêmico Google Acadêmico, Science Direct e Microsoft Academic.

No momento da Pesquisa os livros eletrônicos do Atheneu não estava acessível (Falha técnica) e para os demais fora utilizado inicialmente a chave de pesquisa "*Sin-*

gle Sample Face Recognition". Porém, os mecanismos de busca do Ebrary, Wiley, Taylor & Francys, Duke University Press e o Royal Society não retornaram resultados relevantes (que abordasse o tema proposto). Os resultados iniciais do Cambridge também não continham conteúdo relevante ao tema proposto e não foi encontrado um mecanismo que disponibilizasse uma filtragem que possibilitasse retornar apenas documentos que abordasse o tema.

Dos resultados obtidos, foram considerado apenas os que foram publicados nos últimos 7 anos, com Qualis [1] (Sistema brasileiro de avaliação de periódicos mantido pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)) A1 ou A2, utilizando imagens 2D e que citassem a chave de pesquisa no título ou no resumo. A Tabela 2.1 mostra os resultados utilizados na revisão, levando em consideração que os resultados finais não consideram as repetições de artigos. A Tabela 2.2 mostra a distribuição dos artigos por ano e por Qualis.

Tabela 2.1: Resultados dos mecanismos de busca.

Mecanismo de Busca	Resultados	Selecionados
Microsoft Academic	196	39
Periódicos da CAPES	199542	1
IEEE	355	9
Springer	119368	0
Science Direct	141028	6
Google Acadêmico	25800	3
Total		58

Fonte: Própria.

[1] acessado em: http://qualis.ic.ufmt.br/ e https://sucupira.capes.gov.br/sucupira/

Tabela 2.2: Resultados dos mecanismos de busca.

Ano	A1	A2
2023	2	0
2021	1	0
2020	4	2
2019	2	1
2018	3	3
2017	15	1
2016	6	8
2015	6	0
2014	5	0
2013	10	1
Total	47	12

Fonte: Própria.

O *Microsoft Academic* foi o primeiro mecanismo de busca a ter os resultados analisados, sendo também o que teve a melhor relação entre a quantidade de resultados após os filtros e resultados selecionados. Os resultados dos outros mecanismos de busca não consideram as repetições, sendo assim, os números refletem os artigos encontrados que não foram retornados pelos outros mecanismos de busca. Dentre os artigos selecionados pode-se observar que a maioria tem Qualis A1, sendo o resultado mais alto atribuído pela CAPES.

2.2 Técnicas utilizadas

Durante as últimas décadas vários métodos foram propostos, e junto com eles algumas formas de classificações para as técnicas de SSFR, entre elas: (1) a classificação

em métodos baseado em geometria ou aparência [6, 11]; (2) a classificação de em holístico, características locais ou hibrido [2, 3, 12–14]; (3) a classificação baseada em imagem sintéticas, em aprendizagem de espaços multi-visão e na extração de características robustas a pose [10]; (4) classificação em métodos com ou sem base de dados auxiliar, nos quais as técnicas sem base auxiliar pode ser sub-classificado em métodos com base em características robustas, em sintetização de imagens e na divisão em blocos locais[15]; (5) classificação de CHU *et al.* [4], que classifica em métodos supervisionado, não-supervisionados e semi-supervisionados, em que o supervisionado pode ser subdividido em técnicas baseadas em amostragem virtual, em base de dados auxiliar genérica e no particionamento em blocos.

Observando os métodos revisados e as classificações propostos, apenas as duas primeiras classificações são realmente abrangentes, pois muitas das técnicas revisadas se enquadram em mais de uma das classificações propostas. Sendo assim, os métodos pode ser classificados com base em geometria (métodos mais antigos e não encontrado na revisão) e métodos com base em aparência (todos os artigos revisados). Dentre as técnicas com base em aparência, podem-se dividir em: (1) Holísticos, em que é utilizado a face inteira como entrada do método, utilizando de características e relacionamentos globais; (2) locais, em que a face é particionada em blocos, sendo esses com sobreposição ou não, assim utilizando características e relacionamentos locais; (3) Híbridos, que é composto da junção dos métodos anteriormente citados. As demais classificações encontradas na revisão não são bem delimitadas, com por exemplo a classificação de KAN *et al.* [11] que divide em: (1) métodos que utilizam apenas a imagens da base de dados com uma única imagem por pessoa (também chamada de galeria); (2) métodos que utiliza a galeria para sintetizar imagem; (3) métodos que utilizam de base de dados auxiliar para complementar a falta de informação da galeria.

Trabalhos como o de DENG *et al.* [16] utilizam tanto base auxiliar, quanto sintetização de imagens, não se limitando em apenas uma das classes de KAN *et al.* [11]. Dentre os artigos revisados também é possível observar que muitos utilizam uma base auxiliar para complementar a falta de informação devido se ter apenas uma única imagem por pessoa, como mostrado na Tabela 2.3. Outra observação é que muitos trabalhos utilizam da geração de imagens para aumentar a base de dados

de treinamento, assim como o particionamento das imagens em blocos, que também gera um aumento da base de dados de treinamento, utilizando os blocos particionados como a galeria. Também é percebido que há equilíbrio entre a quantidade de métodos holísticos e locais, assim como alguns com abordagens hibridas.

Tabela 2.3: Métodos com e sem base auxiliar (base genérica).

Ano	Com base auxiliar	Sem base auxiliar
2023	PANG *et al.* [17]	-
2021	PANG *et al.* [8]	-
2020	PLUMERAULT *et al.* [9], YANG *et al.* [18], KARRAS *et al.* [19], LEE *et al.* [20]	-
2019	TRAN *et al.* [7], PANG *et al.* [21], ZHAO *et al.* [22]	-
2018	LIU e WASSELL [15], DENG *et al.* [16]	CHU *et al.* [4], OH *et al.* [10], ZHOU *et al.* [23], ZHANG *et al.* [24]
2017	YANG *et al.* [12], YU *et al.* [25], PARCHAMI *et al.* [26], LIONG e HAIBIN YAN [27], JI *et al.* [28], HUANG *et al.* [29], HU [30], GAO *et al.* [31], DENG *et al.* [32]	PANG *et al.* [13], ZHU *et al.* [33], PEI *et al.* [34], HU *et al.* [35], HONG *et al.* [36], GUO *et al.* [37], CHU *et al.* [38]
2016	HAGHIGHAT *et al.* [39], GU *et al.* [40]	SONG *et al.* [14], LIU *et al.* [41, 42]
2015	ZHUANG *et al.* [43], HU *et al.* [44], GAO *et al.* [45, 46], DING *et al.* [47], CAI *et al.* [48], BASHBAGHI *et al.* [49]	HU *et al.* [50], JUEFEI-XU *et al.* [51], MACHADO [52], ZHAO *et al.* [53]
2014	YIN *et al.* [54], DENG *et al.* [55]	YAN *et al.* [56], LIU *et al.* [57], BORGI *et al.* [58]
2013	KAN *et al.* [11], YANG *et al.* [59], KVETON e VALKO [60], DENG *et al.* [61]	YING LI *et al.* [6], LU *et al.* [62, 63], ZHU *et al.* [64]

Outro ponto relevante é a predominância dos métodos que têm como base o *Sparse Representation Classifier* (SRC), seguido por redes neurais e *Linear Discriminant Analysis* (LDA) e, sendo as técnicas inspiradas no primeiro mais que o dobro da quantidade que as inspiradas no segundo. O método de classificação mais utilizando é a comparação entre o erro de reconstrução das imagens de teste por meio de combinações lineares da base de treinamento, normalmente chamado de galeria. Em Segundo lugar vem a classificação por vizinhança, em que a imagem de teste é projetada em um subespaço e a classificação é dada pelo vizinho mais próximo da projeção da imagem de teste.

Tabela 2.4: Relação entre bases de dados e artigos revisados.

Base de dados	Quantidade
AR	31
FERET	27
LFW	23
YALE B	13
MULTI-PIE	15
CMU-PIE	11
CAS-PEAL	10
EXTEND YALE B	8
ORL	3
FRGV	4
FEI	2
SCFace	2
COX	2

Fonte: Própria.

Dentre as bases de dados públicas utilizadas para os testes temos a LFW como a mais desafiadora e a AR como a mais utilizada, sendo seguida por FERET, LFW, YALE B, MULT-PIE e CMU-PIE. A Tabela 2.4 mostra a relação entre bases de dados mais utilizadas e a utilização nos artigos revisados. Por se tratar das bases de dados mais utilizadas, não são listadas na tabela as 11 bases que foram utilizada apenas por um único artigo, sendo elas: FRGG 2.0; HFB; CASIA-WebFace; ChokePoint; JAFFE; FG-NET; UMIST; CK+; GeorgiaTech; MIT-CBCL; AT&T; CFP.

Um dos primeiros artigos selecionados para revisão foi o de DENG *et al.* [61], que já tratava do uso de SRC em seu método chamado *Superposed Sparse Representation based Classification* (SSRC). A principal ideia do trabalho era a escolha criteriosa de um dicionário, representado a imagem de teste como uma combinação linear esparsa das imagens de treino. Segundo os autores, com o crescimento do SRC, muitas variações foram criadas, cada uma incrementando o conceito de representação esparsa ou estendendo-o. Porém, o SRC foi muito criticado por precisar de muitas imagens e ser extremamente sensível a qualidade da imagem. Para superar os problemas citados, o trabalho utiliza uma abordagem que divide a matriz esparsa que é utilizada para representar a imagem de teste como uma combinação linear em dois dicionários, um com um protótipo, criado a partir dos centroides e o outro a partir da diferença dos centroides das imagens da galeria, o dicionário de variação, que representa as variações intraclasses, como variações de iluminação e expressão. No caso de uma galeria com uma única imagem por pessoa, esse dicionário de variação é extraído de uma base de dados auxiliar. A classificação é realizada por meio da comparação dos erros de reconstrução da imagem de teste.

Também baseado em SRC, o método de ZHUANG *et al.* [65] focam na transferência de aprendizagem, utilizando um dicionário focado em iluminação para transferência de variações. BORGI *et al.* [58] utilizam de uma *Shearlet Network* para gerar uma nova representação mais discriminante da galeria e da imagens de teste, e assim como SRC a representação da imagem de teste é reconstruída como uma combinação linear das novas representações da galeria.

LIU *et al.* [57] utilizaram de métodos locais, particionando as imagens em blocos e treinando tanto um classificador SRC quanto um classificador *Collaborative Repre-*

sentation Classification (CRC). Os resultados mostram o SRC com um desempenho um pouco melhor. CAI *et al.* [48] utilizaram um método baseado no SSRC, Porém ele utiliza os centroides das classes para criação do dicionário do protótipo e divide o dicionário da variação em dois dicionários, um com a similaridade intraclasse e outro com a diferença da variação intraclasse.

Também similar ao SSRC o trabalho de DING *et al.* [47] utilizaram um dicionário de variações, porém esse dicionário é construído por meio da base genérica e da galeria, não apenas pela base genérica. GAO *et al.* [46] utilizaram o particionamento em blocos e representação esparsa, porém, os blocos não são tratados isoladamente, para preservar as características globais. A imagem é tratada como uma matriz, onde cada bloco é representado como uma coluna dessa matriz e as partes menos descriminantes como uma reconstrução por meio do dicionário de variação intraclasse.

ZHUANG *et al.* [43] utilizaram uma abordagem similar a SSRC, porém eles consideram que as outras variações não são tão importantes quanto a iluminação. Devido à observação anterior, os autores divide as imagens em dois componentes: uma matriz de baixo *rank* que codifica a indentidade do indivíduo e uma matriz esparsa (ou dicionário) que representa as possibilidades de iluminação que é extraído de uma base auxiliar. Similar ao anterior, GU *et al.* [40] baseiam-se em SSRC, só que inicialmente ele usa pontos chaves para poder dividir o rosto em dois, utiliza *Scale-Invariant Feature Transform* (SIFT) para definir os pesos das regiões e divide a imagem em blocos. Para cada bloco é aplicado o filtro de Gabor e extraído o dicionário com o protótipo. O mesmo processo é utilizando na base auxiliar para extrair o dicionário de variação. Os autores destacam que o método não precisa de alinhamento da face antes do método, e que é mais rápido que métodos 3D ou que fazem alinhamento.

LIONG e HAIBIN YAN [27] utilizaram o mesmo princípio do SSRC, porém o dicionário de variação é aprendido a partir da combinação da galeria e da base auxiliar como um único conjunto de dados. Criando um dicionário que minimiza a diferença entre os conjuntos, facilitando a transferência das variações intraclasse da base auxiliar para a galeria.

LIU *et al.* [41] utilizaram de uma abordagem baseada em características locais,

dividindo as imagens da galeria em blocos e cada bloco organizado em estruturas locais, que são formadas pelos blocos e a vizinhança local dos blocos. Os autores consideram que todas as estruturas locais que pertencem à mesma pessoa residem no mesmo subespaço linear, dessa forma considera que a estrutura central de uma imagem de teste de um indivíduo encontra-se na galeria deve ser uma combinação linear de todas as estruturas locais da galeria. Para melhorar o resultado, os autores utilizam de uma técnica multi-fase para remover as classes com menos probabilidade de ser a correta, minimizando a entropia do sistema.

SONG *et al.* [14] utilizaram CRC para gerar o modelo e classificar a imagem, com a diferença que o dicionário de variação é construído por meio da própria galeria. Todas as imagens são divididas em blocos e em cada bloco é utilizado filtros de Gabor com escalas e orientações diferentes, dos quais, são extraído o dicionário de variações.

DENG *et al.* [32] também utilizaram os dicionários de protótipo e variação, sendo o primeiro feito por meio dos centroides e o segundo pela diferença dos centroides. Em seguida é computada a matriz de projeção *Principal Component Analysis* (PCA) pra redução de dimensionalidade e a mesma é aplicada aos dicionários. A classificação das imagens é feita por meio do menor erro de reconstrução.

HUANG *et al.* [29] utilizaram de SRC para classificação das imagens, porém as imagens que são utilizadas para criação dos dicionários são primeiramente passadas por projeções *Kernel Discriminant Analysis* (KDA) que são aprendidas de uma base auxiliar. Os dicionários são criados por meio da junção das projeções da base auxiliar com a galeria, a imagem de teste também é projetada por meio do KDA antes de ser calculado o erro residual da reconstrução da imagem e o rotulamento.

Baseado em CRC o trabalho de JI *et al.* [28] utilizaram as imagens da base auxiliar com mais afinidade com a imagem da galeria para gerar um subconjunto de referência, com o qual é utilizado para treinar a matriz de probabilidade dos rótulos. Para classificação de uma imagem de teste é utilizado um subconjunto de referência para ser usado na classificação e na matriz de probabilidade dos rótulos.

YU *et al.* [25] utilizaram um método baseado em SRC que cria um dicionário específico para oclusões a partir de uma base auxiliar, que é aprendido com base no PCA. Para detectar e desconsiderar pixeis discrepantes devido a oclusão, é desenvol-

vida uma estratégia de medição de erro multi-escala, que produz uma representação esparsa, robusta e altamente discriminativa. LIU e WASSELL [15] utilizaram de uma base auxiliar para extrair dicionários representando vários pontos de vista, como iluminação, expressão e pose. Os autores sintetizam novas imagens utilizando os dicionários, estendendo a galeria e utilizando o SRC padrão para classificá-las.

KAN *et al.* [11] propõem o *Adaptive Discriminant Analysis* (ADA), que tem como base o LDA, o qual encontra as projeções mais discriminantes por meio da razão da matriz de dispersão intraclasse e a matriz de dispersão interclasse. Devido à galeria SSPP (conjunto de treinamento com apenas uma imagem de referência por pessoa) só ter uma imagem, a matriz de dispersão intraclasse tende a ser zero. Para contornar essa situação o ADA utiliza uma base auxiliar genérica (base genérica) para extrair a matriz de dispersão intraclasse. O método utiliza de *K-Nearest Neighbors* (KNN) ou regressão Lasso para escolher as imagens da base genérica que mais se assemelha à imagem da galeria e por meio delas gerar as matrizes de dispersão intraclasse. As matrizes geradas por meio dos indivíduos da base genérica são unidas realizando uma soma aritmética delas. A matriz resultante é utilizada para gerar o conjunto de projeções discriminantes. Também baseado em LDA temos os trabalhos de YING LI *et al.* [6] que utilizaram de projeções randômicas para aumentar a base de treinamento.

Baseado em LDA, HU *et al.* [50] propõem a sintetização de imagens por meio de uma decomposição LU, em seguida utiliza de uma projeção para um subespaço utilizando 2D-LDA e classifica por meio dos vizinhos mais próximos nesse subespaço. HU *et al.* [44] também utilizaram a de LDA para fazer a transferência de aprendizado, porém utiliza uma estratégia de alinhamento de *manifold* para juntar a matriz de dispersão da base auxiliar com a galeria, gerar um subespaço em que é projetado as imagens de teste como a galeria e usado a vizinhança para classificar, método por ele chamado de *Discriminative Transfer Learning* (DTL). HU [30] propõe a aplicação de uma regularização esparsa para o DTL para torná-lo mais robusto a outliers e ruído. Método chamando de *Discriminative Transfer Learning with Sparsity Reguarization* (DTLSR).

Com base no particionamento em blocos, o método de PANG *et al.* [13] utilizaram de *graph embedding* para extrair informações discriminantes suficientes de dois

subespaço heterogêneo de características. É utilizado um método baseado em LDA que incorpora dois *Manifolds Embedding* para aprender representações heterogenias dos blocos das imagens, um que preserva os relacionamentos entre blocos da mesma pessoa e outro que suprime o relacionamento dos blocos de pessoas diferentes. Os autores ainda criam uma métrica de distância bloco para bloco e bloco para *manifold* e faz uso de uma estratégia de combinação de resultados pela junção dos votos majoritários das métricas dos dois subespaços.

Também baseado em blocos, CHU *et al.* [4] utilizaram da simetria facial para melhorar o reconhecimento e aumentar a base de treinamento. Os autores dividem a imagem da face em duas metades, a quais são dividida em blocos e são alinhados semanticamente. Os blocos são processados por uma transformação *Whitening* (similar ao PCA) e para cada bloco é criada uma matriz de projeção discriminante para cada bloco. A classificação é feita por vizinhos mais próximos no novo subespaço e os resultados de cada blocos são unidos por voto majoritário para chegar a uma classificação final.

LU *et al.* [62] apresentaram o *Discriminative Multimanifold Analysis* (DMMA), em que os autores dividem imagens da galeria em blocos (sem sobreposição) e representa esses blocos por meio de *Manifolds* (variedades), por meio deles encontram matrizes de projeção que maximizem a distância entre os *manifolds* da galeria. Para a classificação, a imagem de teste segue o mesmo processo de divisão e representação, a classificação é feita por meio da distância entre os *Manifolds*. YAN *et al.* [56] utilizaram uma abordagem similar, porém, esta abordagem gera vários blocos com características diferentes (e.g. intensidade dos pixeis, *Local Binary Patterns* (LBP) e Gabor).

Utilizando uma abordagem baseada em particionamento, o trabalho de ZHANG *et al.* [24] visa aprender um conjunto de filtros lineares discriminantes via *manifold*, esses filtros são aplicados nos blocos das imagens e os resultados são binarizados. Os blocos resultantes são organizados como um histograma e a classificação é feita por interseção de histogramas.

GAO *et al.* [45] utilizaram de *Deep Learning* para aprender redes auto codificadores supervisionadas (*auto-encoders*), essa rede é utilizada para remover ruídos das imagens, assim, as variações das imagens de teste são tratadas como ruídos. A

rede é utilizada para reconstruir a imagem de teste como sendo uma imagem sem as variações, e o resultado da reconstrução é comparado com a galeria e é classificada como a imagem com o menor erro de reconstrução.

GUO *et al.* [37] também propuseram um método baseado em redes neurais, que utiliza de uma rede rasa, de apenas duas camadas (sendo apenas uma camada escondida) para evitar o ajuste excessivo (*overfitting*) aos dados. Os autores afirmam utilizar características tanto holísticas quanto locais, e ser robusto à iluminação, expressão e pose. Na primeira camada é utilizada uma abordagem em blocos que seleciona características orientadas ao objetivo por meio de um conjunto *fuzzy rough*. Para contornar o problema das abordagem em blocos de sofrer com as regiões menos discriminantes, a segunda camada extrai informações estruturais globais das características selecionadas dos blocos por meio de *auto-encoders* esparsos, que remove a redundância dos dados. A saída da segunda camada são as características finais obtidas que serão utilizadas para classificação, no trabalho foram realizados testes utilizando uma camada *SoftMax* e testes utilizando uma versão linear do *Support Vector Machine* (SVM).

Utilizando redes neurais profundas similar a *trunk-branch ensemble* e as redes siamesas que contêm mais de uma sub-rede conectada a rede principal, PARCHAMI *et al.* [26] propõem a *Cross-Correlation Matching Convolutional Neural Network* (CCM-CNN), uma rede neural profunda que utiliza a função de perda baseada em *triplet-loss*. A rede consiste em três ramificações, uma para a imagem de referência, outra para uma segunda imagem com variações da mesma pessoa de referência e uma terceira com uma pessoa que não é a pessoa de referência. Essa rede é inicialmente treinada com uma base genérica maior, utilizando o *triplet-loss*, que funciona maximizando a similaridade entre as imagens da mesma pessoa e minimizando a similaridade com a pessoa diferente. Após o treino, a ramificação com a pessoa diferente da rede é removida da rede e a rede é refinada com a galeria como sendo a imagem de referência e imagens sintetizadas por operações de transformação como sendo a segunda imagem com variações.

YANG *et al.* [12] utilizaram de uma abordagem baseada em *Convolutional Neural Network* (CNN), a qual utiliza uma base genérica para treinar uma rede chamada de *Adaptive Convolution Feature* (ACF) para cada bloco das imagens (blocos seleci-

onados por pontos chaves). Uma vez treinada, a última camada da rede (*SoftMax*) é removida, e os *embeddings* resultantes da última camada são utilizados para criar o dicionário utilizado pelo CRC, o qual realiza a classificação final. OH *et al.* [10] utilizaram de uma rede neural de apenas uma camada treinada de forma não iterativa para extração de características Gabor sem necessidade de todas as convoluções. A última camada da rede utiliza uma projeção *Whitened Principal Component Analysis* (WPCA) para reduzir a dimensionalidade dos dados, que são classificados por meio de *Total Error Rate* (TER).

WANG *et al.* [66] fizeram uso da premissa que as variações intraclasse podem ser compartilhadas entre os indivíduos. Assim, ele faz uso de uma base genérica para extrair as variações e construir um modelo linear para cada indivíduo, utilizando as variações dos vizinhos mais próximo na base genérica. Dessa forma, ele transforma um problema em uma regressão linear resolvível por mínimos quadrados, onde a classificação é feita pelo menor erro de reconstrução. DENG *et al.* [55] também utilizaram regressão linear para a classificação, porém, deixando as imagens da galeria em distâncias iguais no subespaço de características, técnica chamada de *Equidistant Embedding*.

No trabalho de YANG *et al.* [59] é proposto a junção de arcabouço de aprendizagem de projeção adaptativa e o dicionário de variação esparsa. Extraindo do conjunto de treinamento genérico um subconjunto de referência (mais próximo da galeria) e um subconjunto de variação. O aprendizado da projeção adaptativa visa correlacionar o subconjunto de variação com a galeria, enquanto o aprendizado do dicionário visa aprender um dicionário compacto com bases esparsas de uma grande matriz de variação, que é a projeção da variação intraclasse do conjunto de treinamento genérico sobre a matriz de projeção aprendida.

PEI *et al.* [34] fizeram uso de um método não paramétrico, utilizando pirâmides de decisão para o reconhecimento de faces em SSPP. As imagens são particionadas em blocos e de cada bloco é extraído características que formam o conjunto de treinamento, com qual é construído a pirâmide de decisão. As imagens de teste são particionadas da mesma forma que a de treino e o método obtém a predição por meio da pirâmide.

ZHU *et al.* [33] abordam o problema da variação da iluminação em SSPP. O tra-

balho visa construir um método que elimina o efeito da iluminação extraindo duas características invariantes à iluminação: *Logarithm Gradient Orientation* (LGO) e *Logarithm Gradient Magnitude* (LGM). Por meio das características é gerado uma representação de histogramas, a qual é classificado por meio do vizinho mais próximo.

Inicialmente proposto para trabalhar com textos, ZHOU *et al.* [23] utilizaram uma variação do *Probabilistic Latent Semantic Analysis* (pLSA), o *Discriminat Probabilistic Latent Semantic Analysis* (DpLSA). O método consiste em particionar a imagem em blocos e extrair características *Bag-of-Word* (BoW) de cada bloco. A partir da concatenação das características dos blocos é aprendida uma representação semântica que é utilizada para criar um modelo probabilístico que é usado na classificação das imagens.

O trabalho de HONG *et al.* [36] abordaram a sintetização de imagens em modelos 3D utilizando *Domain Adaptation* (DA), que consiste em criar um mapeamento entre um domínio de fonte e um domínio alvo (nesse caso a galeria e as imagens de teste). Dessa forma, sendo possível criar um classificador no domínio fonte e utilizá-lo no domínio alvo. Para aumentar a galeria é utilizado uma técnica baseada em *Supervised Descent Method* (SDM) para gerar as imagens sintetizadas em 3D. A partir das imagens é criado uma *deep DA network* para extrair as características e classificar, porém, ao utilizar a base de dados LFW os autores utilizaram a saída da última camada da rede como entrada para um classificador SVM.

HU *et al.* [35] utilizaram de modelos 3D, porém fez uso de uma aplicação de terceiros, o FaceGen Modeller[1]. Os modelos gerados são rotacionados e os resultados são desfocados usando *Gaussian blurred* para simular a degradação gerada pelas câmeras de segurança. As imagens geradas são utilizadas para alimentar os classificadores SRC, LDA, SIFT, PCA e *Stacked Denoising Auto-Encoder* (SDAE).

JUEFEI-XU *et al.* [51] Utilizaram de 3D-*Generic Elastic Models* (GEM) para criar um modelo 3D da face e com isso rotaciona a imagem e coleta exemplos para treino da região dos olhos, pois afirma que é a região menos afetada pelas expressões e iluminação. Ele ainda utiliza de *modified active shape model* para encontrar os pontos chaves da face que serão mapeados para o modelo 3D genérico e Walsh

[1]http://facegen.com/index.htm

LBP nas imagens geradas antes da utilização do *Coupled Max-Pooling Kernel Class-Dependence Feature Analysis* (CoMax-KCFA), que cria um espaço discriminante por meio do domínio da frequência e do domínio do espaço.

HAGHIGHAT *et al.* [39] focam no trabalho com imagens em ambiente não controlado *"in the wild"*. Nesse cenário, para ter um método robusto à variação de pose ele utiliza de *Active Appearance Model* (AAM) aprendido a partir de uma base auxiliar em ambiente não controlado e utilizando um método próprio de inicialização do modelo AAM. As imagens de teste são alinhadas para uma imagem frontal por meio de um método baseado em AAM e então é extraído *Histogram of Oriented Gradients* (HOG) e Gabor que são unidos por análise de correlação canônica para ter um conjunto de características mais discriminantes.

Também utilizando 3D-GEM, DENG *et al.* [16] sintetizaram imagens em poses e iluminações (utilizando *Adaptive Quotient Image* (AQI)) diferentes. Para cada pose sintetizada, é aplicada uma técnica de regressão linear para transformar os exemplos sintetizados de um indivíduo em um único ponto no espaço métrico criado. Para o reconhecimento, inicialmente é estimada a pose da imagem de teste, e em seguida é utilizada a matriz de projeção da pose mais próxima para realizar a classificação.

Métodos como os de KVETON e VALKO [60] e ZHU *et al.* [64] utilizam de imagens em tempo real para atualização do modelo, o que não faz parte o escopo deste trabalho, pois a atualização depende de verdadeiros positivos resultante do próprio sistema. E uma falha na classificação pode colocar o sistema em colapso. Outro trabalho que aborda SSPP, mas se distancia do escopo deste trabalho é o de LU *et al.* [63], pois supõe-se que há um conjunto de imagens de teste a qual necessita ser rotulada, diferente do que é desejado neste trabalho, que é: dado uma imagem de teste, reconhece-la com base na galeria SSPP, sem o auxílio de imagens extras de indivíduos que estejam na galeria. Assim como no trabalho anterior, CHU *et al.* [38] utilizaram de uma base de dados não rotulada, a qual é propagada os rótulos para se usar como base de treinamento.

YIN *et al.* [54] ressaltam o problema da "maldição da dimensionalidade", mesmo sendo imagens pequenas, quando são agrupadas geram vetores ou matrizes gigantescas (e.g. Uma imagem de 32x32 em tons de cinza tem 1024 valores de pixel e em uma base de 100 indivíduos tem 102400 valores a serem tratados). O traba-

lho utiliza uma regressão linear dupla para o SSPP, porém os autores utilizaram imagens adicionais não rotuladas das imagens da galeria para propagar o rótulo e utilizar para o treino, o que foge do escopo deste trabalho, mas ele destaca que o *Regularized Discriminant Analysis* (RDA) é melhor que o LDA.

BASHBAGHI *et al.* [49] utilizaram como galeria imagens de uma base própria que contém a imagem da pessoa a ser reconhecida (galeria) e de outras pessoas como plano de fundo. As imagens tanto da galeria como dos outros indivíduos são particionadas e em cada partição é utilizada várias técnicas de extração de características. Então essas partições são utilizadas para treinar um conjunto de SVM maximizando a distância entre a galeria e os outros indivíduos do plano de fundo, método que também utiliza mais que a simples galeria SSPP. O método de CHU *et al.* [38] aborda o problema SSPP, porém se diferencia do escopo por trabalhar com imagens em baixa resolução.

MACHADO [52] propõe a utilização de *Factor Analysis* (FA) em uma forma 2D, utilizando-o como matriz e não como vetor, para maximizar a representação da correlação dos pixeis, e capturar informações mais importantes sobre os relacionamentos espaciais. FA é um método multivariante, que busca representar um conjunto de variáveis p como uma combinação linear de m construtos hipotéticos chamados fatores. Segundo o autor é um método de redução de dimensionalidade mais discriminante que o PCA.

ZHAO *et al.* [53] têm como base um ambiente de conferência, onde cada participante se registra com uma imagem e o sistema os reconhecem em outras imagens obtidas na conferência. Esse sistema utiliza cinco pontos chaves encontrados na imagem para extrair blocos da imagem que contêm os cantos da boca, olhos e ponta do nariz. Esses blocos são suavizados utilizando uma função gaussiana e extraído a probabilidade posterior bayesiana para classificação.

Também foi encontrado na literatura que técnicas de reconstrução facial 3D e transferência de iluminação têm sido empregadas para aprimorar conjuntos de dados de referência [67]. Da mesma forma, o método Uniform Generic Representation (UGR) combina representações locais e globais genéricas para lidar com variações em pose, iluminação e oclusão [68]. Abordagens híbridas, como a combinação de características elaboradas manualmente com características de aprendizado profundo de

CNNs, também demonstraram potencial em melhorar a precisão do reconhecimento [69].

Dos artigos revisados, o mais recente e promissor foi o de PANG *et al.* [8], que utilizou de representações desamanhadas e redes generativas adversárias para protótipos da imagens da galeria, preservando a identidade do indivíduo. O trabalho contava com 3 redes neurais, uma rede funcionando como um Codificador, outra como um Gerador/Decodificador e a última como um Discriminador. O Codificador gera vetor com uma representação da imagem de entrada em um espaço latente; O gerador utiliza o vetor do codificador, adicionado de um ruído aleatório para gerar um protótipo que preserve a identidade do indivíduo de entrada; A última rede é uma rede multi-tarefa, que recebe como entrada a saída do gerador e executa três tarefas: identifica se a imagem é real ou falsa; identifica o indivíduo na imagem; e identifica se a imagem possui variação ou não.

PANG *et al.* [17] utilizou do conceito de redes generativas aliado ao conceito de SRC, criando quatro redes neurais: Codificador; Gerador; Discriminador; e o Discriminador da variação. Assim, a imagem de treino é passada para o codificador, que, por sua vez, tem como saída dois vetores: (1)um vetor latente que representa o protótipo da imagem e (2) um vetor latente que representa a variação da imagem. O gerador é utilizado três vezes: a primeira para gerar a imagem contendo o protótipo a partir do vetor latente do protótipo; A segunda para gerar a imagem contendo as variações a partir do vetor latente da variação e por último a imagem contendo a reconstrução da imagem de entrada a partir da soma dos códigos latentes do protótipo e da variação. Das imagens geradas pela rede geradora, a imagem da variação é enviada para o Descritor de variação, o qual irá identificar o indivíduo da imagem. A última rede é uma rede multi-tarefa, que recebe como entrada a reconstrução da imagem e a identifica, assim como recebe o protótipo gerado e distingue se ele é real ou falso. Em seguida, a imagem do protótipo e da variação são utilizadas na equação do SRC.

Seguindo um princípio parecido ao anterior, porém utilizando também o conceito de SRC. PANG *et al.* [17] utilizou novamente o conceito de redes generativas

Outros autores que utilizam de probabilidade são LIU *et al.* [42], que divide a imagem em blocos com sobreposição para aumentar a galeria e minimizar a entropia

do sistema, retirando iterativamente as classes que têm menos probabilidade de ser a correta. A classificação é realizada por meio de uma eurística multi-fase.

2.3 Considerações finais

Neste capítulo foram apresentados os principais métodos utilizados na tarefa de reconhecimento de faces com apenas uma amostra por pessoa, bem como as bases de dados mais utilizadas nesses estudos. A partir da revisão, é possível observar a predominância de abordagens baseadas em *Compressive Sensing*, com destaque para o SRC e suas derivações, conforme expresso na Equação 2.1:

$$y = P\alpha + V\beta + z \tag{2.1}$$

Em que y é o vetor da imagem de teste que se deseja classificar, P é o de todas as amostras faciais de treinamento conhecidas (galeria), α é o vetor de coeficientes esparsos que indica como y é combinado linearmente a partir de P, V é o dicionário de variações, β é o vetor de coeficientes que expressa a contribuição das variações de V e é o ruido ou erro residual.

Esses métodos utilizam bases auxiliares para estimar variações de pose, iluminação e expressão, permitindo que a reconstrução da imagem de entrada seja realizada mesmo com poucas amostras por identidade. Além disso, destacam-se os avanços promovidos por técnicas baseadas em *Deep Learning*, em especial aquelas que combinam aprendizado supervisionado com redes neurais convolucionais profundas para extração de características discriminativas.

Outro ponto relevante identificado na revisão é o uso recorrente de bases auxiliares genéricas e da base LFW, composta por imagens de faces obtidas em ambientes não controlados. A maioria dos estudos utiliza a versão alinhada dessa base, e muitos dos melhores resultados nela obtidos fazem uso de técnicas baseadas em reconstrução 3D ou redes neurais, frequentemente em experimentos de verificação, e não de reconhecimento pleno.

A análise dos trabalhos também revelou que métodos como o VD-GAN buscam melhorar a capacidade de geração de protótipos a partir de uma única imagem por identidade. No entanto, esses modelos enfrentam limitações ao não separar

explicitamente os fatores de identidade e variação no espaço latente.

Diante dessas limitações, justifica-se a proposta de uma nova arquitetura, o AD-VAE, que combina a capacidade de representação *disentangled* dos VAE com o poder de síntese fidedigna dos GAN. A arquitetura proposta introduz uma função de perda supervisionada ($\mathcal{L}_C$) e um treinamento dividido em duas etapas, buscando superar os desafios do reconhecimento facial em ambientes não controlados com apenas uma imagem por pessoa. O próximo capítulo apresenta, em detalhes, a metodologia adotada e a implementação do AD-VAE.

Tabela 2.5: Utilização da base de dados LFW.

Autores	Quantidade	Auxiliar	Versão alinhada
ZHOU *et al.* [23]	158	0	Sim
OH *et al.* [10]	158	0	Sim
ZHANG *et al.* [24]	158	0	Não
YANG *et al.* [12]	50/901	108	Sim
HU [30]	1680	0	Não
HU *et al.* [35]	45	0	Sim
HONG *et al.* [36]	50	0	Sim
GUO *et al.* [37]	158	0	Não
LIU *et al.* [41]	158	0	Sim
LIU *et al.* [42]	158	0	Sim
HAGHIGHAT *et al.* [39]	50	0	Não
JUEFEI-XU *et al.* [51]	3000	1600	Não
HU *et al.* [44]	1680	0	Sim
GAO *et al.* [46]	80	78/1600	Sim
DING *et al.* [47]	151	250	Sim
YAN *et al.* [56]	158	0	Não
YANG *et al.* [59]	136	0	Sim
PANG *et al.* [8]	50	108	Sim

Fonte: Própria.

Capítulo 3

Antecedentes

Nesta sessão serão apresentados os principais técnicas que o método proposto utiliza como base.

3.0.1 Generative Adversarial Network (GAN)

Em tradução livre, as redes generativas adversárias (GAN) consiste em um jogo, chamado por GOODFELLOW *et al.* [70] de um jogo de mínimo e máximo. A GAN é formada por duas redes neurais profundas, a primeira rede é chamada de discriminador e tem como objetivo distinguir se uma dada imagem é real ou falsa. A segunda rede é chamada de gerador, e tem como objetivo gerar uma imagem que seja real o suficiente para enganar o discriminador, devido este jogo de uma rede tentando superar a outra elas são chamadas de adversárias. Enquanto uma tenta minimizar o erro, a outra tenta maximizar o acerto (*MinMax playing game*).

Nomeando a rede do gerador como G e a rede do discriminador como D, temos que o jogo de enganar e não ser enganado segue a seguinte função de valor V(G, D):

$$\begin{aligned}\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{\text{data}}(x)}[\log D(x)] \\ +\mathbb{E}_{z\sim p_z(z)}[\log\left(1 - D(G(z))\right)]\end{aligned} \tag{3.1}$$

Onde p_{data} é a distribuição das imagens de treino e p_z a distribuição do ruído (normalmente amostrado de uma distribuição gaussiana). Na prática, para prover um gradiente mais robusto o G pode ser treinado para maximizar $\log D(G(z))$. Sendo

assim, a equação 3.1 pode ser reformulada como:

$$\begin{aligned}\max_D V_D(G,D) = \mathbb{E}_{x\sim p_{\text{data}}(x)}[\log D(x)] \\ +\mathbb{E}_{z\sim p_z(z)}[\log(1-D(G(z)))]\end{aligned} \tag{3.2}$$

$$\max_G V_G(G,D) = \mathbb{E}_{z\sim p_z(z)}[\log(D(G(z)))] \tag{3.3}$$

3.0.2 Disentangled Representation learning-Generative Adversarial Network (DR-GAN)

O método é baseado no GAN[70], fazendo o uso de redes neurais geradora e discriminadora. O *Disentangled Representation learning-Generative Adversarial Network* (DR-GAN) realiza a sintetização de imagens invariante a poses, utilizando do conceito de Desembaraçamento ou do inglês "*Disentangled*", que resumidamente visa separar características específicas da representação dos dados. Essa separação faz com que o método não necessite se preocupar em gerar uma representação dessas características específicas, focando a atenção apenas no que interessa.

Diferente do GAN, que utiliza apenas um vetor de ruído como entrada de uma rede neural que gera imagem a partir desse ruído, o DR-GAN utiliza uma estrutura similar a um Autoencoder no lugar do Gerador. O método utiliza uma rede Codificadora G_{enc} que recebe uma imagem X como entrada e gera um código latente $f(x)$ e uma rede Decodificadora G_{dec} que recebe como entrada o código latente $f(x)$ concatenado a um ruído aleatório z e um vetor de pose c para gerar uma imagem $\hat{X}$ com a pose determinada no vetor de pose c.

Assim como no GAN, o método utiliza uma rede discriminadora, porém, ao invés de uma rede neural com uma única saída (real/falsa), o discriminador do DR-GAN funciona como uma rede neural multi-tarefa. O discriminador D tem como saída o id do indivíduo de teste, se a imagem é real ou falsa (GAN tradicional) e a pose do indivíduo. Podendo ser representada como $D^{id}, D^{gan}, D^{pose}$, a figura 3.1 mostra a arquitetura do DR-GAN.

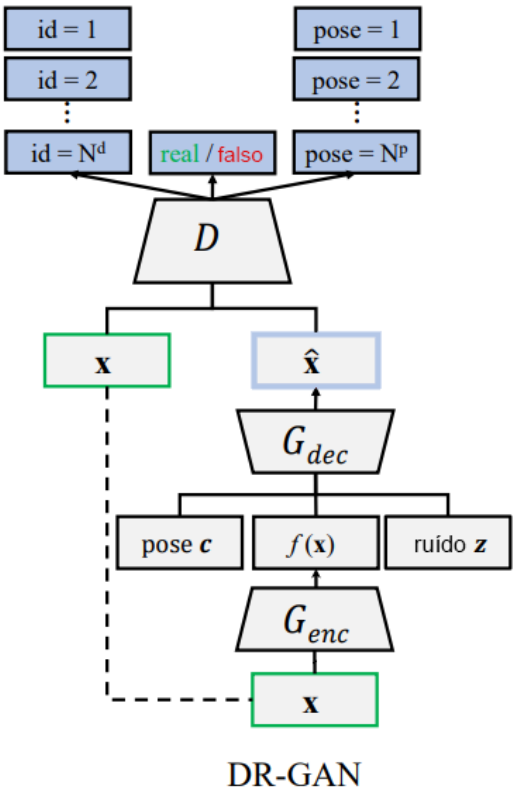

Figura 3.1: Arquitetura DR-GAN

Fonte: TRAN *et al.* [7].

3.0.3 Variation Disentangled Generative Adversarial Autoencoder Network (VD-GAN)

O VD-GAN[8] tem como base o DR-GAN, utilizando também a rede geradora como um Autoencoder e a rede discriminadora como uma rede multi-tarefa. O método tem como objetivo o reconhecimento de face com apenas uma imagem para treino por indivíduo, realizando o reconhecimento a partir da geração de um protótipo que conserve a identidade do indivíduo de forma que o indivíduo possa ser reconhecido por meio desse protótipo.

A técnica utiliza um Codificador G_{enc} que gera um código latente $f(x)$ de uma imagem X, $f(x) = G_{enc}(X)$. O decodificador G_{dec} tem como entrada o código latente gerado concatenado a um ruído aleatório $z \sim p(z)$ (sendo $p(z)$ uma distribuição Gaussiana) para gerar o protótipo $\hat{X}$ preservando a identidade do indivíduo, $\hat{X} = G_{dec}(f(x), z) = G_{dec}(G_{enc}(X), z)$.

Assim como o DR-GAN a rede discriminadora D é uma rede multi-tarefa, com 3 saídas: D^{id} que identifica a identidade do indivíduo; D^{gan} que identifica se a imagem é real ou falsa; e D^{var} que identifica se a imagem tem ou não variações. Diferente do DR-GAN que fazia o desembaraçamento apenas da pose, o VD-GAN considera qualquer mudança como variação (incluindo pose). A figura 3.2 mostra a arquitetura do VD-GAN.

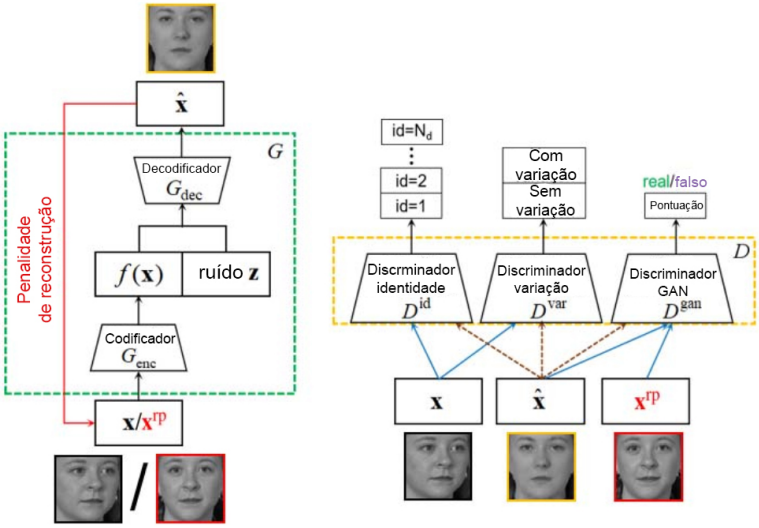

Figura 3.2: Arquitetura VD-GAN

Fonte: PANG *et al.* [8].

O método treina o autoencoder $G(\ G_{enc}, G_{dec})$ com a seguinte função objetivo:

$$\max_G V_G = V_G^{gan} + \mu_1 V_G^{id} + \mu_2 V_G^{var} - \mu_3 V_G^{rec} \tag{3.4}$$

Onde μ_1, μ_2 e μ_3 são os pesos dos hiper-parâmetros para o objetivo híbrido V_G. Os quatro sub-objetivos são definidos a seguir:

$$V_G^{id}(G, D^{id}, \mathbf{x}, \mathbf{z}) = \mathbb{E}_{x,y_{id},z}[\log D_{y_{id}}^{id}(G(\mathbf{x}, \mathbf{z}))] \tag{3.5}$$

$$V_G^{var}(G, D^{var}, \mathbf{x}, \mathbf{z}) = \mathbb{E}_{x,y_{var},z}[\log D_{y_{var}}^{var}(G(\mathbf{x}, \mathbf{z}))] \tag{3.6}$$

$$V_G^{gan}(G, D^{gan}, \mathbf{x}, \mathbf{z}) = \mathbb{E}_{x,z}[\log D^{gan}(G(\mathbf{x}, \mathbf{z}))] \tag{3.7}$$

$$V_G^{rec}(G, \mathbf{x}_{rp}, \mathbf{z}) = \mathbb{E}_{\mathbf{x}_{rp},\mathbf{z}}[\frac{1}{2} \left\| \mathbf{x}_{rp} - G(\mathbf{x}_{rp}, \mathbf{z}) \right\|_F^2] \tag{3.8}$$

onde $\mathbf{x}, \mathbf{x}_{rp}, \mathbf{y}_{id}, \mathbf{y}_{var}$ provém dos dados de treinamento $X = \{[x^1, x^1_{rp}, y^1_{id}, , y^1_{var}], ..., [x^n, x^n_{rp}, y^n_{id}, , y^n_{var}]\}$.

O discriminador é treinado com a seguinte função objetivo:

$$\max_D V_D = V_D^{gan} + \lambda_1 V_D^{id} + \lambda_2 V_D^{var} \tag{3.9}$$

onde λ_1 e λ_2 são os pesos dos hiper-parâmetros, V_D^{id}, V_D^{var} e V_D^{gan} são definidos a seguir:

$$V_D^{id}(D^{id}, \mathbf{x}) = \mathbb{E}_{x,y_{id}}[\log D^{id}_{y_{id}}(\mathbf{x})] \tag{3.10}$$

$$V_D^{var}(D^{var}, \mathbf{x}) = \mathbb{E}_{x,y_{var}}[\log D^{var}_{y_{var}}(\mathbf{x})] \tag{3.11}$$

$$\begin{aligned} V_D^{gan}(D^{id}, \mathbf{x}) = \mathbb{E}_{\mathbf{x}_{rp}}[\log D^{gan}(\mathbf{x}_{rp})] \\ +\mathbb{E}_{\mathbf{x},\mathbf{z}}[\log(1 - D^{gan}(G(\mathbf{x}, \mathbf{z})))] \end{aligned} \tag{3.12}$$

3.0.4 Variational Autoencoder (VAE)

O *Variational Autoencoder* (VAE)[71] difere dos Autoencoder tradicionais por não utilizar o codificador (*encoder*) para gerar um código latente (Como utilizado no DR-GAN e VD-GAN). O codificador do VAE em como saída a média e a variância da distribuição de probabilidade do espaço latente da imagem de entrada. O método assume que uma imagem x de um conjunto de treinamento X é resultado de uma função determinística $f(z)$ sendo z uma variável randômica $z \sim p(z)$ do espaço latente Z de forma que $f : (z, \epsilon) \rightarrow x$, sendo ϵ um ruído estocástico.

A probabilidade de observando z conhecer x é estimada pelo decodificador $p_{\theta_d} : z \mapsto p_{\theta_d}(x \mid z)$ parametrizado por θ_d e a probabilidade de z ser a código latente de x é estimado pelo codificador $q_{\theta_e} : x \mapsto q_{\theta_e}(z \mid x)$ parametrizado por θ_d. Sendo o conjunto de dados $X = (x^{(1)}, ..., x^{(n)})$ com n número de imagens, os parâmetros do modelo são obtidos maximizando a log-verossimilhança das variáveis observadas: $\log p_{\theta_d}(x^{(i)}) = \log \int_Z p_{\theta_d}(x^{(i)} \mid z)p(z)dz$. O $\log p_{\theta_d}(x^{(i)})$ é computado maximizando o limite inferior tratável (*Evidence Lower Bound* (ELBO)). O treinamento do VAE é realizado com a seguinte função de perda:

$$\begin{aligned}\mathcal{L}_{VAE}(\theta_e, \theta_d; x)) = \mathbb{E}_{q_{\theta_e}(z|x)}[-\log p_{\theta_d}(x \mid z)] \\ +KL(q_{\theta_e}(z \mid x) \,||\, p(z)))\end{aligned} \tag{3.13}$$

onde p_{θ_d} é normalmente utilizada como uma distribuição gaussiana $\mathcal{N}(x; \mu_{\theta_d}(z), Id)$ e KL é a divergência de Kullback-Leibler [72, 73]. A equação 3.13 estima o erro de reconstrução e força a distribuição do espaço latente a corresponder à distribuição $p(z)$. Normalmente $p(z)$ é uma distribuição gaussiana padrão $\mathcal{N}(z; 0, Id)$.

3.0.5 Adversarial Variational Autoencoder (AVAE)

Assim como o DR-GAN o *Adversarial Variational Autoencoder* (AVAE) [9] foi proposto para sintetização de imagens. O método realiza a junção do VAE com o GAN, para isso, ele utiliza quatro redes neurais em um processo dividido em duas partes, uma parte VAE e uma parte GAN. A parte VAE funciona de forma similar ao VAE clássico, contendo um codificador (*encoder*) E_{θ_e} e um decodificador (*decoder*) D_{θ_d}.

O modelo é parametrizado por $q_{\theta_e}(z|x) = \mathcal{N}(z; \mu_{\theta_e}(x), \sum_{\theta_e})$, sendo $\sum_{\theta_e}$ a matriz diagonal $diag(\sigma^2_{\theta_e})$. A probabilidade a priori do código latente é $p(z) = \mathcal{N}(z; 0, Id)$ e $p_{\theta_d}(x|z) = \mathcal{N}(x; \mu\theta_d(z), Id)$. Assim, o $\mathcal{O}VAE(\theta_e, \theta_d; x)$ pode ser estimado por meio do método de Monte-Carlo. O casamento entre a função de distribuição do espaço latente e a distribuição a priori $p(z)$ é realizado utilizando a divergência de Kullback-Leibler na forma da função de perda $KL(q_{\theta_e}(z|x)||p(z))$ que é igual a:

$$\frac{1}{2} \sum_{j=1}^{dim(Z)} \sigma^2_{Encj} + \mu^2_{Enc}(x)_j - 1 - \log \sigma^2_{Encj} \tag{3.14}$$

Assim, sendo z amostrado de $q_{\theta_e}(z|x)$, a função de perda da parte VAE para uma amostra é descrita a seguir:

$$\begin{aligned}\mathcal{L}_{VAE}(\theta_e, \theta_d; x)) = \frac{1}{2} \left\| \mu_{\theta_d}(z)) - x \right\|^2 \\ +\frac{1}{2} \sum_{j=1}^{dim(\mathcal{Z})} \sigma^2_{\theta_e j} + \mu^2_{\theta_e}(x)_j - 1 - \log \sigma^2_{\theta_e j}\end{aligned} \tag{3.15}$$

Na parte do Gerador (Parte GAN) o autor utiliza três redes neurais, o codificador VAE treinado no passo anterior e duas novas redes: o Gerador G_{θ_g} e o Discriminador C_{θ_c}. A rede geradora tem como entrada o código latente z concatenado com um vetor de erro ξ amostrado de uma distribuição $\mathcal{N}(0, Id)$. z codifica as informações capturadas pelo codificador, enquanto ξ codifica a variação não capturada pelo codificador. A rede C_{θ_c} funciona como o Discriminador normal do GAN, distinguindo se a imagem é real ou falsa. Assim, a imagem é gerada por $\hat{x} = G_{\theta_g}(z, \xi)$, tendo como função de perda 3.16.

$$\mathcal{L}_G(\theta_g) = \mathbb{E}[-\log p_{\theta_e}(z|x)] + \mathrm{KL}(p_{\theta_g}(x)|p(x)) + H_{\theta_g} + C \tag{3.16}$$

Capítulo 4

Metodologia

Neste capítulo é abordada a apresentação do *Adversarial Disentangled Variational Autoencoder* (AD-VAE) e detalhes do método proposto.

4.1 Método Proposto

O método proposto *Adversarial Disentangled Variational Autoencoder* (AD-VAE) é uma combinação da efetividade de sistemas baseados em VAE em aprender representações desembaraçadas [74], com a sintetização de imagens com alta fidelidade das técnicas baseadas em GAN, assim como descrito por LEE *et al.* [20]. Similar a PANG *et al.* [8] utilizamos o desembaraçamento das variações e as representações de identidade discriminativas por meio das redes neurais profundas baseadas em GAN, porém, complementamos as representações desembaraçadas por meio do codificador que utiliza o VAE como base.

A arquitetura AD-VAE foi selecionada por sua capacidade inerente de aprender representações de identidade desentrelaçadas de fatores de variação como pose, iluminação e expressão, um aspecto crucial para o reconhecimento facial com uma única amostra por pessoa (SSPP-FR). Uma de suas principais vantagens sobre outros métodos de aprendizado profundo com melhores resultados reside no fato de não exigir grandes volumes de dados de treinamento por indivíduo, o que é ideal para o cenário de SSPP-FR. Essa abordagem permite a geração de protótipos faciais robustos e em condições neutras a partir de uma única imagem, enriquecendo efetivamente os dados de treinamento e superando a escassez de amostras intra-classe. A com-

binação sinérgica de um Variational Autoencoder, para uma representação latente probabilística, e de uma Generative Adversarial Network, para a síntese realista de imagens, confere ao AD-VAE uma vantagem significativa sobre outros métodos de aprendizado profundo, que frequentemente falham em desvincular efetivamente a identidade das variações em cenários não controlados.

4.1.1 Arquitetura da Parte VAE

Assim como o AVAE, a arquitetura do AD-VAE pode ser dividida em duas partes, uma parte VAE e uma parte GAN. A primeira parte da arquitetura é ilustrada na figura 4.1 e a segunda parte é ilustrada na figura 4.2.

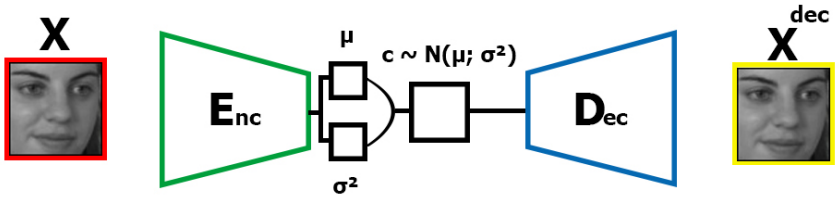

Figura 4.1: A primeira parte da arquitetura proposta do AD-VAE. Que trabalha como um VAE, onde $\mathbf{x}$ é a imagem oriunda do conjunto de dados $\mathbf{X}$, e $\mathbf{x}^{dec}$ denota a reconstrução de $\mathbf{x}$ feita pelo decodificador D_{ec}. O codificador E_{nc} tem como entrada a imagem x e como saída a média μ e a variância σ^2 da distribuição de probabilidade do esapaço latente de $\mathbf{x}$. A partir dessa da distribuição $\mathcal{N}(\mu, \sigma^2)$ é amostrado um vetor latente c sendo $c \sim \mathcal{N}(\mu, \sigma^2)$, que serve como entrada do decodificador D_{ec} que tem como saída a imagem reconstruída $\mathbf{x}^{dec}$.

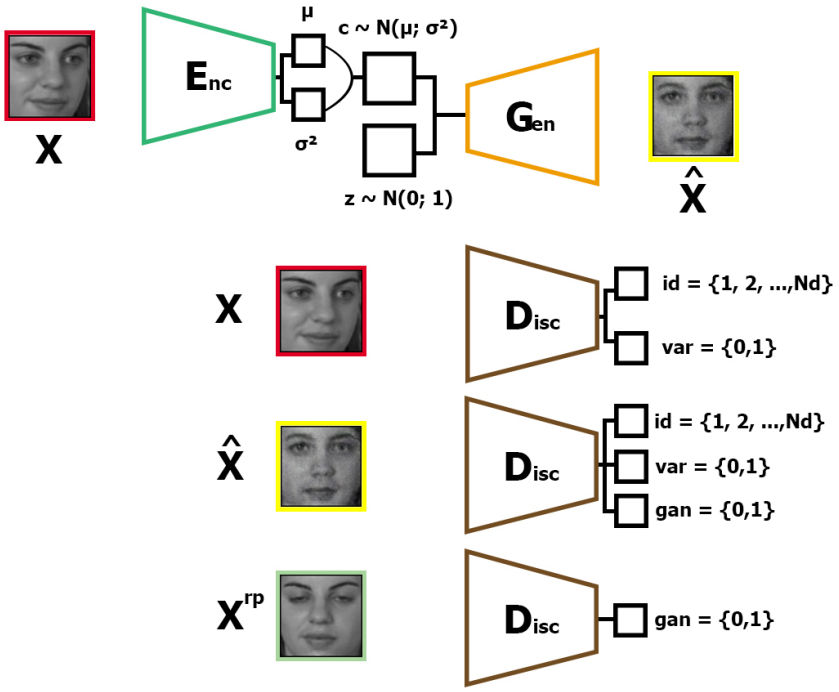

Figura 4.2: A segunda parte da arquitetura proposta do AD-VAE. Onde $\mathbf{x}$ provem do conjunto de dados $\mathbf{X}$, $\mathbf{x}^{rp}$ denota a protótipo real da imagem $\mathbf{x}$, $\hat{\mathbf{x}}$ é o protótipo gerado a partir da imagem $\mathbf{x}$. O codificador treinado na primeira parte (Parte VAE) E_{nc} gera a partir de uma imagem $\mathbf{x}$ a média μ e a variância σ^2. Em seguida é extraído um vetor c do espaço latente $c \sim \mathcal{N}(\mu, \sigma^2)$, esse vetor latente é concatenado com um vetor de ruído aleatório z, oriundo de uma distribuição normal $z \sim \mathcal{N}(0, 1)$. Os vetores c e z são concatenados e passado como entrada para o gerador G_{en} que por sua vez tem como saída o protótipo $\hat{\mathbf{x}}$ de $\mathbf{x}$. O discriminador D_{isc} é usado para: (1) identificar o id do indivíduo e se há variações da amostra $\mathbf{x}$; (2) identificar o id do indivíduo, se a imagem é real ou falsa e se há variações da amostra $\hat{\mathbf{x}}$; (3) discriminar se $\mathbf{x}_{rp}$ é real ou falsa.

Considerando os avanços no uso de VAE e GAN, como o VAE-GAN para síntese de cenas internas 3D de [75], o VAE-GAN híbrido de [76] e o Patch VAE-GAN de [72], propomos um Autoencoder Variacional Desentrelado Adversarial (AD-VAE). Projetado especificamente para enfrentar os desafios do SSPP FR, aprendendo representações de identidade robustas, a arquitetura do framework é formada por quatro redes neurais profundas, com o treinamento dividido em duas partes. Essas partes são treinadas sequencialmente. A primeira parte utiliza duas redes neurais de forma similar ao VAE original. Usando o codificador $E_{nc}(x)$ com uma imagem x de uma distribuição de dados $x \sim P_{data}(x)$ obtemos como saída média μ e a variância σ^2 do espaço latente de x. Utilizando a distribuição de probabilidade formada pela média e a variância adquirida, é amostrado um vetor c de $c \sim \mathcal{N}(\mu, \sigma^2)$. Esse vetor amostrado c é utilizado como entrada do decodificador $D_{ec}(c)$ para gerar a reconstrução

da imagem x que foi dada como entrada para o codificador E_{nc}.

A parte VAE tem como objetivo treinar o codificador para aprender a distribuição do espaço latente da imagem de entrada que contenha representações independentes e discriminativas da imagem, de forma a preservar a identidade do indivíduo. Assim como PLUMERAULT *et al.* [9], o erro de reconstrução é estimado pelo método de Monte Carlo e a divergência de Kullback-Leiber $KL(E_{nc}(z|x) \,||\, p(z))$, que é definido como:

$$\frac{1}{2} \sum_{j=1}^{dim(Z)} \sigma^2_{E_{ncj}} + \mu^2_{E_{nc}}(x)_j - 1 - \log \sigma^2_{E_{ncj}} \tag{4.1}$$

A função de custo completa da parte VAE é definida como:

$$\begin{aligned} \mathcal{L}_{VAE}(E_{nc}, D_{ec}; x)) = \frac{1}{2} \left\| \mu_{D_{ec}}(z)) - x \right\|^2 \\ + \frac{1}{2} \sum_{j=1}^{dim(Z)} \sigma^2_{E_{ncj}} + \mu^2_{E_{nc}}(x)_j - 1 - \log \sigma^2_{E_{ncj}} \end{aligned} \tag{4.2}$$

4.1.2 Arquitetura da Parte GAN

Após o treino da parte VAE, a segunda parte é treinada utilizando o codificador E_{nc} treinado no passo anterior, formando uma estrutura codificador-gerador similar a um *Autoencoeder* tradicional. O gerador G_{en} utiliza como entrada o vetor de código latente c amostrado da distribuição $c \sim \mathcal{N}(\mu_{E_{nc}}, \sigma^2_{E_{nc}})$ gerada pelo codificador E_{nc} a partir da imagem x, o vetor c é concatenado a um vetor de ruído aleatório z amostrado de uma distribuição gaussiana $z \sim \mathcal{N}(0, 1)$.

Nomeando a função que faz a amostragem do vetor do espaço latente gerado pelo codificador como $c = f(E_{nc}(x)) = f(\mu_{\mathbf{E}_{nc}}, \sigma_{\mathbf{E}_{nc}})$, temos que $\hat{x} = G_{en}(f(E_{nc}(x)), z)$, sendo $\hat{x}$ o protótipo gerado com a mesma identidade da imagem x.

4.1.3 Funções de Perda

Para treinamento do gerador G_{en} são utilizados as seguintes cinco funções objetivas:

$$\max_{\mathbf{G}_{en}} V_{\mathbf{G}_{en}} = V^{gan}_{\mathbf{G}_{en}} + \mu_1 V^{id}_{\mathbf{G}_{en}} + \mu_2 V^{var}_{\mathbf{G}_{en}} - \mu_3 V^{rec}_{\mathbf{G}_{en}} + \mathcal{L}_{\mathcal{C}} \tag{4.3}$$

onde μ_1, μ_2 e μ_3 são os pesos dos hiper parâmetros para $V_{\mathbf{G}_{en}}$. Os sub-objetivos são definidos pelas funções a seguir:

$$V^{id}_{\mathbf{G}_{en}}(\mathbf{G}_{en}, \mathbf{D}_{is}{}^{id}, \mathbf{c}, \mathbf{z}) = \mathbb{E}_{c,y_{id},z}[\log \mathbf{D}_{is}{}^{id}_{y_{id}}(G(\mathbf{c}, \mathbf{z}))] \tag{4.4}$$

$$V^{var}_{\mathbf{G}_{en}}(\mathbf{G}_{en}, \mathbf{D}_{is}{}^{var}, \mathbf{c}, \mathbf{z}) = \mathbb{E}_{c,y_{var},z}[\log \mathbf{D}_{is}{}^{var}_{y_{var}}(G(\mathbf{c}, \mathbf{z}))] \tag{4.5}$$

$$V^{gan}_{\mathbf{G}_{en}}(\mathbf{G}_{en}, \mathbf{D}_{is}{}^{gan}, \mathbf{c}, \mathbf{z}) = \mathbb{E}_{c,z}[\log \mathbf{D}_{is}{}^{gan}(\mathbf{G}_{en}(\mathbf{c}, \mathbf{z}))] \tag{4.6}$$

$$V^{rec}_{\mathbf{G}_{en}}(\mathbf{G}_{en}, \mathbf{x}_{rp}, \mathbf{z}) = \mathbb{E}_{\mathbf{x}_{rp},\mathbf{z}}[\frac{1}{2} \left\| \mathbf{x}_{rp} - \mathbf{G}_{en}(f(\mathbf{E}_{nc}(\mathbf{x}_{rp})), \mathbf{z}) \right\|^2_F] \tag{4.7}$$

$$\mathcal{L}_{\mathcal{C}}(\mathbf{G}_{en}, \mathbf{x}, \hat{\mathbf{x}}) = \frac{1}{2} \left\| \frac{\mu_{\mathbf{E}_{nc}}(\mathbf{x}_{rp}) \ - \mu_{\mathbf{E}_{nc}}(\hat{\mathbf{x}})}{\sigma_{\mathbf{E}_{nc}}(\hat{\mathbf{x}})} \right\|^2 \tag{4.8}$$

onde $\mathbf{x}, \mathbf{x}_{rp}, \mathbf{y}_{id}, \mathbf{y}_{var}$ são oriundos do conjunto de dados de treinamento $X = \{[x^1, x^1_{rp}, y^1_{id}, , y^1_{var}], ..., [x^n, x^n_{rp}, y^n_{id}, , y^n_{var}]\}$. Sendo x^i uma imagem aleatória do indivíduo i, x^i_{rp} o protótipo real do indivíduo i (protótipo é uma imagem em posição, expressão e iluminação neutra ou próximas a neutra do indivíduo em questão), y^i_{id} é o rótulo que identifica o indivíduo i (id) e y^i_{var} é o rótulo que define se a imagem x^i têm alguma variação (o campo variação distingue se a imagem é uma imagem neutra ou não).

As funções de custo dos sub-objetivos do gerador $\mathbf{G}_{en}$ têm como objetivos específicos:

- $V^{id}_{\mathbf{G}_{en}}$: Possibilitar que o classificador $\mathbf{D}_{is}{}^{id}$ gera uma imagem protótipo $\hat{\mathbf{x}}$ que possa ser rotulada com a mesma identidade y^{id} de $\mathbf{x}$.

- $V^{var}_{\mathbf{G}_{en}}$: Possibilitar que o classificador $\mathbf{D}_{is}{}^{var}$ possa detectar se há ou não variação na imagem protótipo $\hat{\mathbf{x}}$ e na imagem de entrada $\mathbf{x}$.

- $V^{gan}_{\mathbf{G}_{en}}$: Enganar o classificador $\mathbf{D}_{is}{}^{gan}$ para classificar a imagem protótipo $\hat{\mathbf{x}}$ como sendo uma imagem real.

- $V^{rec}_{\mathbf{G}_{en}}$: Permitir que o gerador produza uma imagem $\hat{\mathbf{x}}$ que seja o mais próximo possível da imagem protótipo real $\mathbf{x}_{rp}$

- $\mathcal{L}_{\mathcal{C}}$: Proporcionar que o gerador produza uma imagem $\hat{\mathbf{x}}$, de forma que a distribuição a priori de $\mathbf{E}_{nc}(\hat{\mathbf{x}})$ seja perto o suficiente da distribuição a priori

de $\mathbf{E}_{nc}(\mathbf{x}_{rp})$.

A última rede neural do método proposto $\mathbf{D}_{is}$ é treinada utilizando a seguinte função objetivo:

$$\max_{\mathbf{D}_{is}} V_{\mathbf{D}_{is}} = V^{gan}_{\mathbf{D}_{is}} + \lambda_1 V^{id}_{\mathbf{D}_{is}} + \lambda_2 V^{var}_{\mathbf{D}_{is}} \tag{4.9}$$

onde λ_1 e λ_2 são os parâmetros de compensação, e as funções V^{id}_D, V^{var}_D e V^{gan}_D são definidas nas equações a seguir:

$$V^{id}_{\mathbf{D}_{is}}(\mathbf{D}_{is}{}^{id}, \mathbf{x}) = \mathbb{E}_{x,y_{id}}[\log \mathbf{D}_{is}{}^{id}_{y_{id}}(\mathbf{x})] \tag{4.10}$$

$$V^{var}_{\mathbf{D}_{is}}(\mathbf{D}_{is}{}^{var}, \mathbf{x}) = \mathbb{E}_{x,y_{var}}[\log \mathbf{D}_{is}{}^{var}_{y_{var}}(\mathbf{x})] \tag{4.11}$$

$$\begin{aligned} V^{gan}_{\mathbf{D}_{is}}(\mathbf{D}_{is}{}^{id}, \mathbf{x}) = \mathbb{E}_{\mathbf{x}_{rp}}[\log \mathbf{D}_{is}{}^{gan}(\mathbf{x}_{rp})] \quad + \\ \mathbb{E}_{\mathbf{x},\mathbf{z}}[\log (1 - \mathbf{D}_{is}{}^{gan}(\mathbf{G}_{en}(\mathbf{f}(\mathbf{E}_{nc}(\mathbf{x})), \mathbf{z})))] \end{aligned} \tag{4.12}$$

As funções dos sub-objetivos de $\mathbf{D}_{is}$ têm os seguintes objetivos:

- V^{id}_D: Estimar a identidade correta da imagem de entrada $\mathbf{x}$, conforme indicada em y_{id}.
- V^{var}_D: Determinar a ocorrência correta de variação na imagem de entrada $\mathbf{x}$, conforme rotulada em y_{var}.
- V^{gan}_D: Identificar a imagem protótipo real $\mathbf{x}_{rp}$ como real e prever uma imagem protótipo gerada $\hat{\mathbf{x}}$ como falsa.

4.1.4 Fluxo de Treinamento

Como especificado anteriormente, as 4 redes são treinadas sequencialmente, sendo o codificador $\mathbf{E}_{nc}$ e o decodificador $\mathbf{D}_{ec}$ treinados por meio da equação (4.2). Enquanto a rede geradora $\mathbf{G}_{en}$ e a rede discriminadora $\mathbf{D}_{is}$ são treinadas por meio da equação (4.3) e $\mathbf{D}_{is}$ equação (4.9) respectivamente. Como resultado do treinamento o codificador $\mathbf{E}_{nc}$ aprende uma distribuição de probabilidade aproximada do espaço latente de x, de forma a conseguir uma representação mais discriminativa da imagem de entrada x por meio do treinamento baseado em VAE. A partir do espaço latente

do codificador $\mathbf{E}_{nc}$ o gerador $\mathbf{G}_{en}$ aprende a criar um protótipo $\hat{x}$ em posição neutra ou mais próxima possível da condição de neutralidade que preserva características da identidade do indivíduo de x.

A partir dos protótipos gerados é possível identificar o indivíduo comparando o protótipo gerado a partir da entrada e o protótipo da imagem da galeria. No próximo capítulo é descrito a implementação do método, assim como as bases e os resultados dos experimentos.

Initialize parameters of the models: $\theta_e, \theta_d, \theta_g, \theta_c$

while training do

{Forward pass.}

$x^{\text{real}} \leftarrow$ batch of images sampled from the dataset.

$z_{\mu}^{\text{real}}, z_{\sigma}^{\text{real}} \leftarrow E_{\theta_e}(x^{\text{real}})$

$z^{\text{real}} \leftarrow z_{\mu}^{\text{real}} + z_{\sigma}^{\text{real}} \odot \epsilon$ with $\epsilon \sim N(0,1)$

$x^{\text{rec}} \leftarrow D_{\theta_d}(z^{\text{real}})$

$x^{\text{gen}} \leftarrow G_{\theta_g}(z^{\text{real}}, \xi)$ with $z^{\text{real}}, \xi \sim N(0,1)$

$C^{\text{real}}, C^{\text{gen}} \leftarrow C_{\theta_c}(x^{\text{real}}), C_{\theta_c}(x^{\text{gen}})$

{Compute losses gradients and update parameters.}

$\theta_e \leftarrow \nabla_{\theta_e} L_{\text{VAE}}(\theta_e, \theta_d); \quad \theta_g \leftarrow \nabla_{\theta_g} L_G(\theta_g)$

$\theta_d \leftarrow \nabla_{\theta_d} L_{\text{VAE}}(\theta_e, \theta_d); \quad \theta_c \leftarrow \nabla_{\theta_c} L_C(\theta_c)$

end while

(4.13)

$$L_{\text{VAE}}(\theta_e, \theta_d; x) = \frac{1}{2}||\mu_{\theta_d}(z) - x||^2 + \frac{1}{2}\sum_{j=1}^{dim(Z)} \sigma_{\theta_{ej}}^2 + \mu_{\theta_e}^2(x)_j - \log \sigma_{\theta_{ej}}^2 \tag{4.14}$$

$$L_G(\theta_g) = {\theta_g}^{gan} + \mu_1 {\theta_g}^{id} + \mu_2 {\theta_g}^{var} - \mu_3 {\theta_g}^{rec} + \mathcal{L}_{\mathcal{C}} \tag{4.15}$$

$$L_C(\theta_c) = {\theta_c}^{gan} + \lambda_1 {\theta_c}^{id} + \lambda_2 {\theta_c}^{var} \tag{4.16}$$

Capítulo 5

Resultados e Discussões

Este capítulo apresentamos as configurações do conjunto de dados dos experimentos, os detalhes de implementação e a eficácia do método proposto por meio dos resultados experimentais e comparação com técnicas do estado-da-arte.

5.1 Descrição das bases de dados

Para os testes do método proposto foram utilizado de cindo bases de dados entre as mais utilizadas na revisão de literatura. Sendo uma delas em ambiente não controlado, a LFW, que é a mais desafiadora e mais utilizada entre as bases de dados não controladas "*Wild*", todas as bases de dados e os protocolos utilizados são:

- **AR** [77] que consiste em 126 indivíduos, contendo 26 variações entre expressão, iluminação e oclusão diferentes por indivíduo. Deste conjunto de dados, foi utilizado um subconjunto com 100 identidades. Das quias foram escolhidos aleatoriamente 50 identidades para o conjunto de treinamento e 50 para o conjunto de teste.

- **Extend Yale B (E-YaleB)** [78] consiste em 38 indivíduos sob uma ampla gama de condições de iluminação, desde condições de iluminação leves até severas. Devido ao baixo número de indivíduos, como feito por PANG *et al.* [8], introduzimos o subconjunto de iluminação da base dados AR na E-YaleB para ampliar o número de indivíduos. Foram escolhidos aleatoriamente 100

indivíduos do conjunto de dados misto para o conjunto de treinamento e os outros 38 indivíduos para o conjunto de teste.

- **FERET** [79] consiste em 1.199 indivíduos com uma variedade de gênero, idade e etnia. Deste conjunto de dados utilizamos um subconjunto de 200 indivíduos contendo apenas quatro variações de pose. Foram escolhidos aleatoriamente 150 indivíduos para o conjunto de treinamento e os outros 50 para o conjunto de teste.

- **CAS-PEAL** [80] consiste em 1.040 indivíduos com variações como poses, oclusões e idades. Deste conjunto de dados, foram utilizados um subconjunto com 300 indivíduos das categorias normais e acessórios, com uma imagem neutra e outras seis usando diferentes acessórios, óculos e chapéus. Foram escolhidos aleatoriamente 200 indivíduos para o conjunto de treinamento e os outros 100 para o conjunto de teste.

- **LFW** [81] consiste em 5.749 indivíduos coletados em um ambiente não controlado, com uma ampla variedade de expressões, poses, iluminações e outras variações. Desta base de dados foram utilizado um subconjunto de 158 indivíduos com mais de dez imagens por indivíduo da versão alinhada de LFW, a LFW-a. Para avaliação, foram escolhidos 50 indivíduos contendo imagens neutras para o conjunto de teste e os outros 108 para o conjunto de treinamento.

De acordo com [82], os conjuntos de dados AR e LFW exibem variações faciais mais complexas em comparação com outros, tornando-os particularmente desafiadores para o SSPP FR. Aplicamos algumas etapas de pré-processamento nos dados para garantir a consistência em todos os conjuntos de dados:

- **Redimensionamento de imagem**: Todas as imagens são redimensionadas para 64x64 pixels para corresponder às dimensões de entrada da rede.

- **Normalização**: Os valores dos pixels são normalizados para o intervalo [0, 1] para melhorar a convergência durante o treinamento.

- **Alinhamento**: Para o conjunto de dados LFW, usamos a versão alinhada (LFW-a) para reduzir variações causadas por desalinhamentos.

- **Tratamento de valores ausentes**: Todos os conjuntos de dados usados neste estudo contêm dados completos, sem valores ausentes, eliminando a necessidade de imputação de dados.

A definição da resolução das imagens de entrada em 64x64 pixels visa equilibrar a fidelidade da representação facial com a eficiência computacional. Assim, resolução preserva as características macroscópicas essenciais para a distinção de identidades, como a estrutura geral do rosto e a disposição dos seus principais componentes. A adoção de uma dimensão reduzida melhora o custo computacional inerente ao processamento de imagens, viabilizando um treinamento mais ágil e com menor consumo de recursos de memória.

As redes processam os dados pré-processados da seguinte forma:

- **Geração de código latente**: O codificador (Enc) gera a média (μ) e a variância (σ) da distribuição do espaço latente. O código latente (c) é então amostrado a partir dessa distribuição usando o truque de reparametrização, o que garante que c seja diferenciável em relação aos parâmetros da rede. Essa diferenciabilidade permite otimização baseada em gradientes durante o treinamento. Um vetor de ruído (z) é amostrado independentemente de uma distribuição Gaussiana para modelagem de variação.

- **Concatenação de características**: O gerador (Gen) combina c e z em um único vetor de entrada (c, z) para criar protótipos que preservam a identidade com variações controladas.

- **Dimensionalidade da representação**: Para todos os conjuntos de dados, a dimensão latente (Ldim) é definida como 100, garantindo uma representação consistente das características entre os conjuntos de dados.

5.2 Detalhes de implementação

A Tabela 5.1 e a Tabela 5.2 mostram a arquitetura das redes do AD-VAE. Na Tabela 5.1, pode-se observar que a estrutura das redes $\mathbf{E}_{nc}$ e $\mathbf{D}_{is}$ são semelhantes, com uma diferença na última camada. Após a penúltima camada, que redimensiona a imagem por meio da camada *Flatten*, o $\mathbf{E}_{nc}$ usa o resultado achatado em duas

Codificador $\mathbf{E}_{nc}$ and	**Descriminador $\mathbf{D}_{is}$**	
Layer	input/ output	Filter / Stride / Pad
Conv2d-1	3 / 64	4 x 4 / 2 /1
BatchNorm2d	64 / 64	
LeakyReLU		
Conv2d-2	64 / 128	4 x 4 / 2 /1
BatchNorm2d	128 / 128	
LeakyReLU		
Conv2d-3	128 / 256	4 x 4 / 2 /1
BatchNorm2d	256 / 256	
LeakyReLU		
Conv2d-2	256 / 512	4 x 4 / 2 /1
BatchNorm2d	128 / 128	
LeakyReLU		
$\mathbf{E}_{nc}$ Camadas finais	**Flatten**	
Fullconected-μ	output = L_{dim}	**FullConected-σ**
$\mathbf{D}_{is}$ Camadas finais	**Flatten**	
Fullconected	output = $N_{dim} + 2$	

Tabela 5.1: Estrutura das redes $\mathbf{E}_{nc}$ e $\mathbf{D}_{is}$.

camadas totalmente conectadas, uma camada para gerar a média μ e outra para a variância σ^2. A dimensão de saída dessas duas camadas é L_{dim}, que é uma dimensão do código latente c e do vetor de ruído z usados pelo AD-VAE. Para o $\mathbf{D}_{is}$, a última camada é uma camada totalmente conectada com dimensão de saída $N_{dim}+2$, onde N_{dim} é o número de identidades (${\mathbf{D}_{is}}^{id}$) do conjunto de treinamento. As outras duas posições (+2) são usadas para distinguir a ocorrência de variação (${\mathbf{D}_{is}}^{var}$) e se a imagem de entrada é real ou falsa (${\mathbf{D}_{is}}^{gan}$).

A arquitetura de rede adotada neste trabalho é fundamentada no modelo DCGAN, proposto por Radford et al. [83]. Esta é uma escolha metodológica estratégica que, ao empregar uma base de eficácia comprovada, assegura a robustez e a estabilidade do treinamento. Tal abordagem permite que a análise e a validação experimental se concentrem exclusivamente nas contribuições centrais deste trabalho.

Assim como as duas redes mencionadas anteriormente, as redes $\mathbf{G}_{en}$ e $\mathbf{D}_{ec}$ também são semelhantes, como mostrado na Tabela 5.2. Neste caso, a diferença está no

Gerador $\mathbf{G}_{en}$ and	**Decodificador $\mathbf{D}_{ec}$**	
Layer	input/ output	Filter / Stride / Pad
Generator $\mathbf{G}_{en}$	First Layer	
Fullconected	$L_{dim} * 2$ / 8192	
Decoder $\mathbf{D}_{ec}$	First Layer	
Fullconected	L_{dim} / 8192	
Reshape **BatchNorm2d** **ReLU**	8192 / (512 x 4 x 4) n 512 / 512	
ConvTranspose2d-1 **BatchNorm2d** **ReLU**	512 / 256 256 / 256	4 x 4 / 2 /1
ConvTranspose2d-2 **BatchNorm2d** **ReLU**	256 / 128 128 / 128	4 x 4 / 2 /1
ConvTranspose2d-3 **BatchNorm2d** **ReLU**	128 / 64 64 / 64	4 x 4 / 2 /1
ConvTranspose2d-1 **Tanh**	64 / 3	4 x 4 / 2 /1

Tabela 5.2: Estrutura das redes $\mathbf{D}_{ec}$ e $\mathbf{G}_{en}$.

Base de dados	Identidade de treinamento	Identidade de teste	N_{dim}	L_{dim}
E-YaleB&AR	100	38	100	100
CAS-PEAL	200	100	200	100
FERET	150	50	150	100
AR	50	50	50	100
LFW	108	50	108	100

Tabela 5.3: Partição do conjunto de dados e configuração de parâmetros

topo da rede. A primeira camada de $\mathbf{G}_{en}$ usa uma entrada de dimensão $L_{dim} * 2$, pois a entrada é uma concatenação do código latente c e do vetor de ruído z, ambos com dimensão igual a L_{dim}, sendo $input = concat(\mathbf{c}_{Ldim}, \mathbf{z}_{Ldim})$. Para o $\mathbf{D}_{ec}$, a primeira camada usa uma entrada de dimensão L_{dim}, pois apenas o código latente c é usado como entrada para a camada totalmente conectada.

As imagens são inicialmente pré-processadas, redimensionando-as para 64x64 pixeis. O treinamento usa descida de gradiente estocástica (*Stochastic Gradient Descent* (SGD)) em mini-lotes (*Mini-Batch*) com um tamanho de mini-lote de 16. Os pesos de todas as camadas são inicializados a partir de uma distribuição gaussiana centrada em zero e com desvio padrão de 0,002. O otimizador utilizado é o Otimizador Adam [83] com betas (0,5, 0,999) e taxas de aprendizado de 0,0002, 0,0002, 0,0001 e 0,0003 para $\mathbf{E}_{nc}$, $\mathbf{D}_{ec}$, $\mathbf{G}_{en}$ e $\mathbf{D}_{is}$, respectivamente.

Para todos os conjuntos de dados, foram definidos a dimensão de L_{dim} como 100, e os parâmetros μ_1, μ_2, μ_3 na Eq. 4.3 e λ_1, λ_2 na Eq. 4.9 como 2,0, 0,5, 0,1, 2,0 e 0,5, respectivamente, conforme usado por [8]. A Tabela 5.3 detalha todas as configurações de parâmetros e a partição do conjunto de dados.

Cada experimento foi repetido cinco vezes com diferentes inicializações aleatórias. As médias e os desvios padrão apresentados nas tabelas de resultados foram calculados com base nessas execuções, permitindo uma avaliação mais robusta da estabilidade e desempenho do método proposto.

Figura 5.1: Os protótipos gerados pelo AD-VAE, (a) é uma amostra de imagem com variações, (b) é o protótipo da imagem (a), e (c) é o protótipo real de (a). À direita, temos o nome do conjunto de dados e a descrição da variação da face.

5.3 Avaliação em Reconhecimento de faces com uma única amostra por pessoa

Esta seção descreve os resultados da avaliação do AD-VAE na tarefa de reconhecer faces que tem apenas uma imagem de referência em expressão, iluminação e pose próxima ao neutro. Inicialmente, o experimento foi realizado em conjuntos de dados controlados, AR, CAS-PEAL, FERET e E-YaleB, e por último base de dados não controlada "*Wild*"LFW. Definimos os métodos e configurações das bases de dados como em [8]. A Figura 5.1 mostra exemplos de imagens das bases de dados, mostrando a imagem de teste $\mathbf{x}$, o protótipo da imagem de teste gerado pelo AD-VAE $\hat{\mathbf{x}}$ e o protótipo real da imagem de teste $\mathbf{x}_{rp}$.

Para comparação em bases de dados controladas, foram utilizados nove métodos da literatura: PCA [84], VAE [71], SRC [85], CRC [86], PCRC [87], DMMA [62], SVDL [59], SLRC [16] e S^3RC[31]. Para os métodos que utilizavam de uma base genérica para aprender as variações que uma face poderia ter, foram utilizado o

subconjunto de treinamento como base de dados genéricas para esses casos.

Para DMMA e PCRC, que realiza a divisão das imagem em pequenos pedaços, *patch*, eles não são sobreposto e tem o tamanho de 16x16 pixeis. Os outros parâmetros de DMMA são definidos como $K_1 = 30, K_2 = 2, K = 2$ e $\sigma = 10$. O parâmetro de regularização λ de SRC, CRC, SLRC e S^{3}RCé fixado em 0,01. Para SVDL, os parâmetros são definidos como $\lambda_1 = 0,001, \lambda_2 = 0,01$ e $\lambda_3 = 0,0001$. A medida de similaridade de protótipos gerados em PCA e VAE é a métrica de distância do cosseno e o KNN com k=1 é usado como classificador. Para VD-GAN e AD-VAE, é gerado um protótipo da imagem de consulta y e um protótipo de cada $\mathbf{x}_{rp}$. Em seguida é utilizado o KNN com distância do cosseno para classificar os protótipos gerados do sujeito contra os protótipos gerados de toda a galeria. Considerando $P(x)$ como um protótipo gerado a partir de x na Eq. 5.1, os protótipos gerados são classificados pela Eq. 5.2.

$$P(x) = \mathbf{G}_{en}(f(\mathbf{E}_{nc}(x)), z) \tag{5.1}$$

$$ID(\mathbf{y}) = \arg\min_{k} dist(P(y), P(\mathbf{x}_{rp})) \tag{5.2}$$

A Tabela 5.4 apresenta a acurácia de reconhecimento nas quatro bases de dados controladas. Os resultados mostram que o AD-VAE supera todos os outros métodos nas quatro bases de dados. O método proposto alcança taxas de acurácia superiores às do VD-GAN, que por sua vez supera os métodos de aprendizado de dados genéricos. Assim como o VD-GAN, O método proposto apresenta um desempenho superior em relação aos métodos baseados em modelos de superposição linear quando se trata de variações não lineares, tais como a variação de pose na base de dados FERET. O AD-VAE supera os outros devido à aprendizagem de variação desemaranhada durante o treinamento do codificador e gerador, resultando em um vetor latente mais representativo de cada identidade.

Mesmo com semelhanças, o treinamento do codificador pelo VAE, a arquitetura das redes neurais e a função de perda adicional $\mathcal{L}_{\mathcal{C}}$ conseguiram superar o VD-GAN.

Métodos	AR	E-YaleB&AR	CA-PEAL	FERET
PCA	42.4 ± 2.2	58.5 ± 2.3	51.3 ± 1.0	40.5 ± 31
VAE	44.9 ± 1.1	59.9 ± 1.1	51.4 ± 0.9	55.0 ± 2.3
SRC	49.6 ± 2.4	64.0 ± 3.8	62.3 ± 1.4	51.5 ± 2.6
CRC	50.8 ± 4.8	63.5 ± 1.4	69.5 ± 2.7	43.0 ± 4.1
DMMA	51.9 ± 1.9	55.4 ± 1.1	59.2 ± 0.6	57.5 ± 1.2
PCRC	74.1 ± 3.7	80.7 ± 5.4	75.8 ± 0.6	24.0 ± 2.4
SVDL	76.0 ± 0.8	$88.1 \pm 1,8$	78.7 ± 1.2	67.0 ± 1.7
SLRC	76.6 ± 1.8	88.8 ± 2.6	78.2 ± 3.3	68.0 ± 3.8
S^3RC	77.8 ± 2.6	88.2 ± 1.5	80.3 ± 3.3	73.0 ± 2.1
VD-GAN	79.7 ± 0.8	90.6 ± 2.5	81.2 ± 2.2	90.5 ± 0.8
AD-VAE	$84.9 \pm 1.5\%$	$94.6 \pm 1.8\%$	$94.5 \pm 1.6\%$	96.0 ± 1.0 %

Tabela 5.4: Acurácia de reconhecimento (%) e o desvios padrão de diferentes métodos nas bases de dados E-YaleB&AR, CAS-PEAL, AR e FERET para SSPP FR.

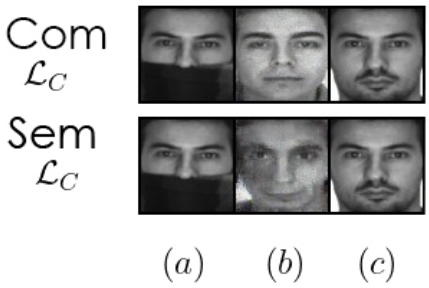

Figura 5.2: Os protótipos gerados pelo AD-VAE com e sem a função de perda $\mathcal{L}_\mathcal{C}$, (a) é a imagem de amostra com variações, (b) é o protótipo da imagem (a), e (c) é o protótipo real de (a).

De acordo com os autores do VD-GAN, o VAE não era competitivo por ser um método não supervisionado, entretanto, o AD-VAE utiliza as vantagens do VAE de forma supervisionada, alcançando taxas melhores que o VD-GAN. A Figura 5.2 mostra o protótipo usando a $\mathcal{L}_\mathcal{C}$ e o protótipo sem a $\mathcal{L}_\mathcal{C}$.

Para testar em um conjunto de dados não controlado, avaliamos o AD-VAE comparando-o com o VD-GAN$_{Lcnn}$ (versão do VD-GAN que utiliza o LightCNN-29 como codificador na rede geradora) e quatro métodos recentes baseados em redes neurais profundas: JCR-ACF [12], Regular-face [22], Arc-face [88] e CJR-RACF [18]. A Tabela 5.5 apresenta as taxas de reconhecimento de rank-1 de todos os métodos para SSPP FR, mostrando que o AD-VAE supera todos os outros métodos, incluindo o VD-GAN com as modificações no codificador.

Métodos	Taxa de reconhecimento(%)
JCR-ACF	86,0%
Regular-face	83,7%
Arc-face	92,3%
CJR-RACF	95,5%
VD-GAN$_{Lcnn}$	98,4%
AD-VAE	$99.6 \pm 1.2\%$

Tabela 5.5: Taxas de Reconhecimento (%) de diferentes métodos baseados em aprendizado profundo no conjunto de dados LFW para SSPP FR

Capítulo 6

Conclusões

Neste trabalho, propusemos o arcabouço AD-VAE, que, até onde sabemos, é o primeiro a utilizar o VAE no problema de Reconhecimento de Faces com uma Amostra por Pessoa (SSPP FR). O AD-VAE aprende a construir protótipos representativos de identidade a partir tanto de conjuntos de dados controlados quanto de conjuntos de dados não controlados, ou "wild", de SSPP FR. A versão padrão do AD-VAE, sem a adição de redes neurais pré-treinadas em substituição a qualquer uma de suas redes profundas, superou todas as técnicas de SSPP FR testadas, demonstrando a robustez do método proposto. O AD-VAE lida efetivamente com grandes variações, como pose (FERET), iluminação e oclusão (AR), condições de iluminação complexas (EYaleB) e variações mistas do conjunto de dados selvagem LFW. Esses resultados confirmam que a combinação de VAE com GAN é uma abordagem promissora para o reconhecimento de faces com apenas uma imagem por indivíduo.

Além disso, o AD-VAE demonstrou ser uma ferramenta robusta, capaz de lidar com variações críticas, como as encontradas nos conjuntos de dados FERET, AR, EYaleB e LFW. Isso posiciona o AD-VAE como uma forte candidata para aplicações práticas de reconhecimento facial em cenários do mundo real. O uso de autoencoders variacionais, além disso, abre novas possibilidades para separar características de identidade e variação nas tarefas de SSPP FR, o que é fundamental em aplicações biométricas. O framework alcançou altas taxas de reconhecimento sem depender de redes externas pré-treinadas, oferecendo uma solução eficiente e econômica para SSPP FR.

Os experimentos demonstraram que o AD-VAE supera os métodos do estado da

arte com diferenças estatisticamente relevantes nas taxas de acurácia, alcançando 99.6% no conjunto LFW, mesmo sem uso de redes pré-treinadas.

Respondendo assim à hipótese de pesquisa levantada de que "a combinação de um codificador de Variational Autoencoder (que utiliza distribuições de probabilidade para geração de imagens) com uma GAN que incorpora o desentrelaçamento de características (similar ao VD-GAN) pode gerar imagens de protótipos em condições neutras (pose frontal, expressão séria, iluminação regular e ausência de oclusões) para o reconhecimento facial com uma única amostra por pessoa"o que é comprovada pelos resultados obitidos

6.1 Trabalhos Futuros

Para aprimorar ainda mais a capacidade de aprendizado do AD-VAE, sugerem-se várias direções para trabalhos futuros. A primeira possível melhoria seria a incorporação de novas arquiteturas de rede. Uma abordagem promissora seria o uso de espaço latente intermediário combinado com a normalização de instância adaptativa, como implementado no StyleGAN2 [19]. Outra melhoria seria adotar uma etapa separada de treinamento da parte VAE, como feito no modelo ID-GAN [20], o que permitiria um controle mais refinado sobre a geração de faces e melhoraria a precisão do reconhecimento.

Além disso, futuros trabalhos poderiam explorar o uso de redes neurais pré-treinadas, como já feito por outros autores, ou até mesmo utilizar as técnicas emergentes para geração de imagens baseadas em modelos avançados, como o DALL-E [89], CLIP [90], e o aclamado GPT-3 [91]. A adoção dessas abordagens pode resultar em uma melhoria substancial nas capacidades de generalização e na qualidade das representações de identidade e variação. Outro aprimoramento possível seria a aplicação de modelos de difusão, como o utilizado no Dual Condition Face Generator (DCFace) [92], para métodos de transferência de estilo. O uso de datasets maiores, como o GAN-Control [93], também poderia levar a um aumento no desempenho do AD-VAE, ampliando sua aplicabilidade e robustez. Essas direções representam um vasto campo para inovação e aprimoramento do reconhecimento facial com uma amostra por pessoa, tornando o AD-VAE uma ferramenta cada vez mais eficaz para

desafios práticos no reconhecimento de identidade.

Referências Bibliográficas

[1] LAHASAN, B., LUTFI, S. L., SAN-SEGUNDO, R. "A survey on techniques to handle face recognition challenges: occlusion, single sample per subject and expression", *Artificial Intelligence Review*, set. 2017. ISSN: 0269-2821, 1573-7462. doi: 10.1007/s10462-017-9578-y. Disponível em: <http://link.springer.com/10.1007/s10462-017-9578-y>.

[2] ZHAO, W., CHELLAPPA, R., PHILLIPS, P. J., etãl. "Face recognition: A literature survey", *ACM Computing Surveys*, v. 35, n. 4, pp. 399–458, dez. 2003. ISSN: 03600300. doi: 10.1145/954339.954342. Disponível em: <http://portal.acm.org/citation.cfm?doid=954339.954342>.

[3] TAN, X., CHEN, S., ZHOU, Z.-H., etãl. "Face recognition from a single image per person: A survey", *Pattern Recognition*, v. 39, n. 9, pp. 1725 – 1745, 2006. ISSN: 0031-3203. doi: https://doi.org/10.1016/j.patcog.2006.03.013. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0031320306001270>.

[4] CHU, Y., ZHAO, L., AHMAD, T. "Multiple feature subspaces analysis for single sample per person face recognition", *The Visual Computer*, jan. 2018. ISSN: 0178-2789, 1432-2315. doi: 10.1007/s00371-017-1468-4. Disponível em: <http://link.springer.com/10.1007/s00371-017-1468-4>.

[5] MINAEE, S., ABDOLRASHIDI, A., SU, H., etãl. "Biometrics recognition using deep learning: a survey", *Artificial Intelligence Review*, v. 56, n. 8, pp. 8647–8695, jan 2023. ISSN: 1573-7462. doi: 10.1007/s10462-022-10237-x. Disponível em: <http://dx.doi.org/10.1007/s10462-022-10237-x>.

[6] YING LI, WEI SHEN, XUN SHI, etãl. "Ensemble of Randomized Linear Discriminant Analysis for face recognition with single sample per person". In: *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, pp. 1–8. IEEE, abr. 2013. ISBN: 978-1-4673-5546-9 978-1-4673-5545-2 978-1-4673-5544-5. doi:

10.1109/FG.2013.6553755. Disponível em: <http://ieeexplore.ieee.org/document/6553755/>.

[7] TRAN, L., YIN, X., LIU, X. "Disentangled Representation Learning GAN for Pose-Invariant Face Recognition". In: *In Proceeding of IEEE Computer Vision and Pattern Recognition*, Honolulu, HI, July 2017.

[8] PANG, M., WANG, B., CHEUNG, Y.-M., etãl. "VD-GAN: A Unified Framework for Joint Prototype and Representation Learning From Contaminated Single Sample per Person", *IEEE Transactions on Information Forensics and Security*, v. 16, pp. 2246–2259, 2021. doi: 10.1109/TIFS.2021.3050055.

[9] PLUMERAULT, A., BORGNE, H. L., HUDELOT, C. "AVAE: Adversarial Variational Auto Encoder". 2020. Disponível em: <https://arxiv.org/abs/2012.11551>.

[10] OH, B.-S., TOH, K.-A., TEOH, A. B. J., etãl. "An Analytic Gabor Feedforward Network for Single-Sample and Pose-Invariant Face Recognition", *IEEE Transactions on Image Processing*, v. 27, n. 6, pp. 2791–2805, jun. 2018. ISSN: 1057-7149, 1941-0042. doi: 10.1109/TIP.2018.2809040. Disponível em: <http://ieeexplore.ieee.org/document/8301521/>.

[11] KAN, M., SHAN, S., SU, Y., etãl. "Adaptive discriminant learning for face recognition", *Pattern Recognition*, v. 46, n. 9, pp. 2497–2509, set. 2013. ISSN: 00313203. doi: 10.1016/j.patcog.2013.01.037. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0031320313000769>.

[12] YANG, M., WANG, X., ZENG, G., etãl. "Joint and collaborative representation with local adaptive convolution feature for face recognition with single sample per person", *Pattern Recognition*, v. 66, pp. 117–128, jun. 2017. ISSN: 00313203. doi: 10.1016/j.patcog.2016.12.028. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0031320316304496>.

[13] PANG, M., CHEUNG, Y.-M., WANG, B., etãl. "Robust Heterogeneous Discriminative Analysis for Single Sample Per Person Face Recognition". In: *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pp. 2251–2254. ACM Press, 2017. ISBN: 978-1-4503-4918-5. doi: 10.1145/3132847.3133096. Disponível em: <http://dl.acm.org/citation.cfm?doid=3132847.3133096>.

[14] SONG, T., WANG, X., YANG, M., etãl. "Triple local feature based collaborative representation for face recognition with single sample per person". In: *2016 IEEE International Conference on Image Processing (ICIP)*, pp. 3234–3238. IEEE, set. 2016. ISBN: 978-1-4673-9961-6. doi: 10.1109/ICIP.2016.7532957. Disponível em: <http://ieeexplore.ieee.org/document/7532957/>.

[15] LIU, Y., WASSELL, I. J. "A New Face Recognition Algorithm based on Dictionary Learning for a Single Training Sample per Person." mar. 2018. Disponível em: <https://www.repository.cam.ac.uk/handle/1810/274243>.

[16] DENG, W., HU, J., WU, Z., etãl. "From one to many: Pose-Aware Metric Learning for single-sample face recognition", *Pattern Recognition*, v. 77, pp. 426–437, maio 2018. ISSN: 00313203. doi: 10.1016/j.patcog.2017.10.020. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0031320317304259>.

[17] PANG, M., WANG, B., YE, M., etãl. "DisP+V: A Unified Framework for Disentangling Prototype and Variation From Single Sample per Person", *IEEE Transactions on Neural Networks and Learning Systems*, v. 34, n. 2, pp. 867–881, 2023. doi: 10.1109/TNNLS.2021.3103194.

[18] YANG, M., WEN, W., WANG, X., etãl. "Adaptive Convolution Local and Global Learning for Class-Level Joint Representation of Facial Recognition With a Single Sample Per Data Subject", *IEEE Transactions on Information Forensics and Security*, v. 15, pp. 2469–2484, 2020. doi: 10.1109/TIFS.2020.2965301.

[19] KARRAS, T., AITTALA, M., HELLSTEN, J., etãl. "Training Generative Adversarial Networks with Limited Data". In: *Proc. NeurIPS*, 2020.

[20] LEE, W., KIM, D., HONG, S., etãl. "High-Fidelity Synthesis with Disentangled Representation". 2020. Disponível em: <https://arxiv.org/abs/2001.04296>.

[21] PANG, M., CHEUNG, Y.-M., WANG, B., etãl. "Robust heterogeneous discriminative analysis for face recognition with single sample per person", *Pattern Recognition*, v. 89, pp. 91–107, 2019.

[22] ZHAO, K., XU, J., CHENG, M.-M. "RegularFace: Deep Face Recognition via Exclusive Regularization". In: *2019 IEEE/CVF Conference on Computer*

Vision and Pattern Recognition (CVPR), pp. 1136–1144, 2019. doi: 10.1109/CVPR.2019.00123.

[23] ZHOU, D., YANG, D., ZHANG, X., etãl. "Discriminative Probabilistic Latent Semantic Analysis with Application to Single Sample Face Recognition", *Neural Processing Letters*, jun. 2018. ISSN: 1370-4621, 1573-773X. doi: 10.1007/s11063-018-9852-2. Disponível em: <http://link.springer.com/10.1007/s11063-018-9852-2>.

[24] ZHANG, W., XU, Z., WANG, Y., etãl. "Binarized features with discriminant manifold filters for robust single-sample face recognition", *Signal Processing: Image Communication*, v. 65, pp. 1–10, jul. 2018. ISSN: 09235965. doi: 10.1016/j.image.2018.03.003. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0923596518302157>.

[25] YU, Y.-F., DAI, D.-Q., REN, C.-X., etãl. "Discriminative multi-scale sparse coding for single-sample face recognition with occlusion", *Pattern Recognition*, v. 66, pp. 302–312, jun. 2017. ISSN: 00313203. doi: 10.1016/j.patcog.2017.01.021. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0031320317300225>.

[26] PARCHAMI, M., BASHBAGHI, S., GRANGER, E. "CNNs with cross-correlation matching for face recognition in video surveillance using a single training sample per person". In: *2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pp. 1–6. IEEE, ago. 2017. ISBN: 978-1-5386-2939-0. doi: 10.1109/AVSS.2017.8078554. Disponível em: <http://ieeexplore.ieee.org/document/8078554/>.

[27] LIONG, V. E., HAIBIN YAN. "Domain transfer sparse representation for single sample face recognition". In: *2017 IEEE International Conference on Multimedia e Expo Workshops (ICMEW)*, pp. 663–668. IEEE, jul. 2017. ISBN: 978-1-5386-0560-8. doi: 10.1109/ICMEW.2017.8026289. Disponível em: <http://ieeexplore.ieee.org/document/8026289/>.

[28] JI, H.-K., SUN, Q.-S., JI, Z.-X., etãl. "Collaborative probabilistic labels for face recognition from single sample per person", *Pattern Recognition*, v. 62, pp. 125–134, fev. 2017. ISSN: 00313203. doi: 10.1016/j.patcog.2016.08.007. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0031320316302205>.

[29] HUANG, K.-K., DAI, D.-Q., REN, C.-X., etãl. "Learning Kernel Extended Dictionary for Face Recognition", *IEEE Transactions on Neural Networks*

and Learning Systems, v. 28, n. 5, pp. 1082–1094, maio 2017. ISSN: 2162-237X, 2162-2388. doi: 10.1109/TNNLS.2016.2522431. Disponível em: <http://ieeexplore.ieee.org/document/7407377/>.

[30] HU, J. "Discriminative transfer learning with sparsity regularization for single-sample face recognition", *Image and Vision Computing*, v. 60, pp. 48–57, abr. 2017. ISSN: 02628856. doi: 10.1016/j.imavis.2016.08.007. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0262885616301317>.

[31] GAO, Y., MA, J., YUILLE, A. L. "Semi-Supervised Sparse Representation Based Classification for Face Recognition With Insufficient Labeled Samples", *IEEE Transactions on Image Processing*, v. 26, n. 5, pp. 2545–2560, maio 2017. ISSN: 1057-7149, 1941-0042. doi: 10.1109/TIP.2017.2675341. Disponível em: <http://ieeexplore.ieee.org/document/7865903/>.

[32] DENG, W., HU, J., GUO, J. "Face Recognition via Collaborative Representation: Its Discriminant Nature and Superposed Representation", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1, 2017. ISSN: 0162-8828, 2160-9292. doi: 10.1109/TPAMI.2017.2757923. Disponível em: <http://ieeexplore.ieee.org/document/8053795/>.

[33] ZHU, J.-Y., ZHENG, W.-S., LU, F., etãl. "Illumination invariant single face image recognition under heterogeneous lighting condition", *Pattern Recognition*, v. 66, pp. 313–327, jun. 2017. ISSN: 00313203. doi: 10.1016/j.patcog.2016.12.029. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0031320316304502>.

[34] PEI, T., ZHANG, L., WANG, B., etãl. "Decision pyramid classifier for face recognition under complex variations using single sample per person", *Pattern Recognition*, v. 64, pp. 305–313, abr. 2017. ISSN: 00313203. doi: 10.1016/j.patcog.2016.11.016. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0031320316303740>.

[35] HU, X., PENG, S., WANG, L., etãl. "Surveillance video face recognition with single sample per person based on 3D modeling and blurring", *Neurocomputing*, v. 235, pp. 46–58, abr. 2017. ISSN: 09252312. doi: 10.1016/j.neucom.2016.12.059. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0925231217300012>.

[36] HONG, S., IM, W., RYU, J., etãl. "SSPP-DAN: Deep Domain Adaptation Network for Face Recognition with Single Sample Per Person", *ar-*

Xiv:1702.04069 [cs], fev. 2017. Disponível em: <http://arxiv.org/abs/1702.04069>. arXiv: 1702.04069.

[37] GUO, Y., JIAO, L., WANG, S., etãl. "Fuzzy Sparse Autoencoder Framework for Single Image Per Person Face Recognition", *IEEE Transactions on Cybernetics*, pp. 1–14, 2017. ISSN: 2168-2267, 2168-2275. doi: 10.1109/TCYB.2017.2739338. Disponível em: <http://ieeexplore.ieee.org/document/8017477/>.

[38] CHU, Y., AHMAD, T., BEBIS, G., etãl. "Low-resolution face recognition with single sample per person", *Signal Processing*, v. 141, pp. 144–157, dez. 2017. ISSN: 01651684. doi: 10.1016/j.sigpro.2017.05.012. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0165168417301834>.

[39] HAGHIGHAT, M., ABDEL-MOTTALEB, M., ALHALABI, W. "Fully automatic face normalization and single sample face recognition in unconstrained environments", *Expert Systems with Applications*, v. 47, pp. 23–34, abr. 2016. ISSN: 09574174. doi: 10.1016/j.eswa.2015.10.047. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0957417415007514>.

[40] GU, J., HU, H., LI, H., etãl. "Patch-based alignment-free generic sparse representation for pose-robust face recognition". In: *2016 IEEE International Conference on Image Processing (ICIP)*, pp. 3006–3010. IEEE, set. 2016. ISBN: 978-1-4673-9961-6. doi: 10.1109/ICIP.2016.7532911. Disponível em: <http://ieeexplore.ieee.org/document/7532911/>.

[41] LIU, F., TANG, J., SONG, Y., etãl. "Local structure based multi-phase collaborative representation for face recognition with single sample per person", *Information Sciences*, v. 346-347, pp. 198–215, jun. 2016. ISSN: 00200255. doi: 10.1016/j.ins.2016.02.001. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0020025516300433>.

[42] LIU, F., TANG, J., SONG, Y., etãl. "A multi-phase sparse probability framework via entropy minimization for single sample face recognition". In: *2016 IEEE International Conference on Image Processing (ICIP)*, pp. 3887–3891. IEEE, set. 2016. ISBN: 978-1-4673-9961-6. doi: 10.1109/ICIP.2016.7533088. Disponível em: <http://ieeexplore.ieee.org/document/7533088/>.

[43] ZHUANG, L., CHAN, T.-H., YANG, A. Y., etãl. "Sparse Illumination Learning and Transfer for Single-Sample Face Recognition with Image Corruption

and Misalignment", *International Journal of Computer Vision*, v. 114, n. 2-3, pp. 272–287, set. 2015. ISSN: 0920-5691, 1573-1405. doi: 10.1007/s11263-014-0749-x. Disponível em: <http://link.springer.com/10.1007/s11263-014-0749-x>.

[44] HU, J., LU, J., ZHOU, X., etãl. "Discriminative transfer learning for single-sample face recognition". In: *IAPR International Conference on Biometrics (ICB)*, pp. 272–277. IEEE, maio 2015. ISBN: 978-1-4799-7824-3. doi: 10.1109/ICB.2015.7139095. Disponível em: <http://ieeexplore.ieee.org/document/7139095/>.

[45] GAO, S., ZHANG, Y., JIA, K., etãl. "Single Sample Face Recognition via Learning Deep Supervised Autoencoders", *IEEE Transactions on Information Forensics and Security*, v. 10, n. 10, pp. 2108–2118, out. 2015. ISSN: 1556-6013, 1556-6021. doi: 10.1109/TIFS.2015.2446438. Disponível em: <http://ieeexplore.ieee.org/document/7124463/>.

[46] GAO, S., JIA, K., ZHUANG, L., etãl. "Neither Global Nor Local: Regularized Patch-Based Representation for Single Sample Per Person Face Recognition", *International Journal of Computer Vision*, v. 111, n. 3, pp. 365–383, fev. 2015. ISSN: 0920-5691, 1573-1405. doi: 10.1007/s11263-014-0750-4. Disponível em: <http://link.springer.com/10.1007/s11263-014-0750-4>.

[47] DING, R.-X., DU, D. K., HUANG, Z.-H., etãl. "Variational Feature Representation-based Classification for face recognition with single sample per person", *Journal of Visual Communication and Image Representation*, v. 30, pp. 35–45, jul. 2015. ISSN: 10473203. doi: 10.1016/j.jvcir.2015.03.001. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S1047320315000450>.

[48] CAI, J., CHEN, J., LIANG, X. "Single-Sample Face Recognition Based on Intra-Class Differences in a Variation Model", *Sensors*, v. 15, n. 1, pp. 1071–1087, jan. 2015. ISSN: 1424-8220. doi: 10.3390/s150101071. Disponível em: <http://www.mdpi.com/1424-8220/15/1/1071>.

[49] BASHBAGHI, S., GRANGER, E., SABOURIN, R., etãl. "Ensembles of exemplar-SVMs for video face recognition from a single sample per person". In: *2015 12th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pp. 1–6. IEEE, ago. 2015. ISBN: 978-1-4673-7632-7. doi: 10.1109/AVSS.2015.7301749. Dispo-

nível em: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7301749>.

[50] HU, C., YE, M., JI, S., etãl. "A new face recognition method based on image decomposition for single sample per person problem", *Neurocomputing*, v. 160, pp. 287–299, jul. 2015. ISSN: 09252312. doi: 10.1016/j.neucom.2015.02.032. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0925231215001812>.

[51] JUEFEI-XU, F., LUU, K., SAVVIDES, M. "Spartans: Single-Sample Periocular-Based Alignment-Robust Recognition Technique Applied to Non-Frontal Scenarios", *IEEE Transactions on Image Processing*, v. 24, n. 12, pp. 4780–4795, dez. 2015. ISSN: 1057-7149, 1941-0042. doi: 10.1109/TIP.2015.2468173. Disponível em: <http://ieeexplore.ieee.org/document/7194796/>.

[52] MACHADO, A. M. C. "The 2D factor analysis and its application to face recognition with a single sample per person". In: *2015 23rd European Signal Processing Conference (EUSIPCO)*, pp. 1148–1152. IEEE, ago. 2015. ISBN: 978-0-9928626-3-3. doi: 10.1109/EUSIPCO.2015.7362563. Disponível em: <http://ieeexplore.ieee.org/document/7362563/>.

[53] ZHAO, Y., LIU, Y., LIU, Y., etãl. "Face recognition from a single registered image for conference socializing", *Expert Systems with Applications*, v. 42, n. 3, pp. 973–979, fev. 2015. ISSN: 09574174. doi: 10.1016/j.eswa.2014.08.016. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0957417414004928>.

[54] YIN, F., JIAO, L., SHANG, F., etãl. "Double linear regressions for single labeled image per person face recognition", *Pattern Recognition*, v. 47, n. 4, pp. 1547–1558, abr. 2014. ISSN: 00313203. doi: 10.1016/j.patcog.2013.09.013. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0031320313003853>.

[55] DENG, W., HU, J., ZHOU, X., etãl. "Equidistant prototypes embedding for single sample based face recognition with generic learning and incremental learning", *Pattern Recognition*, v. 47, n. 12, pp. 3738–3749, dez. 2014. ISSN: 00313203. doi: 10.1016/j.patcog.2014.06.020. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0031320314002428>.

[56] YAN, H., LU, J., ZHOU, X., etãl. "Multi-feature multi-manifold learning for single-sample face recognition", *Neurocomputing*, v. 143, pp. 134–143, nov. 2014. ISSN: 09252312. doi: 10.1016/j.neucom.2014.06.012.

Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0925231214007541>.

[57] LIU, F., TANG, J., SONG, Y., etãl. "Local structure based sparse representation for face recognition with single sample per person". In: *2014 IEEE International Conference on Image Processing (ICIP)*, pp. 713–717. IEEE, out. 2014. ISBN: 978-1-4799-5751-4. doi: 10.1109/ICIP.2014.7025143. Disponível em: <http://ieeexplore.ieee.org/document/7025143/>.

[58] BORGI, M. A., LABATE, D., EL'ARBI, M., etãl. "Regularized Shearlet Network for face recognition using single sample per person". In: *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 514–518. IEEE, maio 2014. ISBN: 978-1-4799-2893-4. doi: 10.1109/ICASSP.2014.6853649. Disponível em: <http://ieeexplore.ieee.org/document/6853649/>.

[59] YANG, M., VAN, L., ZHANG, L. "Sparse Variation Dictionary Learning for Face Recognition with a Single Training Sample per Person". In: *2013 IEEE International Conference on Computer Vision*, pp. 689–696. IEEE, dez. 2013. ISBN: 978-1-4799-2840-8. doi: 10.1109/ICCV.2013.91. Disponível em: <http://ieeexplore.ieee.org/document/6751195/>.

[60] KVETON, B., VALKO, M. "Learning from a single labeled face and a stream of unlabeled data". In: *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, pp. 1–8. IEEE, abr. 2013. ISBN: 978-1-4673-5546-9 978-1-4673-5545-2 978-1-4673-5544-5. doi: 10.1109/FG.2013.6553720. Disponível em: <http://ieeexplore.ieee.org/document/6553720/>.

[61] DENG, W., HU, J., GUO, J. "In Defense of Sparsity Based Face Recognition". In: *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 399–406. IEEE, jun. 2013. ISBN: 978-0-7695-4989-7. doi: 10.1109/CVPR.2013.58. Disponível em: <http://ieeexplore.ieee.org/document/6618902/>.

[62] LU, J., TAN, Y.-P., WANG, G. "Discriminative Multimanifold Analysis for Face Recognition from a Single Training Sample per Person", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 35, n. 1, pp. 39–51, jan. 2013. ISSN: 0162-8828, 2160-9292. doi: 10.1109/TPAMI.2012.70. Disponível em: <http://ieeexplore.ieee.org/document/6175025/>.

[63] LU, J., TAN, Y.-P., WANG, G., etãl. "Image-to-Set Face Recognition Using Locality Repulsion Projections and Sparse Reconstruction-Based Simi-

larity Measure", *IEEE Transactions on Circuits and Systems for Video Technology*, v. 23, n. 6, pp. 1070–1080, jun. 2013. ISSN: 1051-8215, 1558-2205. doi: 10.1109/TCSVT.2013.2241353. Disponível em: <http://ieeexplore.ieee.org/document/6417014/>.

[64] ZHU, T., SHEN, F., ZHAO, J. "An incremental learning face recognition system for single sample per person". In: *The 2013 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–6. IEEE, ago. 2013. ISBN: 978-1-4673-6129-3 978-1-4673-6128-6. doi: 10.1109/IJCNN.2013.6707081. Disponível em: <http://ieeexplore.ieee.org/document/6707081/>.

[65] ZHUANG, L., YANG, A. Y., ZHOU, Z., etãl. "Single-Sample Face Recognition with Image Corruption and Misalignment via Sparse Illumination Transfer". In: *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3546–3553. IEEE, jun. 2013. ISBN: 978-0-7695-4989-7. doi: 10.1109/CVPR.2013.455. Disponível em: <http://ieeexplore.ieee.org/document/6619299/>.

[66] WANG, B., LI, W., LI, Z., etãl. "Adaptive linear regression for single-sample face recognition", *Neurocomputing*, v. 115, pp. 186–191, set. 2013. ISSN: 09252312. doi: 10.1016/j.neucom.2013.02.004. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0925231213001239>.

[67] ABDELMAKSOUD, M., NABIL, E., FARAG, I., etãl. "A Novel Neural Network Method for Face Recognition With a Single Sample Per Person", *IEEE Access*, v. 8, pp. 102212–102221, 2020. doi: 10.1109/ACCESS.2020.2999030.

[68] DING, Y., LIU, F., TANG, Z., etãl. "Uniform Generic Representation for Single Sample Face Recognition", *IEEE Access*, v. 8, pp. 158281–158292, 2020. doi: 10.1109/ACCESS.2020.3017479.

[69] ADJABI, I. "Combining hand-crafted and deep-learning features for single sample face recognition". In: *2022 7th International Conference on Image and Signal Processing and their Applications (ISPA)*, pp. 1–6, 2022. doi: 10.1109/ISPA54004.2022.9786302.

[70] GOODFELLOW, I. J., POUGET-ABADIE, J., MIRZA, M., etãl. "Generative Adversarial Networks". 2014. Disponível em: <https://arxiv.org/abs/1406.2661>.

[71] TRAN, L., YIN, X., LIU, X. "Representation Learning by Rotating Your Faces", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 41, n. 12, pp. 3007–3021, 2019. doi: 10.1109/TPAMI.2018.2868350.

[72] MAK, H. W. L., HAN, R., YIN, H. H. F. "Application of Variational Auto-Encoder (VAE) Model and Image Processing Approaches in Game Design", *Sensors*, v. 23, n. 7, pp. 3457, 2023. ISSN: 1424-8220. doi: 10.3390/s23073457. Disponível em: <http://dx.doi.org/10.3390/s23073457>.

[73] GIMENEZ, J. R., ZOU, J. "A Unified f-divergence Framework Generalizing VAE and GAN". 2022. Disponível em: <https://arxiv.org/abs/2205.05214>.

[74] CHEN, X., DUAN, Y., HOUTHOOFT, R., etãl. "InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets". 2016. Disponível em: <https://arxiv.org/abs/1606.03657>.

[75] LI, S., LI, H. "Deep Generative Modeling Based on VAE-GAN for 3D Indoor Scene Synthesis", *International Journal of Computer Games Technology*, v. 2023, pp. 1–11, 2023. ISSN: 1687-7047. doi: 10.1155/2023/3368647. Disponível em: <http://dx.doi.org/10.1155/2023/3368647>.

[76] CHENG, M., FANG, F., PAIN, C., etãl. "An advanced hybrid deep adversarial autoencoder for parameterized nonlinear fluid flow modelling", *Computer Methods in Applied Mechanics and Engineering*, v. 372, pp. 113375, 2020. ISSN: 0045-7825. doi: 10.1016/j.cma.2020.113375. Disponível em: <http://dx.doi.org/10.1016/j.cma.2020.113375>.

[77] MARTINEZ, A., BENAVENTE, R. *The AR face database*. Tech. Rep. 24, Comput. Vis. Center, Barcelona, Spain, Jun 1998.

[78] GEORGHIADES, A., BELHUMEUR, P., KRIEGMAN, D. "From few to many: illumination cone models for face recognition under variable lighting and pose", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 23, n. 6, pp. 643–660, 2001. doi: 10.1109/34.927464.

[79] PHILLIPS, P., MOON, H., RIZVI, S., etãl. "The FERET evaluation methodology for face-recognition algorithms", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 22, n. 10, pp. 1090–1104, 2000. doi: 10.1109/34.879790.

[80] GAO, W., CAO, B., SHAN, S., etãl. "The CAS-PEAL Large-Scale Chinese Face Database and Baseline Evaluations", *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, v. 38, n. 1, pp. 149–161, 2008. doi: 10.1109/TSMCA.2007.909557.

[81] HUANG, G. B., RAMESH, M., BERG, T., etãl. *Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments*. Relatório Técnico 07-49, University of Massachusetts, Amherst, October 2007.

[82] LIU, F., CHEN, D., WANG, F., etãl. "Deep learning based single sample face recognition: a survey", *Artificial Intelligence Review*, v. 56, n. 3, pp. 2723–2748. ISSN: 1573-7462. doi: 10.1007/s10462-022-10240-2. Disponível em: <https://doi.org/10.1007/s10462-022-10240-2>.

[83] RADFORD, A., METZ, L., CHINTALA, S. "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks". 2015. Disponível em: <https://arxiv.org/abs/1511.06434>.

[84] TURK, M., PENTLAND, A. "Face recognition using eigenfaces". In: *Proceedings. 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 586–591, 1991. doi: 10.1109/CVPR.1991.139758.

[85] WRIGHT, J., YANG, A. Y., GANESH, A., etãl. "Robust Face Recognition via Sparse Representation", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 31, n. 2, pp. 210–227, 2009. doi: 10.1109/TPAMI.2008.79.

[86] ZHANG, L., YANG, M., FENG, X. "Sparse representation or collaborative representation: Which helps face recognition?" In: *2011 International Conference on Computer Vision*, pp. 471–478, 2011. doi: 10.1109/ICCV.2011.6126277.

[87] ZHU, P., ZHANG, L., HU, Q., etãl. "Multi-scale Patch Based Collaborative Representation for Face Recognition with Margin Distribution Optimization". In: Fitzgibbon, A., Lazebnik, S., Perona, P., etãl. (Eds.), *Computer Vision – ECCV 2012*, pp. 822–835, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

[88] DENG, J., GUO, J., YANG, J., etãl. "ArcFace: Additive Angular Margin Loss for Deep Face Recognition", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1, 2021. doi: 10.1109/tpami.2021.3087709. Disponível em: <https://doi.org/10.1109%2Ftpami.2021.3087709>.

[89] RAMESH, A., PAVLOV, M., GOH, G., etãl. "Zero-Shot Text-to-Image Generation". 2021. Disponível em: <https://arxiv.org/abs/2102.12092>.

[90] RADFORD, A., KIM, J. W., HALLACY, C., etãl. "Learning Transferable Visual Models From Natural Language Supervision". 2021. Disponível em: <https://arxiv.org/abs/2103.00020>.

[91] BROWN, T. B., MANN, B., RYDER, N., etãl. "Language Models are Few-Shot Learners". 2020. Disponível em: <https://arxiv.org/abs/2005.14165>.

[92] KIM, M., LIU, F., JAIN, A., etãl. "DCFace: Synthetic Face Generation with Dual Condition Diffusion Model". 2023. Disponível em: <https://arxiv.org/abs/2304.07060>.

[93] SHOSHAN, A., BHONKER, N., KVIATKOVSKY, I., etãl. "GAN-Control: Explicitly Controllable GANs". 2021. Disponível em: <https://arxiv.org/abs/2101.02477>.

[94] ZAVAN, F. H. D. B., GASPARIN, N., BATISTA, J. C., etãl. "Face Analysis in the Wild". In: *2017 30th SIBGRAPI Conference on Graphics, Patterns and Images Tutorials (SIBGRAPI-T)*, pp. 9–16. IEEE, out. 2017. ISBN: 978-1-5386-0619-3. doi: 10.1109/SIBGRAPI-T.2017.11. Disponível em: <http://ieeexplore.ieee.org/document/8250221/>.

[95] WANG, D., OTTO, C., JAIN, A. K. "Face Search at Scale", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 39, n. 6, pp. 1122–1136, jun. 2017. ISSN: 0162-8828, 2160-9292. doi: 10.1109/TPAMI.2016.2582166. Disponível em: <http://ieeexplore.ieee.org/document/7494641/>.

[96] MAHMOOD, Z., MUHAMMAD, N., BIBI, N., etãl. "A REVIEW ON STATE-OF-THE-ART FACE RECOGNITION APPROACHES", *Fractals*, v. 25, n. 02, pp. 1750025, abr. 2017. ISSN: 0218-348X, 1793-6543. doi: 10.1142/S0218348X17500256. Disponível em: <http://www.worldscientific.com/doi/abs/10.1142/S0218348X17500256>.

[97] MARTINEZ, A. "Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 24, n. 6, pp. 748–763, jun. 2002. ISSN: 0162-8828. doi: 10.1109/TPAMI.2002.1008382. Disponível em: <http://ieeexplore.ieee.org/document/1008382/>.

[98] MOHAMMADZADE, H., HATZINAKOS, D. "Projection into Expression Subspaces for Face Recognition from Single Sample per Person", *IEEE Transactions on Affective Computing*, v. 4, n. 1, pp. 69–82, jan. 2013. ISSN: 1949-3045. doi: 10.1109/T-AFFC.2012.30. Disponível em: <http://ieeexplore.ieee.org/document/6313589/>.

[99] LEI, Y., GUO, Y., HAYAT, M., etãl. "A Two-Phase Weighted Collaborative Representation for 3D partial face recognition with single sample", *Pattern Recognition*, v. 52, pp. 218–237, abr. 2016. ISSN: 00313203. doi: 10.1016/j.patcog.2015.09.035. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0031320315003660>.

[100] ZHU, P., YANG, M., ZHANG, L., etãl. "Local Generic Representation for Face Recognition with Single Sample per Person". In: Cremers, D., Reid, I., Saito, H., etãl. (Eds.), *Computer Vision – ACCV 2014*, v. 9005, Springer International Publishing, pp. 34–50, Cham, 2015. ISBN: 978-3-319-16810-4 978-3-319-16811-1. doi: 10.1007/978-3-319-16811-1_3. Disponível em: <http://link.springer.com/10.1007/978-3-319-16811-1_3>.

[101] JIA, Q., FANG, C., WEN, D., etãl. "Generating face images under multiple illuminations based on a single front-lighted sample without 3D models". In: *2013 International Conference on Biometrics (ICB)*, pp. 1–6. IEEE, jun. 2013. ISBN: 978-1-4799-0310-8. doi: 10.1109/ICB.2013.6612997. Disponível em: <http://ieeexplore.ieee.org/document/6612997/>.

[102] HONG, S., IM, W., RYU, J., etãl. "SSPP-DAN: Deep domain adaptation network for face recognition with single sample per person". In: *Computer Vision and Pattern Recognition (cs.CV)*, pp. 825–829. IEEE, set. 2017. ISBN: 978-1-5090-2175-8. doi: 10.1109/ICIP.2017.8296396. Disponível em: <http://ieeexplore.ieee.org/document/8296396/>.

[103] MATTHEWS, B. "Comparison of the predicted and observed secondary structure of {T4} phage lysozyme", *Biochimica et Biophysica Acta (BBA) - Protein Structure*, v. 405, n. 2, pp. 442 – 451, 1975. ISSN: 0005-2795. doi: http://dx.doi.org/10.1016/0005-2795(75)90109-9. Disponível em: <http://www.sciencedirect.com/science/article/pii/0005279575901099>.

[104] YOUDEN, W. J. "Index for rating diagnostic tests", *Cancer*, v. 3, n. 1, pp. 32–35, 1950. ISSN: 1097-0142. doi: 10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO2-3. Disponí-

vel em: <http://dx.doi.org/10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3>.

[105] PHILLIPS, P., WECHSLER, H., HUANG, J., etãl. "The FERET database and evaluation procedure for face-recognition algorithms", *Image and Vision Computing*, v. 16, n. 5, pp. 295 – 306, 1998. ISSN: 0262-8856. doi: https://doi.org/10.1016/S0262-8856(97)00070-X. Disponível em: <http://www.sciencedirect.com/science/article/pii/S026288569700070X>.

[106] HEO, J., SAVVIDES, M. "3-D Generic Elastic Models for Fast and Texture Preserving 2-D Novel Pose Synthesis", *IEEE Transactions on Information Forensics and Security*, v. 7, n. 2, pp. 563–576, April 2012. ISSN: 1556-6013. doi: 10.1109/TIFS.2012.2184755.

[107] FENG, Y., WU, F., SHAO, X., etãl. "Joint 3D Face Reconstruction and Dense Alignment with Position Map Regression Network". In: *ECCV*, 2018.

[108] KINGMA, D. P., WELLING, M. "Auto-Encoding Variational Bayes". 2013. Disponível em: <https://arxiv.org/abs/1312.6114>.

[109] HONG, S., IM, W., RYU, J., etãl. "SSPP-DAN: Deep Domain Adaptation Network for Face Recognition with Single Sample Per Person". 2017. Disponível em: <https://arxiv.org/abs/1702.04069>.

[110] HUANG, G. B., MATTAR, M., BERG, T., etãl. "Labeled Faces in the Wild: A Database forStudying Face Recognition in Unconstrained Environments". In: *Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition*, Marseille, France, out. 2008. Erik Learned-Miller and Andras Ferencz and Frédéric Jurie. Disponível em: <https://hal.inria.fr/inria-00321923>.